Full Title: Validating an algorithm from a trunk-mounted wearable sensor for detecting stroke events in tennis

Submission Type: Original Investigation

Authors: Thomas Perri[1,2], Machar Reid[2], Alistair Murphy[2], Kieran Howle[3] and Rob Duffield[1]

Institutional Affiliations:

[1] School of Sport, Exercise and Rehabilitation, Faculty of Health, University of Technology Sydney, Sydney, Australia

[2] Sports Science and Sports Medicine Unit, Tennis Australia, Melbourne, Australia

[3] Catapult Sports, Melbourne, Australia

Corresponding Author and Address:

Thomas Perri
Tennis South Australia
P.O. Box: 43
North Adelaide 5006, South Australia
Australia

Phone: 0439 153 025

Email: TPerri@Tennis.com.au

Abstract Word Count: 196

Text-Only Word Count: 3631

Number of Figures: 0

Number of Tables: 5

Abstract

This study analysed the accuracy of a prototype algorithm for tennis stroke detection from wearable technology. Strokes from junior-elite tennis players over ten matches were analysed. Players wore a GPS unit containing an accelerometer, gyroscope and magnetometer. Manufacturer-developed algorithms determined stoke type and count (forehands, backhands, serves and other). Matches were video recorded to manually code ball contacts and shadow swing events for forehands, backhands and serves and further by stroke classifications (i.e., drive, volley, slice, end-range). Comparisons between algorithm and coding were analysed via ANOVA and Bland-Altman plots at the match-level and error rates for specific stroke-types. No significant differences existed for stroke count between the algorithm and manual coding (p>0.05). Significant (p<0.0001) overestimation of "Other" strokes were observed from the algorithm, with no difference in groundstrokes and serves (p>0.05). Serves had the highest accuracy of all stroke types (≥98%). Forehand and backhand "drives" were the most accurate (>86%), with volleys mostly undetected (58-60%) and slices and end-range strokes likely misclassified (49-51%). The prototype algorithm accurately quantifies serves and forehand and backhand "drives" and serves. However, underestimations of shadow swings and overestimations of "other" strokes suggests strokes with reduced trunk rotation have poorer detection accuracy.

Key Words: racquet sports, accuracy, external load, wearable sport technology

Introduction

Competitive tennis involves combinations of high-intensity intermittent court movement concomitant with the execution of stroke play, and both are important insights for optimal training and match preparation (Elliot, Reid & Whiteside 2019; Reid & Duffield 2014). Quantifying these physical loads via traditional methods, such as wearable microtechnology (i.e., global positioning systems [GPS] and accelerometers), has proven challenging in tennis as the location of the unit at the base of the cervical spine may be limited in inferring the specific mechanical demands of limb-dominant stroke play (Reid et al. 2019). Although wrist-worn or racquet-mounted sensors provide accurate information on basic stroke type volumes (Genevois et al. 2018; Keaney & Reid 2020; Myers et al. 2019; Whiteside et al. 2017), they lack the reporting of player movement loads in their final output, which form an integral part of a tennis player's overall mechanical load (Reid et al. 2019). Accordingly, greater understanding of tennis physical demands require measurements of both on-court activity (i.e., movement load) in combination with the stroke type and volumes (i.e., hitting load) similar to observations on cricket bowling (McNamara et al. 2015). Indeed, recent developments in commercially available, yet unvalidated GPS and micro electro-mechanical systems (MEMS) technologies, could reveal a way forward for tennis in the simultaneous capture of stroke events and movement metrics from a single device.

Studies seeking to understand tennis stroke events and associated physical loading of the upper-body have been performed mostly using wrist-worn sensors. Kos and colleagues (2016) evaluated the recall of a wrist-worn sensor for forehands, backhands and serve strokes during closed and open settings and reported respective accuracies of 96%, 98% and 100%. However, they did not include volleys, slices or smashes in their algorithm, with later research on wrist-worn sensors showing poorer recall of these stroke types (≈80%) compared to groundstrokes

(≥98%) during typical training drills and simulated match-play (Whiteside et al. 2017). This is common in tennis, where the recall of basic stroke types (i.e., forehand, backhand and serve) show acceptable accuracies from wrist-worn sensors (Myers et al. 2019; Whiteside et al. 2017), but perform with reduced accuracy when detecting explicit stroke types (if reported at all). Keaney and Reid (2020) highlight this point, whereby a racquet-mounted sensor showed high accuracy for total strokes, but poor differentiation between specific stroke types (i.e. groundstroke, volley etc.), postulated to result from sensor quality and location. Indeed, the wider issue existing in tennis load measurement is not solely on limitations of stroke event detection, but rather the lack of application of hitting load measures in commercial wearables that already provide insights on whole-body movement.

In team sport, the use of wearable sensors (i.e., GPS devices) are commonly used for quantifying running loads, but also increasingly for the specific detection of other load producing events, such as throwing and collisions (Crang et al. 2021). Indeed, the raw outputs from in-built accelerometer (linear accelerations), gyroscope (angular accelerations) and magnetometer (unit orientation) within commercially available GPS units have been used to develop algorithms to detect acute cricket bowling events with high accuracy (sensitivity >95%), though reductions in accuracy are observed in official matches due to the introduction of other upper-body events such as fielding (Jowitt et al. 2020; McNamara et al. 2015). Similar results have been observed in handball, where high event detection sensitivity (84-100%) was present in a controlled setting but reduced during match-play (sensitivity 52-91%) (Skejo et al. 2021). These results highlight the challenges of identifying multi-dimensional sporting actions when algorithms developed from a single body location are relied upon for event detection (Ishii et al. 2021). Further limitations have been observed in event detection algorithms for rugby and Australian football (AF), where positional differences and tackling technique negatively impacts the true reporting of collision events (Gastin et al. 2014; Reardon et al. 2017). Whilst these limitations reveal the need for rigorous testing of event detection algorithms, the data available from such devices provide richer insights into the mechanical demands of technical actions to benefit athlete preparation and management, which is lacking in tennis (Reid et al. 2019). In this regard, the aims of this study were to; 1) evaluate the accuracy of a prototype algorithm to detect tennis strokes from a commercially available trunk-mounted wearable sensor and 2) evaluate the accuracy of this event detection across different stroke types. It was hypothesised that serves will have the highest detection accuracy due to their distinctive trunk rotation profile, while strokes with less trunk rotation (such as end-range forehands and backhands) and shadow swings would be detected less accurately.

Methods

Ten matches from eight junior-elite male tennis players (age 15.5 ± 1.6y) were analysed in this study. The players were part of Tennis Australia's National Academy program and trained and competed as per the guidelines suggested in Tennis Australia's athlete development matrix including ≈20 h of on-court training per week, ≈6 h off-court training per week and were competing in regular International Tennis Federation (ITF) sanctioned junior tournaments. All players in the study were right-handed with a double-handed backhand. A Human Research Ethics Committee (HREC) gave ethical approval for the methods used in this study (ETH19-4062).

Matches were analysed across an ITF sanctioned Grade 5 tournament and two National Championships in Australia for singles matches held on hard and grass courts during the 2019 season. All matches were best-of-three sets in accordance with the rules of the ITF (ITF 2016). Recording of matches was performed using video cameras (HDR-CX700VE, Sony, Japan) that were positioned 10 m above and 6 m behind the baseline in accordance with previous protocols (Murphy et al. 2014; Perri et al. 2018).

Stroke events were captured via a wearable device (Catapult OptimEye S5, Catapult Sports, Melbourne) with an in-built triaxial accelerometer, magnetometer and gyroscope. The device was worn between the scapulae in the manufacturer-designed harness that minimised movement on the skin (McLean et al. 2018), with a combined weight of 102g. The manufacturer has developed a prototype algorithm (White Paper, Catapult Sports) to recognise the swing and movement pattern of stroke events, which can be used to count strokes for workload monitoring. The prototype algorithm from the manufacturer has been internally investigated through implementing machine learning models that classify four categories of strokes (‘Forehand [FH] Drive’, Backhand [BH] Drive’, ‘Serve’ and ‘Other stroke’) based on absolute rotation yaw values. These unpublished investigations have shown respective accuracies of the aforementioned stroke categories to be 94%, 96.5%, 99.9% and 83.5% (Personal Communication, Catapult Sports). Raw accelerometer data from the wearable units were downloaded via custom software (OpenField 2.3.4, Catapult Sports, Melbourne) and processed by company staff using customised algorithms in RStudio (RStudio, 1.1.463, RStudio, Inc.). All data provided to the manufacturer were de-identified. The processed data detailing the Coordinated Universal Time (UTC) (hh:mm:ss) of each stroke and detail on its respective category was provided to the research team in a comma separated values (.csv) file to be analysed, without knowledge of the match, player or access to video footage. All strokes captured by the wearable device were presented in consecutive chronological rows to then be compared with the video footage.

Manual notation of each match was performed in the months after the conclusion of the final match and were analysed by a coder with five years of experience coding tennis matches and a coefficient of variation (CV) <2% from previous work notating stroke counts and technical errors during tennis training and match-play (Perri et al. 2018). The manually coded strokes were collated in the .csv file with algorithm stroke outcomes. Strokes were coded manually from the video footage in accordance with their basic type of stroke (i.e., forehand, backhand, serve) and further detailed by their specific spin or trajectory (i.e., rally, slice, volley, drop shot) (Table 1) and whether they were in "live play" or in-between points. Strokes that did not meet these general classifications (i.e., an underarm stroke to pass ball back to server) were coded as an "Other stroke". As the Catapult algorithm does not differentiate between smashes and serves, smashes were manually coded as an "Other stroke". Racquet swings, which still resemble a forehand or backhand drive but without ball contact, were coded in respective "forehand" or "backhand" categories (Table 1). Reliability of the coding method was determined through re-coding a randomly selected match that was separated by one month. A total of 624 stroke events existed in the match, with 616 strokes correctly matching the previous coding method (CV = 0.9%).

TABLE 1 NEAR HERE

Data was further prepared for analysis using a customised Microsoft Excel (Microsoft Excel, 16.49, Microsoft, Washington) spreadsheet. Strokes detected by the Catapult algorithm were compared to the manually coded data across multiple levels. The dataset was first analysed to denote whether a stroke event was detected by the wearable device. This was then further scrutinised to classify whether the algorithm correctly identified the type of stroke (i.e., forehand, backhand or serve). In this example, a stroke labelled a "FH Drive" by the algorithm and manually coded as a forehand volley was considered to be correct from the algorithms perspective as it does not discriminate between stroke types beyond rally strokes. Instances where the algorithm detected a forehand, backhand or serve but classified it as an "Other stroke", this was categorised as an incorrect classification. However, if "Other stroke" was recorded by the algorithm and a smash or stroke not meeting the previous criteria was played, this was considered to be correct.

Statistical Analysis

All statistical analysis was performed in the R language (RStudio, 1.1.463, RStudio, Inc.). Initial comparisons between wearable sensor and manual coding at a match level for strokes and shadow swing events were performed via a one-way analysis of variance (ANOVA), with significance set at p = 0.05. To calculate the levels agreement between wearable sensor and manual coding at a stroke level (n=5349), Bland-Altman limits of agreement (LOA) (standard error of the means [SEM]) with 95% confidence intervals (CI) were reported. To calculate absolute and relative measures of error for stroke and swing events across the respective four stroke categories (forehand, backhand, serve and other), the number of correctly classified strokes from the wearable sensor was divided by the total number of events in that category and multiplied by 100. This calculation was also applied to strokes detailed in Table 1 to determine specific limitations of the wearable technology versus manually coded categories.

Results

A total of 5349 stroke/swing events were identified through manual coding, with 5119 events detected by the wearable unit. Within the 5119 events detected by the wearable unit, 204 were classified as false positives via cross-referencing the video footage. At a match-level, Table 2 shows the strokes and shadow swings during match play recorded in the dataset. No significant differences were observed for total strokes (*p*=0.56) as well as the four stroke categories categorised by the wearable sensor (*p*>0.05). Bland-Altman analysis revealed a mean bias for an underestimation of total stroke counts (-67.80±35.96[-93.52 to -42.08]), by the wearable sensor. The wearable sensor showed the lowest mean bias for serves (-4.00±4.24[-7.04 to -0.96], with highest bias levels observed for the "other stroke" category (30.70±17.13[18.45 to 42.95]). Analysis of shadow swing data showed a significant main effect for "Other strokes", with an overestimation by the wearable sensor compared to manual coding (*p*=0.0001). Significant main effects were also observed for forehand, backhand and other shadow swings between the two methods, with the wearable device significantly underestimating shadow swings played (*p*=0.001, *p*=0.002 and *p*=0.002, respectively). Bland-Altman analysis for shadow strokes revealed greatest bias for the "other stroke" category, with an overestimation of strokes from the wearable unit (8.80±7.90[3.15 to 14.45]). Mean bias for total, forehand and backhand shadow swings showed underestimations ranging from -3.10 to -6.10.

*****TABLE 2 NEAR HERE*****

Table 3 shows the stroke and swing events observed in all matches by stroke type and respective absolute error between the wearable device and the video coding. Error rates <20% were observed for total stroke and swing actions for the wearable device, with error rates of ≤11% for forehand and backhand stroke events. Lowest error rates (≤2%) were observed for serve stroke events. Poorest classification (>75% error) was evident in swings with no ball contact and "other strokes".

*****TABLE 3 NEAR HERE*****

Detailed descriptions of the type of forehand stroke detected by the wearable sensor based on manual coding are detailed in Table 4. Forehand strokes classified as “rally” events showed the lowest error rates (6%), with the “dig”, “shadow” and “volley” classifications revealing the highest error (>90%). For “end-range” strokes, a greater proportion of errors existed due to misclassification rather than being undetected. Further, this pattern also existed for forehand slices however, forehand volley strokes were more likely to be undetected by the wearable’s algorithm. Additionally, strokes categorised as a “smash” are presented in Table 4 and reveals a lower error rate when smash strokes are classified as “other strokes” versus being classified as a serve (percent correct=62% vs. 26%, respectively).

*****TABLE 4 NEAR HERE*****

For specific types of backhand strokes, the “drive” category had the lowest error rate (14% error) compared to all other backhand strokes categories. Backhand slice and end-range strokes respectively contributed the next highest proportions to the overall backhand stroke/swing count, with both sub-types revealing errors to be predominantly a result of algorithm misclassification versus being undetected (49% and 56% error, respectively). Similar to forehand strokes, backhand volleys and “dig” actions were not accurately classified by the algorithm (60% and 50% error, respectively).

*****TABLE 5 NEAR HERE*****

Discussion

This study is the first attempt to validate the accuracy of a prototype algorithm from a wearable device to detect stroke events in tennis. The present results indicate that total strokes from the wearable sensor show acceptable accuracy for general stroke classification and count during tennis match-play, albeit with caution for more specific stroke classifications. In particular, strokes with more pronounced trunk rotation, such as the forehand (Landlinger et al. 2010) and backhand drive (Reid & Elliot 2002) and serves (Abrams et al. 2011) show acceptable accuracy in stroke count. However, stroke events such as volleys, end-range and slices experienced higher error rates resulting from non-detection or misclassification, and likely due to the prototype algorithm not being specifically trained to detect these specific stroke patterns. These findings support our hypotheses and demonstrates the usefulness of current technology to quantify high-load actions such as serving. Future enhancements to accurately identify stroke events with increased lower-limb dominance and reduced trunk rotation intensity (i.e., end-range) would provide further insights for load management strategies in tennis.

Accurate technology to detect tennis stroke type and counts in tennis can guide both training prescription and match preparation (Shanley & Myers 2019). The results of the present study indicate no differences in total stroke volume at the match level, which can be interpreted as a strength of the investigated technology. Tennis match-play is typified by the exchange of forehand and backhand drive or “rally” strokes, which contribute ≈73% of total strokes (Whiteside & Reid 2017); though training often emphasises higher volumes of forehand strokes and separate identification of these events is required (Genevois et al. 2018). Nevertheless, these strokes are characterised by distinct kinematic (axial rotation) profiles (Reid & Elliot 2002; Reid, Elliott & Crespo 2013), which may help to explain this sensor’s accuracy in detecting these more frequent aspects of tennis stroke play. Accordingly, implementing this wearable technology in tennis training environments could improve understanding of the mechanical stresses imposed from repeated forehand and backhand groundstrokes (Johansson et al. 2021).

The underestimation of shadow swings suggests that ball impact somehow plays an important role in event detection. Previous work on wrist-worn sensors has identified that 'swing-type' movements of the arm can be captured in the absence of impact (Hadzic, Germic & Filipcic 2021), which may suggest wearing the device on the torso is a limitation in itself rather that resulting from the algorithm. The practical implications of this finding are questionable though, as recent load management practices have not considered "shadow" technical actions in their reporting. Further, this may be seen as a positive from the algorithm's perspective given the prototype was not designed to detect these events but rather quantify the load associated with these movement patterns.

The application of a prototype algorithm from wearable sensor data showed no differences in serve count compared to manual coding. The accuracy of serve detection (≥98%) compares favourably with similar notation using wrist and racquet-mounted sensors (Kos et al. 2016; Yang et al. 2017). This is a notable practical application for tennis given the relevance of the serve to match-play success (Kolman et al. 2019) and trunk and upper-limb injury rehabilitation (Campbell et al. 2014; Sombelon et al. 2017). Hence, coaches and support staff can be confident in using this technology to monitor serve loads for tennis. This is especially relevant given that chronic upper limb and trunk injuries are often linked to the serve, which likely results from inappropriate overload (Pluim et al. 2006). Interestingly, the detection of "smashes" was comparatively poor and likely explained by the less predictable nature of stroke mechanics during match-play. Despite the low contribution of smashes to overall stroke volumes in tennis matches (Whiteside & Reid 2017), the imposed load resulting from overhead strokes (i.e., serve and smash) can have a negative impact on shoulder joint health if not managed appropriately (Kekelekis et al. 2020) and thus, total overhead load may be underestimated from the wearable device.

Forehand and backhand strokes classified as “drives” showed the highest detection accuracy across respective forehand and backhand strokes. This finding is unsurprising given their respective consistency in the direction and magnitude of trunk rotation (Landlinger et al. 2010; Reid & Elliot 2002). Therefore, the present algorithm appears well suited to detecting these stroke types and provides a suitable alternative to wrist-mounted sensors, which also produce high detection accuracies for forehand and backhand drives (Myers et al. 2019; Whiteside et al. 2017). However, the detection accuracy of the algorithm deteriorated for other strokes (i.e., volleys, slices) that have reduced or more variable magnitudes and intensities of trunk rotation (Chow, Carlton, Chae, et al. 1999; Chow, Carlton, Lim, et al. 1999) and presumably affects the distinctiveness of their kinematic signatures. Previous research has identified that although trunk rotation is involved in the execution of forehand volleys, the resultant force patterning is subject to considerable variability as a result of movement at the racquet arm (Crespo 1999; Roetert & Groppel 2001). This movement variability may complicate the precision and reliability of the algorithm’s detection of volleys, while the placement of the sensor (between scapulae) may be suboptimal for strokes characterised by lower magnitudes of trunk rotation. From a training load monitoring perspective, unregistered volleys are likely to have limited impact on training load decisions given they contribute ≈1% of total strokes in elite-level tennis (Whiteside & Reid 2017).

The hypothesis that detection accuracy for low-intensity strokes (ie. low swing velocity or trunk rotation) would be comparatively poor was further supported when evaluating the sensor’s ability to detect “end-range” strokes. End-range shots represent an integrated load profile in tennis given they are played under considerable time-pressure and, generally, after the player has moved at high-speed (Pieper, Exler & Weber 2007). From this perspective, these strokes are especially relevant for load monitoring and management in tennis, though were not detected by the algorithm given the low-level of torso rotation. End-range shots have high mechanical stresses at the lower limb concomitant with the load imposed on upper limb structures to execute the hitting action (Giles & Reid 2021). The fact that end-range shots are often characterised by alterations in upper limb involvement, likely manifesting as increased variability of accompanying trunk rotation may suggest an area for enhancement of the prototype algorithm given its current focus on recognising distinct "drive" groundstroke patterns and serving actions.

Limitations

Whilst this study is novel in that no previous literature has investigated the validity of applying this prototype event detection algorithm to trunk-mounted wearable sensor data in tennis, the study is limited by the homogenous sample of players (i.e., male, right-handed and double-handed backhand). Indeed, the prototype algorithm was developed on data from elite senior males and may not be specific to the developing stroke patterns of junior players. Further, the stroke categories presented in Table 1 may be seen as a possible limitation given the subjective nature of such classifications; however, these were devised through discussion with tennis coaches holding the highest level of qualification in Australia for longer than 10 years in combination with previous expert reports (Crespo & Miley 1998). These descriptions are more detailed than the prototype algorithm and is acknowledged as a limitation of our methods. Additionally, assessing coder reliability from one match and for one coder conforms with previous methods (Gastin et al. 2014), though may be viewed as a potential limitation. The approach to classifying "shadow" swings may also be considered as a limitation given difficulties in developing algorithms to detect actions mimicking detectable events. Lastly, the confidentiality of the proporiety algorithm did not allow for alterations of timing windows or specific signal detection for assessing possible modifications to the current device.

Conclusions

This study aimed to validate the tennis stroke event detection accuracy of a prototype algorithm developed from a trunk-mounted wearable sensor. Overall stroke events in match-play are accurately detected from the sensor, though inaccuracies can result through misclassification or non-detection of volleys, slices, and end-range strokes, which contribute the total of forehand and backhand strokes. Highest accuracies for specific stroke classifications were observed for serves, suggesting support staff can confidently monitor serve loads from the present device. Future enhancements toward the detection of stroke events with limited trunk movement (i.e., volleys, slices) and those occurring at high movement speeds (i.e., end-range) would provide greater insights into understanding the mechanical demands of stroke play.

References

Abrams, G.D., Sheets, A.L., Andriacchi, T.P. & Safran, M.R. 2011, 'Review of tennis serve motion analysis and the biomechanics of three serve types with implications for injury', *Sports Biomechanics*, vol. 10, no. 4, pp. 378-90.

Campbell, A., Straker, L., O'Sullivan, P., Elliot, B. & Reid, M. 2014, 'Lumbar loading in the elite adolescent tennis serve: link to low back pain', *Medicine and Science in Sports and Exercise*, vol. 45, no. 8, pp. 1562-8.

Chow, J.W., Carlton, L.G., Chae, W.-S., Shim, J.-H., Lim, Y.-T. & Kuenster, A.F. 1999, 'Movement characteristics of the tennis volley', *Medicine & Science in Sports & Exercise*, vol. 31, no. 6, pp. 855-63.

Chow, J.W., Carlton, L.G., Lim, Y.-T., Shim, J.-H., Chae, W.-S. & Kuenster, A.F. 1999, 'Muscle activation during the tennis volley', *Medicine & Science in Sports & Exercise*, vol. 31, no. 6, pp. 846-54.

Crang, Z.L., Duthie, G.M., Cole, M.H., Weakley, J., Hewitt, A. & Johnston, R.D. 2021, 'The validity and reliability of wearable microtechnology for intermittent team sports: a systematic review', *Sports Medicine*, vol. 51, no. 3, pp. 549-65.

Crespo, M. 1999, 'What tennis research tells us about…biomechanics of volleys and approach shots', *ITF Coaching and Sport Science Review*, vol. 7, no. 17, pp. 15-6.

Crespo, M. & Miley, D. 1998, *ITF advanced coaches manual*, International Tennis Federation.

Elliot, B., Reid, M. & Whiteside, D. 2019, 'Biomechanics of groundstrokes and volleys', in G. Di Giacomo, T. Ellenbecker & W.B. Kibler (eds), *Tennis Medicine*, Springer, Switzerland, pp. 17-42.

Gastin, P.B., McLean, O.C., Breed, R.V. & Spittle, M. 2014, 'Tackle and impact detection in elite Australian football using wearable microsensor technology', *Journal of Sports Sciences*, vol. 32, no. 10, pp. 947-53.

Genevois, C., Amsallem, C., Brandli, C. & Rogowski, I. 2018, 'Using inertial sensors to monitor on-court tennis training sessions', *ITF Coaching and Sport Science Review*, vol. 25, no. 75, pp. 18-9.

Giles, B. & Reid, M. 2021, 'Applying the brakes in tennis: how entry speed affects the movement and hitting kinematics of professional tennis players', *Journal of Sports Sciences*, vol. 39, no. 3, pp. 259-66.

Hadzic, V., Germic, A. & Filipcic, A. 2021, 'Validity and reliability of a novel monitoring sensor for the quantification of the hitting load in tennis', *PLoS One*, vol. 16, no. 7.

Ishii, S., Yokokubo, A., Luimula, M. & Lopez, G. 2021, 'ExerSense: physical exercise recognition and counting algorithm from wearables robust to positioning', *Sensors*, vol. 21, no. 1.

ITF 2016, *ITF Rules of tennis*, viewed September 1st 2019, <http://www.itf.tennis.com/about/organisation/rules.aspx>.

Johansson, F., Gabbett, T., Svedmark, P. & Skillgate, E. 2021, 'External training load and the association with back pain in competitive adolescent tennis players: results from the SMASH cohort study', *Sports Health*.

Jowitt, H.K., Durussel, J., Brandon, R. & King, M. 2020, 'Auto detecting deliveries in elite cricket fast bowlers using microsensors and machine learning', *Journal of Sports Sciences*, vol. 38, no. 7, pp. 767-72.

Keaney, E.M. & Reid, M. 2020, 'Quantifying hitting activity in tennis with racket sensors: a new dawn or false dawn?', *Sports Biomechanics*, vol. 19, no. 6, pp. 831-9.

Kekelekis, A., Nikolaidis, P.T., Moore, I.S., Rosemann, T. & Knechtle, B. 2020, 'Risk factors for upper limb injury in tennis players: a systematic review', *International Journal of Environmental Research and Public Health*, vol. 17, no. 8.

Kolman, N.S., Kramer, T., Elferink-Gemser, M.T., Hujigen, B.C.H. & Visscher, C. 2019, 'Technical and tactical skills related to performance levels in tennis: a systematic review', *Journal of Sports Sciences*, vol. 37, no. 1, pp. 108-21.

Kos, M., Zenko, J., Vlaj, D. & Kramberger, I. 2016, 'Tennis stroke detection and classification using miniature wearable IMU device', *The 23rd International Conference on Systems, Signals and Image Processing*, Bratislava, Slovakia.

Landlinger, J., Lindinger, S.J., Stoggl, T., Wagner, H. & Muller, E. 2010, 'Kinematic differences of elite and high-performance tennis players in the cross court and down the line forehand', *Sports Biomechanics*, vol. 9, no. 4, pp. 280-95.

McLean, B.D., Cummins, C., Conlan, G., Duthie, G.M. & Coutts, A.J. 2018, 'The fit matters: influence of accelerometer fitting and training drill demands on load measures in rugby league players', *International Journal of Sports Physiology and Performance*, vol. 13, no. 8, pp. 1083-9.

McNamara, D.J., Gabbett, T.J., Chapman, P., Naughton, G. & Farhart, P. 2015, 'The validity of microsensors to automatically detect bowling events and counts in cricket fast bowlers', *International Journal of Sports Physiology and Performance*, vol. 10, no. 1, pp. 71-5.

Murphy, A.P., Duffield, R., Kellett, A. & Reid, M. 2014, 'A descriptive analysis of internal and external loads for elite-level tennis drills', *International Journal of Sports Physiology and Performance*, vol. 9, no. 5, pp. 863-70.

Myers, N.L., Kibler, W.B., Axtell, A.H. & Uhl, T.L. 2019, 'The Sony Smart Tennis Sensor accurately measures external workload in junior tennis players', *Sports Science & Coaching*, vol. 14, no. 1, pp. 24-31.

Perri, T., Norton, K.I., Bellenger, C.R. & Murphy, A.P. 2018, 'Training loads in typical junior-elite tennis training and competition: implications for transition periods in a high-performance pathway', *International Journal of Performance Analysis in Sport*, vol. 18, no. 2, pp. 327-38.

Pieper, S., Exler, T. & Weber, K. 2007, 'Running speed loads on clay and hard courts in world class tennis', *Medicine and Science in Tennis*, no. 2.

Pluim, B.M., Staal, J.B., Windler, G.E. & Jayanthi, N. 2006, 'Tennis injuries: occurrence, aetiology, and prevention', *British Journal of Sports Medicine*, vol. 40, no. 5, pp. 415-23.

Reardon, C., Tobin, D.P., Tierney, P. & Delahunt, E. 2017, 'Collision count in rugby union: a comparison of micro-technology and video analysis methods', *Journal of Sports Sciences*, vol. 35, no. 20, pp. 2028-34.

Reid, M., Cormack, S.J., Duffield, R., Kovalchik, S., Crespo, M., Pluim, B. & Gescheit, D.T. 2019, 'Improving the reporting of tennis injuries: the use of workload data as the denominator?', *British Journal of Sports Medicine*, vol. 53, no. 16, pp. 1041-2.

Reid, M. & Duffield, R. 2014, 'The development of fatigue during match-play tennis', *British Journal of Sports Medicine*, vol. 48, pp. S7-S11.

Reid, M. & Elliot, B. 2002, 'The one- and two-handed backhands in tennis', *Sports Biomechanics*, vol. 1, no. 1, pp. 47-68.

Reid, M., Elliott, B. & Crespo, M. 2013, 'Mechanics and learning practices associated with the tennis forehand: a review', *Journal of Sports Science & Medicine*, vol. 12, no. 2, pp. 225-31.

Roetert, E.P. & Groppel, J.L. 2001, 'Biomechanics of the volley', *ITF Coaching and Sport Science Review*, vol. 9, no. 24, pp. 15-6.

Shanley, E. & Myers, N.L. 2019, 'Understanding load in baseball and tennis', in W.B. Kibler & A.D. Sciascia (eds), *Mechanics, pathmechanics and injury in the overhead athlete*, Springer.

Skejo, S.D., Liaghat, B., Jakobsen, C.C., Moller, M., Bencke, J., Papi, G., Kunwald, N.P. & Sorensen, H. 2021, 'Quantifying throwing load in handball: a method for measuring the number of throws', *SportRxiv*.

Sombelon, G.N., Myers, N.L., Westgate, P., Smith, B.J. & Kibler, W.B. 2017, 'Tennis serve volume and its relationship to injury in professional women's tennis players', *Journal of Athletic Training*, vol. 52, no. 6, p. 185.

Whiteside, D., Cant, O., Connolly, M. & Reid, M. 2017, 'Monitoring hitting load in tennis using inertial sensors and machine learning', *International Journal of Sports Physiology and Performance*, vol. 12, no. 9, pp. 1212-7.

Whiteside, D. & Reid, M. 2017, 'External match workloads during the first week of Australian Open tennis competition', *International Journal of Sports Physiology and Performance*, vol. 12, no. 6, pp. 756-63.

Yang, D., Tang, J., Huang, Y., Xu, C., Li, J., Hu, L., Shen, G., Liang, C.-J.M. & Liu, H. 2017, 'TennisMaster: an IMU-based online serve performance evaluation system', paper presented to the *the 8th Augmented Human International Conference*.

508 **Table 1.** Manually coded stroke definitions

Stroke Type	Definition
Drive	A typical 'topspin' or 'flat' forehand or backhand stroke. Also included 'offensive' lobs.
End-Range	A forehand or backhand stroke, typically played with the racquet arm at full stretch and in a wide position of the court.
Volley	A forehand or backhand stroke played 'on-the-full' with no bounce prior to the stroke.
Drop shot	A disguised forehand stroke that is played with the aim of the ball dropping short into the opposing player's side of the court.
Block	A forehand or backhand stroke often played by the returner in response to a fast serve.
Slice	A forehand or backhand stroke played where the racquet's forward-swing trajectory imparts backspin to the ball.
Dig	Strokes played with limited forward-swing and often are more vertical with a low to high 'redirect' trajectory
Shadow	Any stroke pattern played in absence of a ball being contacted.

509 (Crespo & Miley 1998)

Table 2. Mean ± SD of strokes and shadow swings captured per match from Catapult algorithms versus manually coded strokes and shadow swings in match play

Stroke Type	Catapult	Coded	*p*	Bias (Mean ± SD)	Bias (95% CI)
Total strokes (n/match)	431 ± 219	464 ± 231	0.562	-67.80 ± 35.96	-93.52 to -42.08
Forehand strokes (n/match)	172 ± 114	192 ± 119	0.655	-18.70 ± 27.79	-38.58 to 1.18
Backhand strokes (n/match)	128 ± 64	169 ± 84	0.073	-45.10 ± 22.26	-61.03 to -29.17
Serves/Overhead strokes (n/match)	93 ± 43	92 ± 42	0.780	-4.00 ± 4.24	-7.04 to -0.96
Other strokes (n/match)	39 ± 22**	11 ± 8	0.0001	30.70 ± 17.13	18.45 to 42.95
Total shadow swings (n/match)	10 ± 7	15 ± 9	0.451	-3.10 ± 4.07	-6.01 to -0.19
Forehand shadow swings (n/match)	0 ± 1*	5 ± 4	0.001	-6.10 ± 4.41	-9.25 to -2.95
Backhand shadow swings (n/match)	1 ± 1*	8 ± 6	0.002	-5.80 ± 4.61	-9.10 to -2.50
Serve/Overhead shadow swings (n/match)	0 ± 0	0 ± 0	N/A	N/A	N/A
Other shadow swings (n/match)	8 ± 7*	0 ± 1	0.002	8.80 ± 7.90	3.15 to 14.45

*significantly different from manually coded strokes (p<0.01)
**significantly different from manually coded strokes (p<0.001)

Table 3. Total count of Catapult versus manually coded swing patterns

Stroke Type	Coded	Catapult	Correctly Classified	Incorrectly Classified	%Error [Correctly Classified / Catapult] *100
Total strokes/swings (n)	5349	5119	4123	996	19%
Total strokes/swings (n) – Contact made (n)	5175	5016	4123	893	18%
Total strokes/swings - No contact made (n)	174	103	14	89	86%
Forehand swings (n)	2220	2002	1755	247	11%
Forehand swings - Contact made (n)	2155	1914	1751	163	9%
Forehand swings - No contact made (n)	65	4	4	0	0%
Backhand swings (n)	1994	1485	1358	127	6%
Backhand swings - Contact made (n)	1889	1417	1351	66	5%
Backhand swings - No contact made (n)	105	7	7	0	0%
Serves swings (n)	1015	1027	1005	22	2%
Serves swings - Contact made (n)	1015	1020	1005	15	1%
Serve swings - No contact made (n)	0	0	0	0	0%
Other swings (n)	120	605	100	505	83%
Other swings - Contact made (n)	119	444	102	342	77%
Other swings - No contact made (n)	92	4	3	1	25%

Table 4. Manually coded forehand stroke classifications that were misclassified/not detected by Catapult algorithms

Stroke Type	True Events (n)	Correctly Identified by Catapult (n) [%total]	Total Errors (n) [%total]	Undetected by Catapult (n) [%total]	Misclassified by Catapult (n) [%total]
Forehand drive (n)	1743	1640 [94%]	102 [6%]	25 [1%]	77 [4%]
Forehand slice (n)	73	16 [22%]	57 [78%]	20 [27%]	37 [51%]
Forehand volley (n)	72	4 [7%]	67 [93%]	42 [58%]	25 [35%]
Forehand end range (n)	223	82 [27%]	141 [63%]	53 [24%]	88 [39%]
Forehand drop shot (n)	2	0 [0%]	2 [100%]	1 [50%]	1 [50%]
Forehand block (n)	5	1 [20%]	4 [80%]	2 [40%]	2 [40%]
Forehand dig (n)	3	0 [0%]	3 [100%]	2 [67%]	1 [33%]
Forehand shadow swing (n)	55	3 [5%]	52 [95%]	23 [42%]	29 [53%]
Smash (n) Coded as "Other stroke"	53	33 [62%]	20 [38%]	4 [1%]	15 [28%]
Smash (n) Coded as "Serve"	53	14 [26%]	39 [74%]	4 [1%]	35 [66%]

Table 5. Manually coded backhand stroke classifications that were misclassified/not detected by Catapult algorithms

Stroke Type	True Events (n)	Correctly Identified by Catapult (n) [%total]	Total Errors (n) [%total]	Undetected by Catapult (n) [%total]	Misclassified by Catapult (n) [%total]
Backhand drive (n)	1438	1235 [86%]	197 [14%]	23 [2%]	174 [12%]
Backhand slice (n)	214	74 [35%]	139 [65%]	34 [16%]	105 [49%]
Backhand volley (n)	65	3 [5%]	62 [95%]	39 [60%]	23 [35%]
Backhand end range (n)	145	31 [19%]	117 [81%]	36 [25%]	81 [56%]
Backhand block (n)	6	0 [0%]	6 [100%]	1 [17%]	5 [83%]
Backhand dig (n)	6	1 [17%]	5 [83%]	3 [50%]	2 [33%]
Backhand shadow swing (n)	91	7 [9%]	83 [91%]	29 [32%]	54 [59%]