

# テニスにおける打球音を用いた球種識別の検討

## Detection of Ball Spin Direction Using Hitting Sound in Tennis

山本修己<sup>1\*</sup> 西田健次<sup>1</sup> 糸山克寿<sup>1</sup> 中臺一博<sup>1,2</sup>

<sup>1</sup> 東京工業大学工学院システム制御系

<sup>1</sup> Department of Systems Controlling and Engineering, School of Engineering,  
Tokyo Institute of Technology

<sup>2</sup> ホンダ・リサーチ・インスティテュート・ジャパン

<sup>2</sup> Honda Research Institute Japan Co., Ltd.

**Abstract:** This paper describes the detection of rotation direction using the hitting sound of tennis balls. Since each ball rotation direction has a slightly different rotation direction and trajectory, there should be a difference in the hitting sound. To distinguish the characteristics of ball rotation direction, a database was constructed that combines the hitting sound recorded experimentally with ball rotation direction. Since it is difficult to distinguish audible differences in hitting sounds by ear, it is necessary to identify them using measuring instruments. For this purpose, after extracting the amplitude spectrum by fast Fourier transform of the shot sound, the entire data was normalized and classified by a support vector machine. As a result of evaluating this method, a high accuracy was obtained in every rotation direction. The proposed method also evaluated the hitting sound from a YouTube video in an unknown environment and showed the effectiveness.

になる。ボールの軌道は回転の影響を大きく受けるため、プレイヤーには相手が打った球種を判別する能力が求められる。

## 1 はじめに

近年、世界的にスポーツに科学技術を取り入れる動きが活発化している。複数のカメラを搭載し、コンピュータビジョン技術を用いて選手やボールの動きを追跡できるスマートコート [1, 2] は、サッカーやバスケットボールなど様々なスポーツで活用されている [3]。テニスの審判判断の精度を高めるために開発されたスマートコートシステム「ホークアイ」は、8台の超高速カメラを搭載し、ボールの軌道や着地点を特定して、プロの試合で審判に情報を伝え、試合の円滑な進行を支援できる [4]。しかし、スマートコートシステムは大規模であり、設置の労力やコストを考えると個人での利用は難しい。また、スポーツではコンピュータビジョン技術を用いたフォーム解析が重要な課題となっており [5-8]、これまでに様々なスポーツを対象に多くの研究が報告されている。

テニスは広いコートを一人数でカバーする必要があるため、ボールの軌道を予測することが重要

\*連絡先：東京工業大学工学院中臺研究室  
東京都目黒区大岡山 2-12-1 西 8 号館 W 棟 W310 号室  
E-mail: yamamoto@ra.sc.e.titech.ac.jp

本稿は ic-sports で採択された "Detection of Ball Spin Direction Using Hitting Sound in Tennis" を和訳したもの（一部改変）です。



図 1: スマートテニスセンサ：ラケットのグリップエンドに取り付けることができ、ストロークの球種、速度、回転数などを測定することができる。

Canal-Bruland らは、被験者にプロテニスの試合をビデオで見てもらい、その時のボールの軌道を予測してもらった実験を行い、その結果ボールの軌道を予測する上で、打球音が重要な要素であること報告している [10]。ボールの軌道を予測する際には、球種がスピン、フラット、スライスいずれであるかを認識することで、ボールの大まかな軌道を予測することができる。なお、スピ

ンは、図 2(a) のようにボールの上面をこするように打つことで、ボールの軌道と同方向の回転をかける。これにより、ボール下面の抗力が低下し、ボールが下に落ちる軌道となる。フラットは、図 2(b) のように、ボールを回転させないように打つ。実際には、軽いスピンのかかることが多い。スライスは、図 2(c) のようにラケットを傾けてボールの下面をスライドさせるように打つ。スピンと反対に、ボールの上向き方向に力が働くのでボールが落ちにくくなる。一方で、ボールの軌道と逆方向の回転がかかるため、スピンやフラットよりも速度が低くなる傾向がある。

実際に球種判別が可能な製品として、図 1 に示すスマートテニスセンサが開発されている [9]。スマートセンサは、加速度センサと 3 軸ジャイロセンサを用いて、ボールの速度、球種、回転数を測定することができる。プレーヤーは、スマートテニスセンサを用いて様々な球種を打ち方を効率的に学習することができるが、相手の打った球種に関する情報は得ることができない。さらにセンサをグリップエンドに装着する必要があるため、試合で邪魔になったりプレーに影響が出るといった欠点がある。

浅野らはボールにマーカーを取り付け、高速度カメラを用いた 3 軸の回転角度と回転数を計測する手法を報告した [11]。2 台のカメラでカメラパラメータからボール中心の 3 次元位置を求め、ボールの軌道を推定しているが、カメラの設置が必要であり、オクルージョンが発生する場合は計測が困難である。

本稿では、この問題を解決するために打球音から 3 種類の球種を識別する技術の構築を目指す。特に、プレーヤーのパフォーマンス向上に寄与するのは、サーブよりもストロークであることから、ストローク打球の球種識別にフォーカスする。このために、まず、ストローク打球とその球種を対応づけたデータセットを設計・構築した。次に、打球音から球種を識別する手法として SVM を用いた球種識別器を提案する。提案手法は、学習データが少なく、クラス間のデータ数のバランスが悪くても識別学習が可能な点が特長である。最後に提案手法を用いて、構築したデータセット、ならびに YouTube から抽出した打球音に対して識別実験を行った。結果として、チャンスレートを大きく超える精度が得られ、テニスの球種判別に有効であることを示すことができた。

以下、第 2 節では関連研究、第 3 節では構築したデータベース、第 4 節では球種の識別手法提案、第 5 節では提案法を用いた評価実験の結果と考察についてそれぞれ延べ、第 6 節でまとめる。

## 2 関連研究

本節では、本稿で着目する打球音を扱った関連研究について述べる。

視覚障がい者用の卓球であるサウンドテーブルテニスでは打球音が聞こえなければファウルとするルールが存在している。この判定には、審判の聴覚を頼りに行っているため、ヒューマンエラーを避けることができないという問題があった。小薬ら [12] はこの問題の解決を目指し、デジタルオーディオテープレコーダーで打球音を騒音計を介して録音し、ウェーブレット変換解析を行い、周波数領域の成分に着目して打球音の存在判定を行う手法を提案している。

Zhang らは、テニスの打球音の特性調査を行った [13]。この研究では、サーブ打球音をデュース側とアドバンテージ側からそれぞれ 15 サンプルずつ抽出し、時間領域で波形を重ね合わせて特徴を比較している。具体的には、テレビ映像を録画して打球音を抽出し、各音波形の最初のピークを選手ごとに重ね合わせて、最初のピークの平均振幅と最初のピークから最後のピークまでの到達時間から選手ごとの音の特徴を分析した。結果として、球速と打球音に相関があることを報告している。

これらの研究を総括すると、打球音解析の重要性は広く認識されており、テニスに限らず、研究が行われている。一方でその内容は、打球そのものの判定や音量と球速との相関の解析にとどまっており、球種判定は十分に組み込まれていない。

## 3 打球音データベースの構築

本稿では、従来研究で解決されていない課題として、打球音から球種を識別する技術を扱う。打球音から球種を識別する上で、打球音と球種が紐づいたデータセットが必要である。上述したように、筆者らの知る限り、打球音から球種を判別するタスクは扱われていないため、これを行うためのデータセットも存在しない。この問題を解決するため、テニスの球種判定用データベースの設計および構築を行った。

### 3.1 録音実験

球種識別用のテニス打球音データベースを構築するため、スピン、フラット、スライスの打球音を球種ラベル付きで収録した。

収録条件は以下の通りである。

- 日時：2019/12/10 11:00～13:00

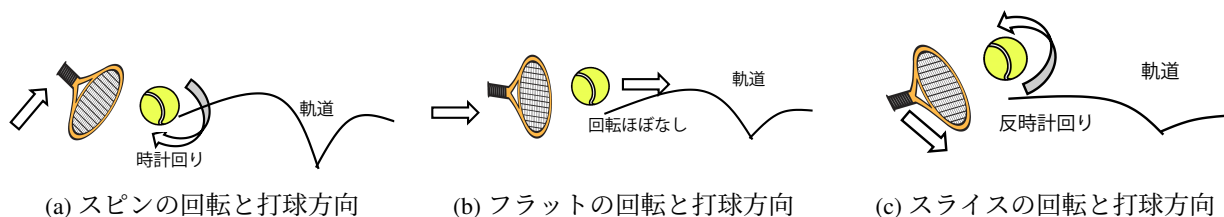


図 2: 3 球種の回転と打球方向

表 1: 録音に使用した設備

機材	型番・設置位置・個数
マイクロホン	TAMAGO-03
マイクロホン位置	ポール近傍（各ポールに1個、計2個）
テニスボール	20 球（新球）
ラケット	SRIXON REVO CV3.0 (SR21802)
PC	16 kHz, 16-bit 録音

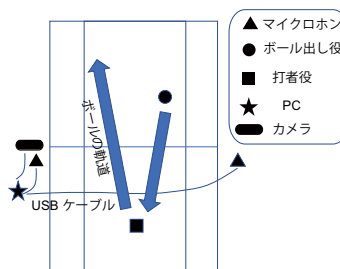


図 3: 実験設定. 矢印は打球の軌道を示す.

- ・ 場所：つくば市二宮公園テニスコート（ハードコート，屋外）
- ・ 天気：晴れ，ほぼ無風.
- ・ 打者：テニス歴 15 年の男性一名

図 3 に収録の状況を示す．打球音データベースは 8ch 音響信号×2 が収録されており，解析にはその中から任意の 1 チャンネルを選択した．また，表 1 に収録に使用した機器の仕様を示す．打球収録は，以下の手順で行った．

1. 球出し役が打者のためにボールを投げる．
2. 打者は，打者が決めた球種と力でボールを打つ．
3. 打者はそれらを記録係に伝える．

計 92 回の試行を行った．

### 3.2 打球音クリップの抽出

各録音について，インパクトの瞬間を含むように 50 ms のクリップを抽出した．この作業は Audacity を使用して，92 個の録音された全ての打球音データに対して手動で行った．このようにして，92 個の打球音とそれに対応する球種ラベルからなる打球パターンデータセットを収集した．92 個の内訳はスピン 46 個，フラット 16 個，スライス 30 個である．

## 4 球種識別手法

この節では打球音から球種を識別する手法を提案する．提案手法は，図 4 に示すように周波数解析，データ正規化，次元削減，2 クラス SVM からなる．以下，各処理について説明する．

### 4.1 周波数解析

入力音は前節で述べたように 50ms で切り出しているため，この切り出した信号に対してフーリエ変換を行う．フーリエ変換は，複雑な音を分解する周波数解析手法である．フーリエ変換は，対象となる信号が周期関数であることを求めるため，一般的な非周期信号に対してフーリエ変換を行う際に，ハンニングやハミングといった窓関数を適用することが多い．今回収録した打球音信号は，50ms の中にはほぼすべての成分が含まれるため，切り出した信号の両端はほぼ振幅が 0 となっている．このため，本稿では窓関数には方形窓を用いた（窓関数を用いない）．また，高速フーリエ変換 (Fast Fourier Transform, FFT) を用いれば，フーリエ変換を高速に実現できるため，FFT を用いた周波数解析を行った．このためには FFT 長を 2 のべき乗に設定する必要がある．構築した打球音データベースのサンプリング周波数は 16kHz であるため，切り出した信号の信号長 50ms は 800 サンプルに相当する．1,024 に足りない分については 0 埋め (0 パディング) を行った．結果としては 0-8kHz (ナイキスト周波数) の周波数成分から

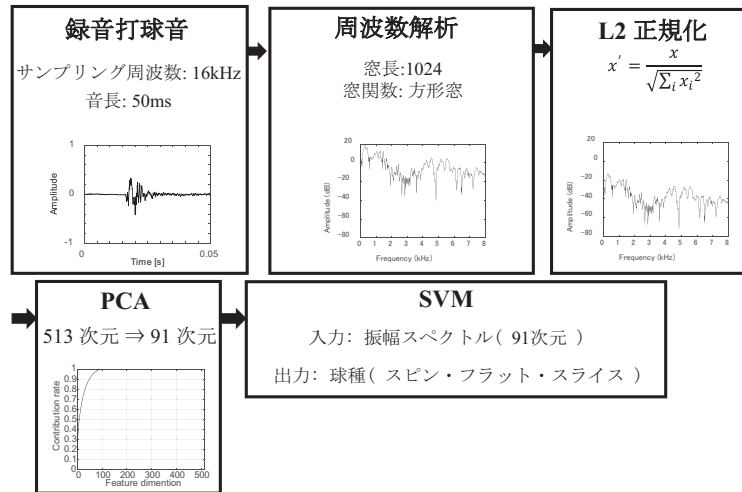


図 4: 提案手法のフローチャート

なる 513 次元の特徴量を各打球音サンプルごとに得た。解析には MATLAB を用いた。

## 4.2 データ正規化

次に、各打球音に対して得られる 513 次元の特徴量  $\mathbf{x} = \{x_i | i = 1 \cdots 513\}$  に対して、インパクト位置の違いによる打球音のばらつきを防ぐために以下の式により L2 ノルム正規化を行い、 $\bar{\mathbf{x}}$  を求めた。

$$\bar{\mathbf{x}} = \left( \sum_{i=1}^{513} x_i^2 \right)^{-\frac{1}{2}} \mathbf{x} \quad (1)$$

## 4.3 主成分分析による次元削減

汎化性能を確保するため主成分分析 (Principal Component Analysis, PCA) [14] を利用した特徴量の低次元化を行った。PCA は互いに直交する主成分を推定する手法で、大きい寄与率を持つ主成分はそれだけよくデータセットを説明することができる。大きい寄与率を持つ主成分を選択することで、低次元でデータセットを表現することができる。得られた 92 個の 513 次元特徴量に対して、主成分分析 (PCA) を適用した。PCA の手順は以下の通りである。 $\mathbf{x}_i (i = 1, \dots, N)$  は  $i$  番目の  $D$  次元データを表す。まず  $\mathbf{x}_i$  から全データ平均を引くことで  $i$  番目のデータの偏差を求める。

$$\tilde{\mathbf{x}}_i = \mathbf{x}_i - \frac{1}{N} \sum_{j=1}^N (\mathbf{x}_j). \quad (2)$$

次に分散共分散行列  $\mathbf{X}$  は以下のように計算できる。

$$\Sigma_{\mathbf{X}} = \frac{1}{N} \sum_{i=1}^N \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i^T. \quad (3)$$

以下の式を用いて、得られた共分散行列  $\mathbf{X}$  に対して固有値分解を行う。

$$\Sigma_{\mathbf{X}} \mathbf{U} = \mathbf{U} \Lambda, (\mathbf{U}^T \mathbf{U} = \mathbf{I}), \quad (4)$$

ここで  $\mathbf{U}$  は  $i$  列目が固有ベクトルである  $D \times D$  の正方行列であり、 $\Lambda$  は対角成分が固有値に対応する対角行列である。このとき、最大固有値の固有ベクトルが最も寄与率の大きい第一主成分に対応する。

図 5 は、513 次元の特徴量に対する累積寄与率を示したものである。第 91 主成分までの累積寄与率が 1 となったため、91 次元に特徴量を低次元化することにした。

## 4.4 SVM

提案手法では、2 クラス分類を行う。例えば、対象がスピンの場合、その音がスピンの打球音であるかスピン以外かを識別する。これにより、スピン、フラット、スライスの 3 種類の 2 クラス識別器を構築する。

PCA により得られた低次元入力ベクトル  $\mathbf{s}$  に対し、識別関数  $y$  は以下の式で定義される。

$$y = \text{sign}(\mathbf{w}^T \mathbf{s} - h), \quad (5)$$

$\mathbf{w}$  は入力に対する重みベクトル、 $h$  は閾値を示す。関数  $\text{sign}(u)$  は符号関数であり、 $u > 0$  のとき、1 を出力し、 $u \leq 0$  のとき、-1 を出力する。言い換えれば、式 (5) は

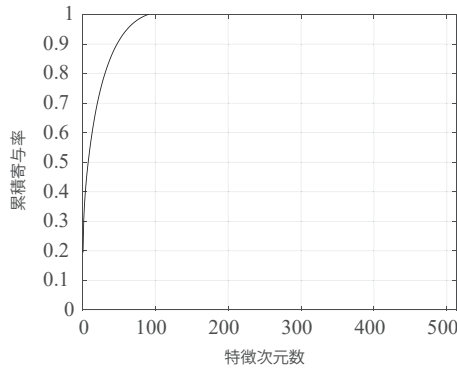


図 5: 構築されたデータベースに対する PCA の累積寄与率. 横軸は特徴次元, 縦軸は選択された特徴次元における累積寄与率.

$\mathbf{w}$  により定義される分離超平面を用いて,  $\mathbf{s}$  で表現される空間を 2 つの部分空間に分離している. SVM [15,16] は, 分離超平面と最も近いサンプルとの距離 (マージン) が最大となる分離超平面を決定する方法である. しかし, 従来の SVM ではすべての入力サンプルは以下の式のように線形分離可能でなければならない.

$$(\mathbf{w}^T \mathbf{s}_i - h) \cdot t_i \geq 1, i = 1, \dots, N, \quad (6)$$

ここで  $t_i$  は,  $i$  番目の入力ベクトル  $\mathbf{s}_i$  に対する正解クラスラベル (1 または  $-1$ ) である.

これは以下の式のようにサンプルは 2 つの超平面により分離されることを示し, 超平面の間にはサンプルは存在しないことを意味している.

$$H1 : \mathbf{w}^T \mathbf{s}_i - h = 1, \quad (7)$$

$$H2 : \mathbf{w}^T \mathbf{s}_i - h = -1, \quad (8)$$

ここで, 各分離超平面間の距離は  $1/\|\mathbf{w}\|$  で定義される.

線形分離可能の制約を緩和するために, H1 と H2 の間に訓練サンプルが入り込むことを許すソフトマージンを導入する. ソフトマージンを導入することで  $\mathbf{s}_i$  に対する距離パラメータ  $\xi_i$  が導入される. このパラメータは  $t_i = 1$  における  $i$  番目のサンプルを以下のように示す.

$$\xi_i = \begin{cases} -\mathbf{w}^T \mathbf{s}_i + h + 1 & (\mathbf{w}^T \mathbf{s}_i - h < 1) \\ 0 & (\text{otherwise}) \end{cases} \quad (9)$$

$t_i = -1$  における  $i$  番目のサンプルは以下のように示す.

$$\xi_i = \begin{cases} \mathbf{w}^T \mathbf{s}_i - h + 1 & (\mathbf{w}^T \mathbf{s}_i - h > -1) \\ 0 & (\text{otherwise}) \end{cases} \quad (10)$$

ソフトマージン SVM は  $\xi = \{\xi_i | i = 1, \dots, N\}$ , が, 式 (11) を満たすという条件のもと, 式 (12) を最小化する

表 2: データ数とクラス重み

識別器	スピンド	フラット	スライス
正解	46	16	30
不正解	46	76	62
$q_i$	1	4.75	2.07

最適化問題として, 定義される.

$$\xi_i \geq 0, t_i \cdot (\mathbf{w}^T \mathbf{s}_i - h) \geq 1 - \xi_i, (i = 1, \dots, N), \quad (11)$$

$$L(\mathbf{w}, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N q_i \xi_i \quad (12)$$

ここで  $C$  は  $\xi$  に対するコストパラメータである. また  $q_i$  は以下の式で定義される  $i$  番目サンプルの重みである.

$$q_i = \begin{cases} 1 & \mathbf{s}_i \in C_l \\ |x \in C_l| / |x \in C_s| & \mathbf{s}_i \in C_s \end{cases} \quad (13)$$

$C_s$  と  $C_l$  はそれぞれデータ数の多いクラスと少ないクラスを示す.

## 5 評価

提案手法は, 構築された打球音データベースと YouTube 動画から抽出された打球音を用いて検証する.

### 5.1 打球音データベースの識別

構築したデータセットを提案手法で評価した. データセットのサンプル数は 92 と少ないため, LOOCV で評価を行った. LOOCV とはサンプルを 2 つに分割し, サンプル中の 1 つのデータで構成される検証データとサンプル中の残りのデータで構成される訓練データに分けて検討する手法である. この手法には  $N-1$  個のサンプルで観測を行うため, 少ないデータでも過適合を防ぐことができるという利点をもつ. ソフトマージンなどのハイパーパラメータは精度が最大化されるように最適化した.

表 3 は 91 次元の特徴量を用いた 2 クラス識別の混同行列を示している. 各表は, 識別の真陽性, 真陰性, 偽陽性, 偽陰性を示している. この表を見ると, 3 球種とも 65% 以上の精度が得られているが, 精度にはそれぞれ特徴がある. フラットの適合率や再現率はスピンドとスライスより低くなっている. これはフラットのデータ数が他の 2 種類よりも, 少ないこと, フラットはスピンドとスライスの中間的な回転数を持っているため判別が難しいことがその理由であると考えられる.

表 3: 打球音データベースを用いて識別した際の混同行列. 92 サンプルは全て 91 次元特徴量で識別.

(a) スピン識別に対する混同行列

	スピン (正解)	その他 (正解)
スピン (予測)	28	13
その他 (予測)	18	33

(b) フラット識別に対する混同行列

-	フラット (正解)	その他 (正解)
フラット (予測)	8	6
その他 (予測)	8	70

(c) スライス識別に対する混同行列

	スライス (正解)	その他 (正解)
スライス (予測)	24	11
その他 (予測)	6	51

表 4: YouTube 打球音識別に対する混同行列. 91 次元の特徴量を使用.

(a) スピン識別に対する混同行列

	スピン (正解)	その他 (正解)
スピン (予測)	2	5
その他 (予測)	8	15

(b) フラット識別に対する混同行列

-	フラット (正解)	その他 (正解)
フラット (予測)	2	1
その他 (予測)	8	19

(c) スライス識別に対する混同行列

	スライス (正解)	その他 (正解)
スライス (予測)	7	9
その他 (予測)	3	11

## 5.2 YouTube クリップの解析

YouTube から現世界 4 位のロジャー・フェデラー選手が 2020 年 1 月の全豪オープン (ハードコート) で練習している動画\*を選定し, その動画の打球音から球種の識別を試みた. 抽出した打球音はスピン, フラット, スライス各 10 本ずつの計 30 本である. 構築したデータベースと同様の処理を行うため, YouTube のサンプリングレートである 44.1 kHz を 16 kHz にリサンプリングし, その後インパクト音を含む 50 ms の打球音を手動で抽出した. 今回は, この 30 の打球音に対して先程の 92 個のデータを学習したモデルを使って予測を行い, その評価を行った. 表 4(a)–(c) にその結果を示す. 精度はどの球種も 60%程度であった. 再現率では, スライスが 70%と最も高い値を示した. これは, スライスは他球種と回転方向が反対であることから他球種とは識別しやすいためと考えられる. しかし, スライスも適合率は低く他の球種も適合率や再現率は低いため, 高い評価指標を得ることができていないのが現状である. この問題は今後の課題としたい.

\*<https://youtu.be/hTn42aJThk8>

## 6 むすび

本稿は, テニスの打球音から球種識別を行う手法を主成分分析とサポートベクトルマシンを用いた手法として提案した. 構築した打球音データベース, YouTube 動画から抽出したプロテニス選手の打球音に対して提案手法を適用し, その有効性を示した. 用いたデータ数が限られているため, 今後は大量のデータを用いて手法の一般性や頑健性を確認する必要がある. 実応用に向けては入力音に混入する雑音への対策も課題である.

## 謝辞

本研究は JSPS 科研費 19K12017, 19KK0260 および 20H00475 の助成を受けた.

## 参考文献

- [1] SecondSpectrum, The next way of seeing sports, <https://www.secondspectrum.com/index.html> (2020)
- [2] Playsight: Smartcourt, <https://www.playsight.com> (2020)
- [3] Seo, S.-W., Kim, M., and Kim, Y.: Optical and acoustic sensor-based 3D ball motion estimation for ball sport simulators, *Proceedings of the 2017 International Conference on Information and Communication Technology Convergence*, Vol. 18, No. 1323. (2018)
- [4] Baodon, Y.: Hawkeye technology using tennis match, *Computer Modelling & New Technologies*, Vol. 18, No. 12, pp. 400–402. (2014)
- [5] Cust, E. E., Sweeting, A. J., Ball, K., and Robertson, S.: Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance, *Journal of Sports Sciences*, Vol. 37, No. 5, pp. 568–600. (2019)
- [6] Appelbaum, L. G., Erickson, G.: Sports vision training: A review of the state-of-the-art in digital training techniques, *International Review of Sport and Exercise Psychology*, Vol. 11, No. 1, pp. 160–189. (2018)
- [7] Okamoto, H., Moro, A., Yamashita, A., and Asama, H.: Toward sports training service with the interactive learning platform, In Sawatani, Y., Spohrer, J. C., Kwan, S. K., and Takenaka, T., editors, *Serviceology for Smart Service System, Selected papers*

of the 3rd International Conference of Serviceology, ICServ 2015, San Jose, CA, USA, 7-9 July 2015, pp. 231–236. Springer. (2015)

- [8] Cao, Z., Hidalgo Martinez, G., Simon, T., Wei, S., Sheikh, Y. A.: OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1. (2019)
- [9] Zepp: Smart tennis sensors, <https://www.secondspectrum.com/index.html> (2020)
- [10] Canal-Bruland, R.: Auditory contributions to visual anticipation in tennis. *Psychology of Sport and Exercise*, Vol. 36, pp. 100–103. (2018)
- [11] Toshiro, Asano., Yuji, Serikawa., Kazuyuki, Ishiguro., Seiji, Itoh.: Analysis of Tennis Ball Rotation and Trajectory by Image Processing, *Journal of the Japan Society for Precision Engineering*; Vol. 82, No. 2, pp. 168–174. (2016)
- [12] Kogusuri, Y., Sato, T., Toyoda, K., Miyato, S.: Development of the holding judgement technology using batted ball sound of sound table tennis, *The Proceeding of the Conference on Information, Intelligence and Precision Equipement : IIP*, No. 8, pp. 49–52. (2008)
- [13] Zhang, D., Yokohama, K., and Yamamoto, Y.: Characteristics of impact sound in tennis service among top-level players, *Nogoya J. Health, Physical, Fitness, Sports*, Vol. 40, No. 1, pp. 37–43. (2017)
- [14] Diamantaras, K. I., Kung, S. Y.: Principal component neural networks: Theory and applications, In Karhunen, J., editor, *Pattern Analysis and Applications*, pp. 74–75, John Wiley & Sons. (1998)
- [15] Scholkopf, B., Burges, C. J. C., and Smola, A. J.: In *Advances in Kernel Methods - Support Vector Learning*. The MIT Press, USA. (1999)
- [16] Vapnik, V. N.: In *Statistical Learning Theory*. John Wiley and Sons. (1998)