

École polytechnique de Louvain

Inertial Detection of Ball-Racket Contacts in Padel Match Play

Towards Non-Visual Match Analysis

Author: **Kacper OLSZEWSKI**
Supervisors: **Cristel PELSSER, Ramin SADRE**
Reader: **Julien GOURGUE**
Academic year 2025–2026
Master [120] in Computer Science

Abstract

In the context of the growing interest in padel, this thesis investigates event detection for padel using a racket-mounted inertial measurement unit (IMU). The long-term motivation is to support future automatic match analysis or scoring systems, but the evaluated scope is limited to detecting ball-racket contacts during amateur match play.

A dedicated racket-mounted IMU recording system was developed to collect continuous padel recordings. The data was synchronized with video and annotated to build a reference dataset of ball-racket contacts, other racket actions, and labels indicating whether contacts occurred during rallies or between points. The IMU method was formulated as a three-stage cascade: detecting short impact segments in the continuous signal, classifying these segments as ball-racket contacts or other racket actions, and classifying detected contacts as occurring during active play or between points.

On the test partition, the first-stage detection retained 99.1% of reference in-play hits. After the full cascade, 89.5% of these hits were recovered as final in-play predictions, with 94.9% precision and 0.122 false in-play predictions per minute. These results indicate that a racket-mounted IMU can provide useful timestamped evidence for in-play ball-racket contacts in continuous amateur padel recordings.

Data and Source Code

The source code used in this thesis is available in the following GitHub repository: <https://github.com/Baboushka/master-thesis-padel>. The repository contains the recording firmware, IMU processing and evaluation code, trained models, result tables, available IMU recordings, and the corresponding manual label tables. The synchronized videos and audio recordings are not included because of privacy constraints and file size. Consequently, the full manual annotation workflow cannot be reproduced from the public repository. However, the released labels allow the reported IMU training and evaluation results to be reproduced from the provided data. The repository is provided to support reproducibility of the evaluated IMU pipeline and future work based on this research direction.

Acknowledgments

I am sincerely grateful to my promoters, Professor Cristel Pelsser and Professor Ramin Sadre, for their guidance, feedback, and time throughout this year. I especially thank them for allowing me to freely explore a subject close to my interests.

I am also grateful to the padel clubs where the recordings were carried out, especially Evere White Star and Tero Padel Club Lime for allowing the experimental setup on their courts.

I would like to thank all the players who participated in the recordings and made the data collection possible. Their time, patience, and willingness to play with the instrumented setup were essential for this project and greatly reduced the stress of working within tight padel-court reservation times.

Deepest thanks to my dear girlfriend Delphine, for her constant support, patience, and help throughout this thesis, including during experiments and proofreading. It would have been much more difficult without you.

I would also like to thank my dear mother, without whom my stomach would have been much emptier, for all the delicious meals she prepared and all the coffee she brewed during this period. I am grateful for her love, care and constant support.

Finally, I would like to thank my friends Emile Van Gysel and Benjamin for proofreading my manuscript and for their thoughtful comments and encouragement, and all my friends and family for their support. This thesis has certainly been a journey, and I could not have made it without all of you. Sincerely, thank you.

Use of AI

Artificial intelligence tools were used as support tools during this thesis, mainly for language editing, grammar correction, phrasing, code debugging, refactoring support, documentation, and plot formatting. All suggestions were reviewed, adapted, and verified by the author before use. AI assistance in the source code is indicated where required, in compliance with the guidelines of the Université Catholique de Louvain [1]. The research questions, methodology, implementation choices, results, analyses, conclusions, and final responsibility for the thesis remain my own.

List of Abbreviations

IMU	Inertial Measurement Unit
ODR	Sensor Output Data Rate
I ² C	Inter-Integrated Circuit
CV	Cross-Validation
CNN	Convolutional Neural Network
SVM	Support Vector Machine
RF	Random Forest

Contents

Abstract	i
Acknowledgments	ii
List of Abbreviations	iii
1 Introduction	1
1.1 Context and motivation	1
1.2 Existing approaches and thesis positioning	2
1.3 Thesis scope, objectives, and contributions	3
1.4 Thesis organization	4
2 Background and Related Work	5
2.1 Padel technology and match-analysis systems	5
2.2 Inertial sensing in racket sports	6
2.2.1 Sensor placement and deployment trade-offs	6
2.2.2 Controlled classification and continuous match detection . .	7
2.2.3 Candidate generation and event localization	7
2.2.4 Model choices, metrics, and generalization	8
2.2.5 Summary	9
2.3 Audio sensing for impact detection	9
2.4 Gap analysis and thesis positioning	10
3 System Overview and Data Collection	11
3.1 Padel court layout and match phases	11
3.2 System concept and thesis scope	12
3.3 Information sources and modality roles	13
3.4 Recording setup and sensor placement	13
3.4.1 IMU placement	14
3.4.2 Airborne microphones	15
3.4.3 Contact microphone	16

3.5	Recording procedure and dataset summary	16
4	IMU Methodology	18
4.1	Role, task definition, and pipeline overview	18
4.2	Hardware and logger architecture	20
4.3	IMU dataset, decoding, and synchronization	21
4.3.1	Recording dataset	21
4.3.2	Decoding and synchronization	22
4.4	Annotation workflow	23
4.4.1	Annotation label definitions and taxonomy	23
4.4.2	Hit reference annotation procedure	25
4.4.3	Detector-candidate annotation procedure	27
4.5	Three-stage hit detection pipeline	28
4.5.1	Candidate event detection	29
4.5.2	Second and third-stage classifications	29
5	Audio Recordings and Qualitative Inspection	31
5.1	Role and scope of the audio subsystems	31
5.2	Recording hardware and recorded sessions	32
5.3	Manual inspection of selected in-play impacts	34
5.4	Waveform and spectrogram examples	36
5.5	Observations	37
6	Evaluation Protocol and Results	39
6.1	Evaluation protocol	39
6.1.1	Data split	40
6.1.2	First-stage detector protocol	40
6.1.3	Second and third-stage classifiers protocol	41
6.1.4	Final evaluation protocol	42
6.2	First-stage detector tuning	43
6.3	Detector candidate dataset	45
6.4	Classification selection	45
6.4.1	Second-stage classification selection	46
6.4.2	Third-stage classification selection	46
6.5	Final evaluation on test partition	47
6.5.1	First-stage detector performance	47
6.5.2	Second and third-stage classifiers performance	48
6.5.3	Cascaded pipeline performance	48
6.6	Error analysis	50
6.7	Summary	51

7	Discussion and Limitations	52
7.1	Interpretation of the IMU pipeline	52
7.1.1	Main result	52
7.1.2	Cascade behavior and remaining errors	52
7.1.3	Test match variation	53
7.2	Limitations of the evaluation	54
7.2.1	Reference labels and annotation	54
7.2.2	Dataset coverage and generalization	54
7.2.3	Conditioned classifier dataset and offline evaluation	55
7.3	Exploratory status of the audio subsystems	55
7.4	Implication for future systems	56
8	Conclusion and Future Work	57
8.1	Conclusion	57
8.2	Future work	58
A	Supplementary material	63
A.1	Hardware details	63
A.1.1	Hardware costs	63
A.1.2	Wiring of the contact microphone	64
A.2	Dataset and annotation details	64
A.2.1	Mapping of players and rackets	64
A.2.2	Hit reference table per session	65
A.3	Evaluation and training details	66
A.3.1	First-stage detector tuning results	66
A.3.2	Development candidate dataset	67
A.3.3	Classifier cross-validation results	67
A.3.4	Development cross-validation confusion matrices	69
A.4	Additional audio inspection examples	69
A.4.1	Neighboring-court contamination	69
A.4.2	Closely spaced rally impacts	69

Chapter 1

Introduction

1.1 Context and motivation

Padel is a racket sport commonly played by four players on an enclosed court with glass walls and metal mesh. Unlike many racket sports, the walls are part of the game, which creates rebounds and several different impacts during play. These include ball-racket impacts, ground bounces, glass rebounds, metal impacts, and net impacts. Among these events, ball-racket contacts are particularly important because they define the player actions and provide temporal structure for rallies.

The sport has grown rapidly in recent years. The International Padel Federation estimated nearly 30 million amateur players worldwide [2]. This growth motivates the development of tools for match analysis, training feedback, and scoring support in amateur clubs. Many existing systems are based on video, which can provide rich information but they usually require camera infrastructure, calibration, storage, and handling of identifiable visual data [3]. Non-visual sensors provide a complementary direction for less visually intrusive match sensing.

A complete automatic scoring and match analysis system would require several layers of processing: detecting relevant physical events, combining them into a consistent timeline, segmenting rallies, attributing events to players and sides, and interpreting the sequence according to padel rules or analysis objectives. This broader goal motivates the study of non-visual sensing as a possible event source. Within this event-sensing layer, this thesis focuses on one necessary component: detecting ball-racket contacts using racket-mounted inertial sensing during amateur padel match play.

Racket-mounted inertial measurement unit (IMU) sensing is relevant because the

racket is directly involved in ball-racket contacts. Compared with court-mounted sensors, it can provide evidence of racket motion and impact dynamic from the player’s perspective. However, detecting relevant ball-racket contacts in continuous match recordings is not trivial. The sensor also captures non-scoring and non-stroke actions, such as racket handling, contacts with the player’s body and surfaces of the court, and between-points actions. An IMU system must therefore do more than detect large motion peaks. It must also distinguish true ball-racket contacts from other racket actions and identify whether detected hits occurred during active play or between-points.

1.2 Existing approaches and thesis positioning

Existing padel technologies can be grouped into several directions. A first category consists of smart-court systems and video-based systems. Commercial systems ¹ such as Playsight, Gamecam, or Playtomic Smart Courts, use court-mounted cameras to record matches, generate highlights, track players, or provide player analytics. Academic work has also explored video-based tracking and analysis for padel. These systems can provide rich information and are mainly designed for infrastructure-based court analysis.

A second category consists of score keeping tools. These systems help players display and manage the score during a match. They usually require user interaction through a phone, watch, or scoreboard. Examples include applications ² such as Padel’em and Scorepadel. These tools support score management, but they serve a different purpose from automatic event detection during play.

A third category are tools centered on player analytics based on smartphone video, racket-mounted sensors, or wearable devices ³. These systems can provide feedback about shots, player performance, and training progress. For example, commercial sensors such as Padelplay advertise automatic shot tracking and shot classification using a motion sensor. Academic work has also studied padel stroke classification from inertial data. These systems demonstrate that racket and wearable sensing are practically relevant for padel. However, public descriptions of commercial systems usually do not disclose details about datasets, annotation protocols, model design, and validation procedures. In addition, much academic work focuses on stroke-type classification, controlled recordings, or pre-segmented recordings.

¹<https://playsight.com/>, <https://www.gamecam.io/>, <https://www.playtomicsmartcourts.com/>

²<https://padelem.com/en>, <https://scorepadel.app/>

³<https://www.padelplay.ai>, <https://www.padelize.ai/>

This thesis focuses on a narrower evaluation gap: it evaluates, in a transparent manner, whether an accessible IMU setup can provide reliable hit evidence during continuous amateur match play. The evaluation explicitly includes events that a practical system must reject or separate from target in-play hits, such as racket actions similar to hits and between-points ball contacts.

1.3 Thesis scope, objectives, and contributions

This thesis evaluates a racket-mounted IMU subsystem for detecting ball-racket contacts in padel match recordings. The broader recording setup also includes airborne microphones and a net-mounted contact microphone as exploratory modalities for future multimodal sensing. In this thesis, these audio subsystems are documented as recording subsystems and used for qualitative observations only. The quantitative evaluation focuses on the IMU subsystem, while a quantitative audio evaluation is left for future work.

The IMU task is formulated as a three-stage event detection pipeline. First, the continuous IMU recording is reduced to short windows around impact-like racket movements. Second, these windows are classified as either a ball-racket contact (HIT) or other racket actions (NO_HIT). Third, detected ball-racket contacts are further separated by play phase as either contacts occurring during active rallies (IN_PLAY) or between points (BETWEEN_POINTS). This structure allows the evaluation to address hit detection, rejection of non-hit racket actions, and separation of in-play and between-points contacts.

The main research objective is to evaluate whether this method can provide reliable in-play hit evidence in continuous amateur match recordings, and accounts for other racket actions and between-points hits. The main contributions of this thesis are:

1. The design and implementation of a relatively low-cost racket-mounted IMU recording system for continuous padel match recordings.
2. An annotated IMU dataset of amateur padel match play, using synchronized video as reference.
3. An assisted annotation workflow for labeling ball-racket contacts, non-hit racket actions, and between-points contacts.
4. A three-stage IMU pipeline for candidate detection, hit/no-hit classification, and in-play/between-points classification.

5. A test evaluation of the cascaded IMU method for in-play hit detection, including final recall, precision, and error analysis.
6. A documented exploratory audio recording setup with airborne microphones and a net-mounted contact microphone.

1.4 Thesis organization

The remainder of this thesis is organized as follows. Chapter 2 reviews related work on padel technology, inertial racket-sport sensing, and audio impact sensing. Chapter 3 presents the recording setup, sensor roles, synchronization procedure, and dataset overview. Chapter 4 describes the IMU subsystem, including the recording setup, dataset, annotation workflow, and the event detection pipeline. Chapter 5 documents the audio recording subsystems and reports the annotation experiment for airborne and contact-microphone signals. Chapter 6 describes the IMU evaluation protocol and results. Chapter 7 discusses the results, limitations, deployment considerations, and implications for future multimodal extensions. Chapter 8 concludes the thesis.

Chapter 2

Background and Related Work

2.1 Padel technology and match-analysis systems

The interest in padel analysis systems is consistent with the rapid international growth of the sport [2]. Existing padel systems related to this thesis can be grouped into three broad categories: smart-court systems, scorekeeping tools, and player-analytics systems.

The first category consists of smart-court and camera-based systems. Commercial systems such as PlaySight, GameCam, and Playtomic Smart Courts advertise functions such as match recording, live streaming, automated highlights, player statistics, instant replay, and AI-assisted match summaries [4], [5], [6]. These systems appear to rely on one or more cameras placed around the court. Some systems also provide player interfaces with statistics such as serve efficiency, heatmaps, shot classification, and ball mapping. However, publicly available product descriptions do not appear to disclose the exact sensing pipeline and model architectures. Since these systems are camera-based, their use may also involve processing identifiable images of players, which raises privacy and data-protection considerations in Europe [3]. This makes them different from lightweight non-visual sensing approaches, which aim to capture specific physical events without relying on video.

The second category consists of scorekeeping and match management tools. These include smartwatches, applications, or digital scoreboards that help players track or display the score during play [7]. Such tools are simple and useful for score management, but they usually require manual interaction from players. They therefore address a different task from automatic event detection.

The third category consists of player-analytics systems based on video, wearable

devices, or racket-mounted sensors. For example, PadelPlay advertises padel shot tracking, shot classification, and performance feedback during play from a racket-mounted inertial sensor [8]. This type of system is directly relevant because it overlaps with part of the sensing objective of this thesis, which is extracting information about padel shots from racket motion. However, public product descriptions usually do not provide enough technical detail about datasets, annotation protocols, model design, validation protocols, or error behavior to assess the method scientifically and build upon their methods. These systems therefore show that there is practical interest in non-visual padel match analysis, but they do not replace a transparent evaluation of an event sensing pipeline.

2.2 Inertial sensing in racket sports

Wearable inertial sensing is an established research direction in sport performance analysis. Several systematic and scoping reviews have described its use for technique analysis, activity classification, and motion classification [9], [10], [11]. In racket sports, inertial measurement units (IMUs) are useful because they measure acceleration and angular velocity, and therefore can capture racket or limb motion around stroke events. Prior studies have used IMU data for stroke detection, event segmentation, and stroke type classification in tennis [12], [13], [14], [15], [16], [17], badminton [18], [19], padel [20], [21], and table-tennis [22]. These studies show that racket-sport strokes create measurable patterns in IMU signals, especially in controlled or semi-controlled recording conditions.

2.2.1 Sensor placement and deployment trade-offs

Sensor placement is an important design choice in sport sensing. It affects both signal quality and practical use during play. Previous studies have placed sensors on the racket [12], [21] or inside the racket, on the wrist [14], [16], [20], [23], on the dominant arm [12], and, more recently, on the passive arm [13]. This suggests that different placements can be viable depending on the application and deployment constraints.

Racket-mounted sensing is relevant for this thesis because a sensor on the racket can capture racket motion and impact dynamics more directly than a body-mounted sensor. In previous work, sensors have often been placed near the center of the racket or near the throat for symmetry [12], [21]. In contrast, this thesis places the IMU slightly off-center near the throat of the padel racket. This placement was chosen as a practical compromise. It aimed to reduce interference with play, such as risk of direct ball impacts, while still capturing useful racket motion during full

match recordings.

2.2.2 Controlled classification and continuous match detection

A key difference between studies is the task being solved. Many racket-sport studies classify stroke type from short windows. In those cases, the stroke is often already isolated or recorded in a controlled setting [12], [20], [21]. In a full match, players do not perform isolated strokes in a fixed order. The sensor stream also contains racket handling, ball retrieval, between-points contacts, and other non-target actions. Park et al. [13] also noted that long tennis sequences can include non-target events such as ball reception and ball pick-up. This supports the view that continuous event detection is more complex than controlled stroke classification, especially in uncontrolled conditions. The task in this thesis is closer to this more challenging setting. The goal is to detect ball-racket contacts in uncontrolled continuous padel match recordings, where the signal contains both target and non-target events.

2.2.3 Candidate generation and event localization

For continuous recordings, it is useful to first reduce the stream to a smaller set of candidate events before classification. Prior work has used window segmentation, thresholding, peak detection, and window extraction before classification [11], [12], [16], [20]. In controlled studies, segmentation is often simpler because strokes are already isolated or occur in constrained conditions. In continuous match recordings, candidate generation becomes a necessary first step before classification. Building on these approaches, a threshold-based detector is used in this thesis as a candidate generation stage rather than as the final event detector. The aim of this first stage is to find ball-racket contacts that occurred in active play with high recall. These candidates are then classified as hit or non-hit. This reduces the amount of data that must be classified while still allowing the second stage to reject non-target events.

Sampling rate and sensor ranges are also important choices. Leddy et al. [11] report that previous sport studies used sampling rates ranging from 20 Hz to 1125 Hz, reflecting differences in movement speed and task requirements. Higher sampling rates can be useful for fast and short-duration movements [11]. They also report large variation in sensor ranges. In this thesis, a high IMU output data rate was selected because racket impacts are brief and dynamic. The accelerometer and gyroscope ranges were set to their maximum available values, ± 32 G and ± 2000 dps, to reduce the risk of saturation during impacts. The final retained

output data rate was 833 Hz, which was the highest setting that worked reliably with the implemented logger. The details are given in Chapter 4.

After candidate generation, previous studies generally use either handcrafted features or raw temporal windows as classifier input. Feature-based methods compute values such as extrema, means, standard deviations, or other statistics from the signal. Other methods use raw windows directly, especially with neural models [13], [16], [18], [20]. This motivates the comparison in this thesis between feature-based and window-based representations.

2.2.4 Model choices, metrics, and generalization

Different models have been used for IMU-based sport classification. Reviews report the use of classical supervised models such as support vector machines (SVM), k-nearest neighbors (KNN), decision trees, random forests (RF), and convolutional neural network (CNN) models [10], [11]. In racket-sport studies, both conventional classifiers and neural models, such as 1D CNN trained directly on raw temporal windows, have been used [12], [13], [14], [16], [18], [20], [24]. Dominguez et al. [20] found that SVM performed similarly to 1D CNNs in a padel stroke classification task, while being suitable for lower-computation settings. These reviews suggest that there is no single best model for all sport-sensing tasks. Performance depends on the dataset size, task, sensor setup, feature representation, and validation protocol. For this reason, this thesis compares several conventional classifiers instead of assuming that a deeper model is necessary. Logistic regression, linear SVM, RBF SVM, and RF are used as lightweight and interpretable baselines. This choice is also consistent with the limited dataset size and the long-term goal of possible embedded use.

Evaluation protocol is also important. More generally, inertial-sensor activity recognition is affected by challenges such as class imbalance, ground-truth annotation, sensor variability [25]. These issues are relevant here because the dataset contains continuous match recordings and imbalanced event classes. Accuracy alone is insufficient when classes are imbalanced or when different error types have different importance. Prior reviews report that recall, precision, F1-score, confusion matrices, and related metrics are commonly used in IMU-based sport classification [10], [11]. This thesis therefore reports several metrics, including positive-class recall, negative-class recall, F1-score, balanced accuracy, and confusion matrices.

The data split is another important issue. Random window-level splits can lead to optimistic performance estimates if similar samples from the same match, participant, or recording session appear in both training and test sets [26]. Despite

this limitation, Kranzinger et al. [10] report that random or unclear splits were still more common than per-subject evaluation in the related racket-sport studies they reviewed. Since this thesis uses continuous match recordings, a match-level split is used to reduce leakage between training and test data from the same recording context. The detailed split design, metrics, and limitations are described in Chapters 6 and 7.

2.2.5 Summary

Prior IMU studies show that racket-sport strokes can be detected or classified from inertial signals. However, much of the previous work focuses on controlled stroke classification rather than continuous padel hit detection in uncontrolled match recordings. This thesis focuses on that gap.

2.3 Audio sensing for impact detection

Audio provides a complementary sensing modality for racket sports. Ball-racket impacts, ground, glass, metal rebounds, and net contacts can all produce short sounds. Audio can therefore help detect events and structure rallies.

Previous work has used audio for racket-sport event detection. Baughman et al. [27] proposed a tennis event-detection pipeline based on acoustic data, including peak detection, feature extraction, atomic event classification, and rally boundary detection. Their system detected events such as hits, foot sounds, applause, and announcements, and used these events to estimate point boundaries. Huang and Cox [28], [29] showed that audio events in sports can also be interpreted using temporal context. This is relevant for padel because a sound event may be informative not only by itself, but also by its position in the rally or point sequence. In padel, Decorte et al. [30] used audio to detect ball-racket contact windows and combined this information with video-based player and ball analysis. Their work shows that audio can provide useful temporal cues for later analysis steps, such as player assignment and positional analysis. This is relevant to this thesis because event detection can be a first step toward broader rally understanding and scoring support.

These studies show that audio is a promising modality for racket-sport event detection, but also that its performance depends on the acoustic setting and on the event type being detected. Unlike a racket-mounted IMU, a microphone captures the full acoustic environment. Prior tennis acoustic work has noted that ball-ground rebound sounds can be weaker than environmental noise, and that sounds from neighboring courts can reduce detection reliability [31]. Additionally, audio-only

detection can suffer from interfering noise and acoustic mismatch [28]. Decorte et al. [30] also reported false positives caused by non-target impact sounds in padel. These included racket contacts with the metal structure, players running into the glass panel, hard stomps on the ground, and racket contacts with the ground after a hit. These issues are especially relevant in padel because the court contains several possible impact surfaces, and because multiple impact events can occur close together in time, such as a racket hit followed by a glass or ground rebound.

For this reason, the audio subsystem is treated as exploratory in this thesis. A complete audio event detector would require a large annotated dataset and a separate quantitative evaluation protocol. This is outside the main scope of the present work, which focuses on racket-mounted inertial hit detection. Instead, the audio recording setup is documented, and selected examples of target events are used for qualitative inspection. A quantitative audio detector is left for future work.

2.4 Gap analysis and thesis positioning

The related work shows that IMU sensing is a promising approach for racket-sport analysis. It also shows that many previous studies focus on controlled stroke classification rather than continuous hit detection during match play. Commercial systems show practical interest in padel analysis, but their public descriptions usually do not provide enough technical detail for scientific evaluation or for building directly on their methods.

This thesis therefore focuses on a narrower problem. It evaluates, in a transparent manner, whether a low-cost racket-mounted IMU can provide evidence of ball-racket contacts during active play in continuous amateur padel recordings. The evaluated pipeline includes candidate event detection, hit/no-hit classification, and play phase classification of predicted hits. This thesis does not aim to reconstruct the full score, build a complete multimodal system, or replace existing commercial tools. Instead, it evaluates one sensing component in a transparent way and documents two exploratory audio subsystems that could support broader match analysis in future work.

Chapter 3

System Overview and Data Collection

This chapter presents the recording setup and data collection procedure used in this thesis. The setup combines racket-mounted inertial sensing, airborne microphones, a net-mounted contact microphone, and synchronized video. The following introduces the relevant court layout and match phases, then clarifies the role of each sensing source, describes the physical sensor placement, summarizes the recording and synchronization procedure, and finally gives an overview of the collected dataset. The detailed IMU pipeline and quantitative evaluation are presented in Chapters 4 and 6. The audio and contact-microphone recordings are documented in Chapter 5.

3.1 Padel court layout and match phases

As stated previously, padel is played on an enclosed court, generally in doubles. A standard court measures 10 m wide and 20 m long, and is divided into two halves by a net. The 3 m high glass panels and metal mesh that enclose the court are part of the playable environment. This creates several types of impact, including ball-racket contacts, ground bounces, glass rebounds, metal impacts, and net contacts. Detailed court specifications and padel rules are provided by the International Padel Federation ¹. A standard court is illustrated later in Figure 3.2.

A point starts with a serve and continues until the rally ends. In this thesis, the term *in-play* refers to the period from the serve to the end of the point. Once a point has ended, the interval between the end of that point and the next serve is defined with the term *between-points*. These terms are used here as annotation labels and are not official rule categories. Longer pauses, setup periods, and unusual

¹https://www.padelfip.com/wp-content/uploads/2025/12/FIP_Rules-of-Padel.pdf

interruptions during or after a set are not considered here as regular between-points periods. This distinction is important because players may still hit, bounce, catch, or pass the ball between points, sometimes producing IMU signals similar to rally strokes, even if they are not part of the active play relevant for scoring. The IMU evaluation therefore separates true ball-racket contacts from other racket actions, as well as in-play hits from between-points contacts.

3.2 System concept and thesis scope

The recording setup was designed with a broader non-visual event sensing concept in mind. In such a system, different sensors observe different physical events during a match and place them on a common timeline. The setup was also meant to be made from accessible components. Their approximate costs are provided in Table A.1. The IMU records racket motion and impact dynamics, the airborne microphones record court impact sounds, the contact microphone records net vibrations, and the video provides a reference for synchronization and annotation. A future fusion layer could investigate how these observations might be combined for event disambiguation, side or player attribution, rally segmentation, or scoring support. Figure 3.1 illustrates this long term concept and the part evaluated in this thesis.

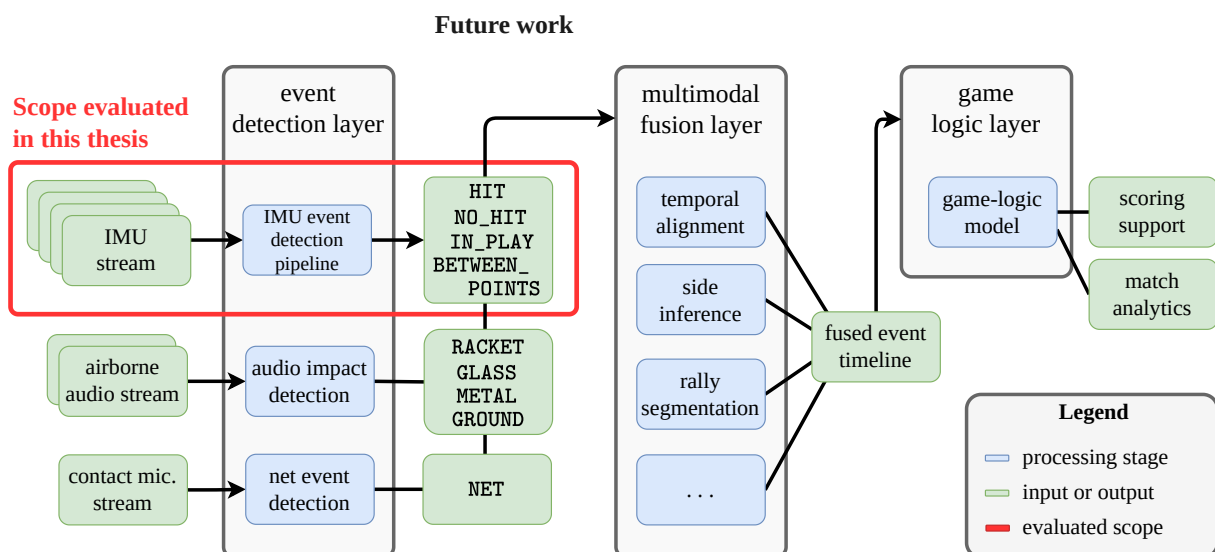


Figure 3.1: Long-term concept for non-visual multimodal sensing. This thesis evaluates the IMU path. Audio, contact-microphone processing, multimodal fusion layer and scoring layer are shown as motivation for future extensions.

In this thesis, only the racket-mounted IMU path is evaluated quantitatively. The labels in the figure are introduced to clarify the system concept. `HIT`, `NO_HIT`, `IN_PLAY`, and `BETWEEN_POINTS` are formally defined in Chapter 4. `RACKET`, `GLASS`, `METAL`, `GROUND` and `NET` refer to physical impact types considered in Chapter 5. The audio and contact-microphone recordings are kept as secondary modalities. They document the broader setup and provide qualitative observations for future multimodal work.

3.3 Information sources and modality roles

The recording setup regroups four sources of information: racket-mounted IMU data, airborne audio, contact microphone audio, and video. Table 3.1 summarizes the role of each source.

Source	Recorded information	Use in this thesis
Racket-mounted IMU	Racket motion and impact dynamics	Main quantitative subsystem for detecting and classifying racket-impact events
Airborne microphone	Acoustic impact sounds in the court	Exploratory modality, documented setup and qualitative inspection
Contact microphone	Vibrations of the net	Exploratory modality, documented setup and qualitative inspection
Video	Visual view of the court	Synchronization and annotation only

Table 3.1: Role of each information source

The sensor streams were recorded continuously during amateur match play. After synchronization, the sensor data and video could be inspected on a common timeline. In this thesis, this common timeline is mainly used for alignment and annotation. In future work, the same principle could be used to investigate multimodal fusion.

3.4 Recording setup and sensor placement

Figure 3.2 shows the placement of the sensors during data collection. The setup was chosen as a compromise between installation time, the signal quality, and minimal

interference during play. The system includes two instrumented rackets, one per team. Therefore, the IMU dataset concerns events recorded on these instrumented rackets, not all rackets on court. It also includes two airborne microphones placed on opposite sides of the court, one contact microphone attached to the net, and one video camera positioned to observe the court. Not all sessions used all modalities simultaneously: the available recordings are summarized in Section 3.5.

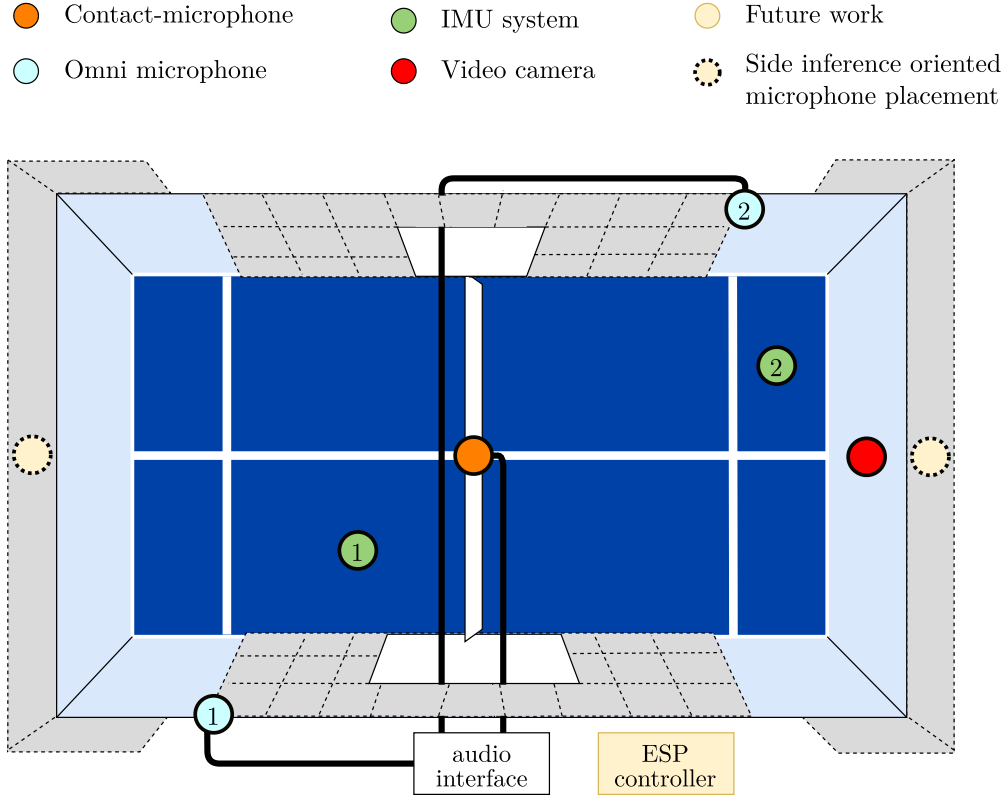


Figure 3.2: Physical setup of each sensor on the padel court

3.4.1 IMU placement

The IMU was mounted directly on the racket as it is directly involved in the target ball-racket contacts. This placement allows the sensor to capture the racket motion and impact dynamics from the player’s side. The sensor was placed near the throat of the racket, in a slightly off-center position. A placement more central would be preferable for symmetry. However, it would have been more exposed to ball impacts and more likely to interfere with player’s grip or play. The chosen

position was therefore a practical compromise between robustness, playability, and signal quality. In practice, players were observed to naturally reverse their racket during play, especially between points. This introduced orientation variability in the recordings. This variability was not controlled formally, but it reflects practical conditions of amateur match recording. Detailed IMU hardware choices and placement motivations are described in Chapter 4. The placement of the subsystem is shown in Figure 3.3.

3.4.2 Airborne microphones

Two airborne omnidirectional microphones were placed on opposite sides of the court. They were positioned near the junction between the glass panels and the metal structure, at approximately 2m height, and oriented toward the inside of the court. This placement was chosen to capture general court impact sounds from both sides, while keeping the installation simple.

In this thesis, the airborne microphones are retained as an exploratory modality. They were therefore not placed for side inference or impact localization, which are left for future work. No direct ball impacts on the microphones were observed during data collection. The detailed audio hardware choices and setup are described in Chapter 5. The placement of one microphone on the court is shown in Figure 3.3.



Figure 3.3: (*left*) IMU subsystem placement (*center*) contact microphone sensor placement (*right*) airborne microphone placement

3.4.3 Contact microphone

A piezoelectric contact microphone was attached to the upper band of the net. This placement was chosen as a simple single sensor position for capturing vibrations of the net structure produced by ball contacts. This sensor is treated as an exploratory modality and is described further in Chapter 5. The placement on the net is shown in Figure 3.3.

3.5 Recording procedure and dataset summary

A GoPro 10 camera was used as a visual reference for synchronization and manual annotation. The camera was positioned to provide a wide view of the court, either near the back glass at around 2 m or near a corner of the court depending on the recording conditions. Figure 3.4 illustrates one possible angle view during recording. It recorded in extra wide mode at 30 Hz, sufficient for manual annotation, while keeping storage manageable. The video was used only as a reference, not as input to the detection pipeline.



Figure 3.4: Example camera view used for manual annotation and synchronization

Each recording session followed the same general procedure. First, the video camera and the active sensors were started. Then, one or several synchronization actions were performed near the start of match play. Depending on the sensor being aligned, these actions included hand claps, racket taps as shown in Figure 3.4, or net taps. Each action was visible in the video and produced a clear peak in the corresponding sensor signal. Synchronization was performed manually by aligning

these reference events with the corresponding signal peak.

The audio channels were recorded through the same audio interface and therefore shared the same sample clock. The audio channel structure and recording details are described in Chapter 5. The IMU data were recorded independently on local microSD cards and synchronized with the video during post-processing. The exact IMU synchronization protocol is explained in Section 4.3. The alignment accuracy was limited by the video frame rate, the sensor sampling rate, and the manual selection of the synchronization peak. At 30 Hz video, one frame corresponds to 33.3 ms. This was sufficient for the used annotation workflow because the video was used to identify the visible event context, and event association was refined using the corresponding sensor signal or candidate window.

Table 3.2 summarizes the number of matches and total recorded duration per modality. Depending on the objective of each session, not all modalities were active in every recording. The detailed IMU dataset is described in Table 4.1, and the audio and contact-microphone recordings are described in Table 5.2. The recordings of a dozen matches used for hardware development, sensor placement tests, logger debugging, and early trials are not included in the counts and datasets reported above.

Sessions	Matches	Total duration	Modality
I01-I20	7	20:18:51	IMU
A01-A06	6	09:06:42	airborne
A01-A06	6	09:06:42	contact

Table 3.2: Summary of the recorded datasets used in the thesis. The airborne audio and contact microphone modalities have the same total duration because they were recorded together through the same audio interface.

The next chapters describe the individual subsystems. Chapter 4 presents the IMU hardware, dataset, annotation workflow, and event detection pipeline. Chapter 5 documents both audio subsystems and reports the observations from selected inspected examples. Chapter 6 describes the evaluation protocol and reports the IMU pipeline results.

Chapter 4

IMU Methodology

This chapter presents the racket-mounted IMU subsystem evaluated in this thesis. The subsystem is designed to detect and classify racket events in continuous padel match recordings. Its objective is to provide timestamped hit evidence, as described in Chapter 3. The chapter first defines the IMU task and gives an overview of the three-stage pipeline, then presents the hardware and logger architecture, the recorded IMU dataset, the decoding and synchronization procedure, the annotation workflow, and the event detection pipeline. The evaluation protocol and results are reported in Chapter 6.

4.1 Role, task definition, and pipeline overview

The role of the IMU subsystem is to provide timestamped evidence of ball-racket contacts from the instrumented racket. Because the sensor is mounted on the racket, it captures both racket motion and impact dynamics. The subsystem is evaluated as a low-level detection component rather than as a complete scoring system. Its output could later be combined with other information sources for multimodal sensing work.

The IMU task is defined as a three-stage pipeline that takes a continuous 6-axis inertial stream as input. The pipeline flow is shown in Figure 4.1. In the first stage, the raw signal is filtered to a set of candidate windows around impact peaks by a detector. Each candidate window is then passed through a hierarchical classification cascade. In the second stage, candidates are classified as either `HIT` or `NO_HIT`. In the reference labels, `HIT` corresponds to a visible ball-racket contact and `NO_HIT` corresponds to any other detected racket action. In the third stage, candidates predicted as `HIT` are further assigned a play phase label, either `IN_PLAY` or `BETWEEN_POINTS`, as defined previously in Section 3.1. The complete annotation

rules are described in Section 4.4.1.

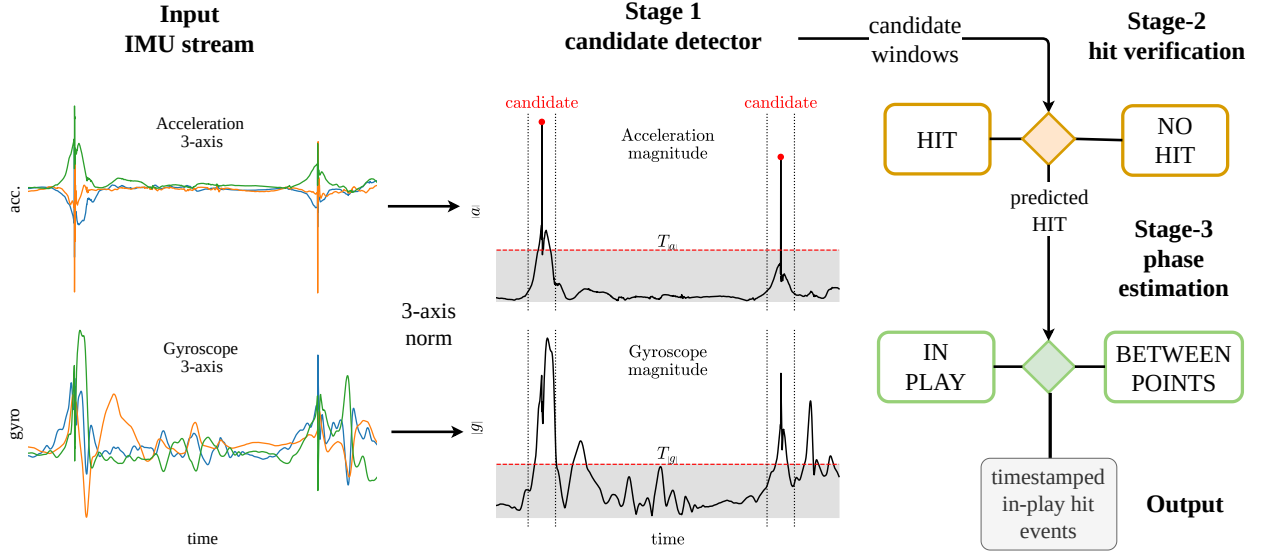


Figure 4.1: Three-stage IMU pipeline

This task definition separates physical contact detection from play-phase estimation, which is useful because future match analysis may depend on the phase of a contact. Indeed, this separation is important because some between-points actions, such as ball passes, racket catches, or controlled racket bounces can produce IMU signals similar to in-play strokes. These events are still ball-racket contacts but they are not part of active play. Labeling them as `NO_HIT` would confuse two different questions: whether a ball-racket contact occurred, and whether it occurred during active play.

The second stage focuses therefore on separating `HIT` from `NO_HIT`, and the third stage focuses on the phase discrimination task applied only to candidates classified as `HIT`. This avoids merging both problems into a single multiclass decision and keeps the role of each stage interpretable. This design also leaves the output more flexible for future work, where all detected ball-racket contacts may be useful, not only those classified as `IN_PLAY`. The phase classifier should be interpreted as a local context estimate and not as a rally segmentation model. More complete play-phase estimation would likely require temporal logic, additional sensing, or game-state logic.

As reviewed in Chapter 2, much prior IMU work in racket sports mainly focused on stroke classification in controlled settings. In contrast, the focus here is on hit verification and play-phase estimation in continuous match recordings.

4.2 Hardware and logger architecture

The IMU module used in this thesis was an Adafruit breakout board based on the STMicroelectronics LSM6DSO32 sensor, containing a three-axis accelerometer and a three-axis gyroscope. The accelerometer and gyroscope ranges were configured to their maximum measurement ranges of $\pm 32g$ and $\pm 2000dps$, respectively. Both output data rates were configured to 833 Hz, which was the highest setting that remained reliable over I²C interface in the present implementation. The rest of the logger was composed of a Seeed Studio XIAO ESP32S3 microcontroller, a 620 mAh Lipo battery, and a Seeed Studio Sense microSD extension board. The components were assembled and soldered into a compact logger system.

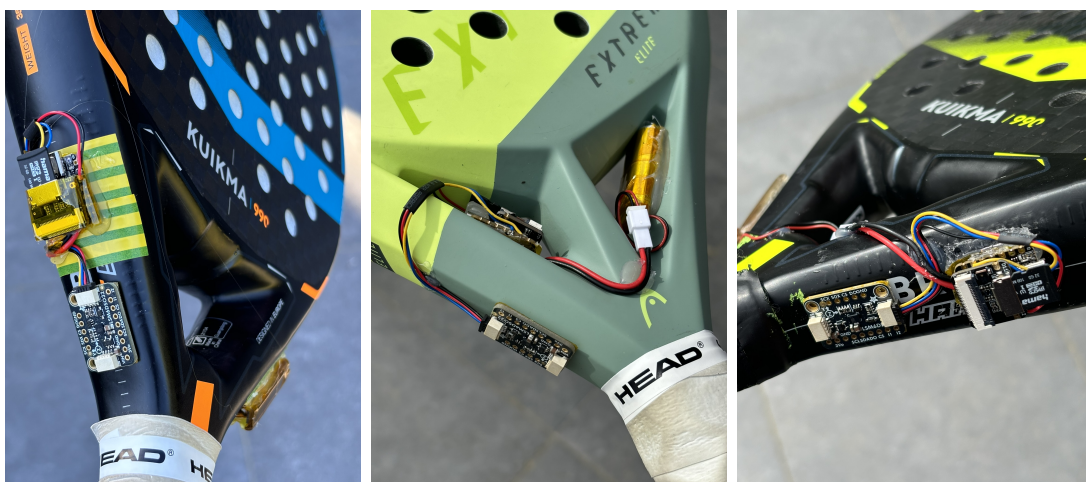


Figure 4.2: Racket-mounted IMU setup without the protection layer

As introduced in Section 3.4.1, the sensor was mounted on the side of the racket, approximately 1.5 cm above the throat junction, as illustrated in Figure 4.2. Preliminary recordings with a custom wrist-mounted sensor were tried but were not formally evaluated. Only the racket-mounted configuration is retained in this thesis.

The implemented logger was designed to record continuously for the duration of an entire match without interruption while minimizing the risk of data loss. To achieve this, the architecture used three buffering stages: the IMU FIFO, a PSRAM ring buffer, and microSD storage. During recording, the LSM6DSO32 produced raw FIFO words containing accelerometer, gyroscope, and timestamp data. The ESP32 polled the FIFO fill level, read available data in I²C bursts, and copied the raw FIFO words into PSRAM ring buffer. The ring buffer absorbed latency from the microSD card where data was written in chunks to a preallocated file on the microSD card over SPI. Each recording file contained a text header with session

metadata, a binary payload with the raw FIFO words, and a text footer with error counters and session status. Figure 4.3 summarizes the logger architecture.

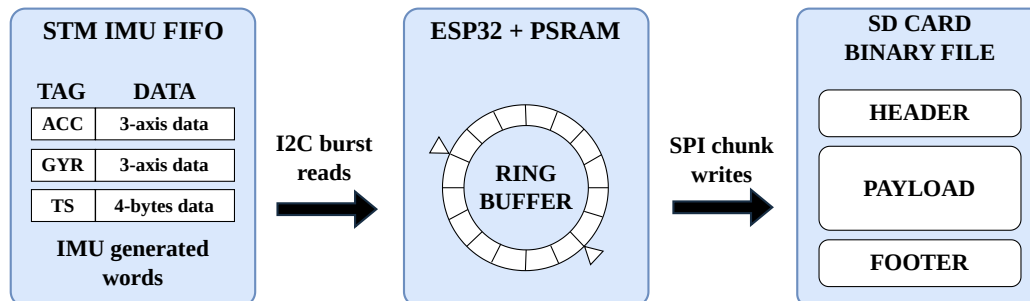


Figure 4.3: Three-stage buffering architecture of the IMU logger

In the retained recordings, including sessions up to 125 minutes, no FIFO overruns, ring buffer overflow, or SD write failures were reported. One recording did not terminate correctly because of battery depletion. Overall, the logger architecture was practically suitable for continuous match-long recordings used in this thesis.

4.3 IMU dataset, decoding, and synchronization

This section describes the IMU recordings and the preprocessing steps used to align them with the reference videos. Each raw logger file was decoded into a sample-level IMU stream, then timestamped, and synchronized with the corresponding video segment. The annotation workflow based on these streams is described in Section 4.4.

4.3.1 Recording dataset

The IMU dataset was collected during real padel play using instrumented rackets. A session was defined as one continuous recording obtained from one instrumented racket. Depending on the recording, a session could contain one set, multiple consecutive sets, or a full match. Some recordings also contained short periods before and after play, such as warm-up, set breaks, or small drills during pauses. These non-target periods are handled through the annotation labels described in Section 4.4.1. The dataset contains 20 sessions, corresponding to 38 cumulated sets, eight players, and three instrumented rackets. In each recorded match, two instrumented rackets were used on opposing teams, producing two IMU streams from the same match context. The goal of this collection was to capture realistic

ball-racket impacts and realistic hard-negative events that occur naturally during play. Because the data was collected in real match conditions rather than under controlled conditions, the number of examples is not perfectly balanced across sessions, players, rackets, or event types.

Each session was associated with metadata including session ID, match ID, player ID, racket ID, set identifier, recording condition, duration, and corresponding video recordings used for manual annotation. Table 4.1 summarizes the recorded sessions.

Match ID	Session ID	# Set	Player ID	Racket ID	Duration
1	I01	2	3	1	00:58:15
	I02	1	3	1	00:23:51
	I06	2	2	2	00:58:41
	I07	1	2	2	00:23:49
2	I03	1	1	1	00:39:18
	I04	1	1	1	00:23:00
	I05	1	4	1	00:26:26
	I08	1	2	2	00:39:39
	I09	2	2	2	00:49:04
3	I10	3	2	1	02:05:19
	I12	2	1	2	01:14:01
	I13	1	5	2	00:44:06
4	I11	1	5	1	00:26:45
	I14	1	4	2	00:26:45
5	I15	3	8	3	01:35:24
	I16	3	9	2	01:36:15
6	I17	3	8	3	01:44:46
	I18	3	12	2	01:43:43
7	I19	3	2	1	01:30:52
	I20	3	1	2	01:28:50
total	20 sessions	38 sets	8 players	3 rackets	20:18:51

Table 4.1: Summary of the recorded IMU sessions. Each session corresponds to one continuous recording from one instrumented racket.

4.3.2 Decoding and synchronization

Each recorded session was decoded from the binary logger format into a sample-level IMU stream. The decoded stream contained three-axis accelerometer and gyroscope data. Because timestamp words were stored only at intervals, sample timestamps

were reconstructed by interpolation between timestamp anchors while preserving FIFO order. This produced an estimated timestamp for each IMU sample.

The decoded IMU stream was then synchronized with the corresponding video recording. Because one IMU session could correspond to one or more video files, synchronization was performed per video segment. The first visible reference event, generally a clap or clear impact, was identified in each video and matched to the corresponding sample peak in the IMU signal. This pair was used as a time anchor to estimate the video timestamp of each IMU sample.

After synchronization, each sample was assigned a video identifier, and an estimated video timestamp. Millisecond-level synchronization was not required for the annotation task because most visible events in the video produced distinctive local peaks in the acceleration magnitude. The synchronization protocol was therefore sufficient for event review and manual annotation.

4.4 Annotation workflow

The annotation workflow produced two related tables: a hit reference table and a final detector candidate table. The hit reference table contains visible ball-racket contacts identified from synchronized video. It also includes ambiguous and out-of-scope **EXCLUDE** events. It was used directly to tune and evaluate first-stage detector configurations and later as a source of **HIT** labels when constructing the detector candidate table.

The final detector-candidate table was created after the first-stage detector was selected and frozen. It contains the actual candidates passed to the classifier stages and therefore includes **HIT**, **NO_HIT**, and **EXCLUDE** labels. This separation is important because detector tuning and classifier training require different reference data: detector tuning requires a reference list of visible hits, and classifier training requires labels for the candidates produced by the selected detector. This two-phase workflow is summarized in Figure 4.4 and its use in the full evaluation protocol is described in Chapter 6.

4.4.1 Annotation label definitions and taxonomy

Both annotation tables used three fields: **label**, **sublabel**, and **phase**. The **label** field defines the main event category. A candidate is labeled **HIT** when it corresponds to a visible ball-racket contact. A candidate is labeled **NO_HIT** when it is a reviewed detector candidate that is not a visible ball-racket contact. This

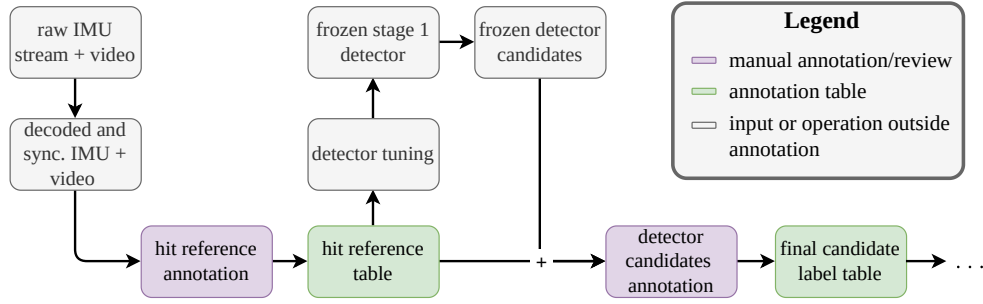


Figure 4.4: Two-phase IMU annotation workflow

class therefore represents hard negative candidates produced by the detector and accepted as valid examples after manual review, and not background signal in the continuous stream. A candidate is labeled **EXCLUDE** when it is outside the target scope or cannot be labeled reliably.

The target periods were rallies and normal between-points periods during sets. Warm-up periods, synchronization actions, drills during pauses, and moments outside the padel-play context were not used as target events. Events from these periods were labeled as **EXCLUDE** in the annotation and candidate tables. The IMU streams were not trimmed manually after recording. Unreliable and out-of-scope events were kept in the reviewed tables with **EXCLUDE** label. For instance, out-of-scope events included racket handling outside the padel court. This preserved the original recording timeline and made the annotation traceable.

The **phase** field describes whether an event occurred during active play or between points. It is used as the target of the third-stage classifier only for candidates labeled **HIT**. Phase labels were kept for **NO_HIT** for error analysis only. Table 4.2 summarizes the main label and phase values used during annotation.

Field	Value	Definition
label	HIT	Visible ball-racket contact in the video.
	NO_HIT	Reviewed detector candidate that is not a visible ball-racket contact.
	EXCLUDE	Synchronization action, ambiguous event, poorly visible event, or event outside the target scope.
phase	IN_PLAY	Event occurring during point play, from serve to point end.
	BETWEEN_POINTS	Event occurring between points (excluding set breaks, warm-up, and moments outside target match context).

Table 4.2: Main annotation fields

The **sublabel** field provides a more detailed description of the event. For **HIT** events, it describes the stroke type or between-points ball-racket action. For **NO_HIT** events, it describes the physical origin of the negative candidate. For **EXCLUDE** events, it describes why the event was removed from the target dataset. Sublabels were used for dataset description and error analysis, but not as classifier targets. The full sublabel taxonomy is given in Table 4.3.

Label	Sublabel	Description
HIT IN_PLAY	forehand	Forehand groundstroke
	backhand	Backhand groundstroke
	forehand_volley	Forehand volley
	backhand_volley	Backhand volley
	lob	Upward stroke intended to lift the ball
	smash	Overhead stroke with high intensity
	serve	Service stroke starting an in-play phase
	other_hit	Other visible in-play ball racket contact not fitting above
HIT BETWEEN _POINTS	ball_retrieval	Ball pickup with the racket when the ball is on the ground
	ball_pass	Any stroke used to intentionally pass the ball to another player or toward the net
	ball_rebound	Controlled ball rebound with racket against ground
	ball_catch	Ball reception or catch on the racket outside active play
NO_HIT	hand_tap	Player tapping the racket with the hand
	racket_glass	Racket contact with glass panel
	racket_metal	Racket contact with metal mesh wall
	racket_ground	Distinct racket contact with the ground
	racket_body	Racket contact with the body of the player
	teammate_check	Racket contact with racket of teammate
	air_swing	Swing motion with no ball contact
	handling	Grip adjustment, and other racket manipulations
EXCLUDE	sync	Sync. action used for video and IMU alignment
	ambiguous	Event for which video did not allow a label decision
	outside_play	Event outside the target scope (warm-up, pauses, drills, ...)

Table 4.3: Sublabel taxonomy used during annotation

4.4.2 Hit reference annotation procedure

The purpose of the hit reference annotation was to identify visible ball-racket contacts in the synchronized video. Because the IMU recordings were continuous and lasted long, a permissive annotation detector was used to help this annotation process. This annotation detector was only an annotation helper and was separate from the final first-stage detector evaluated later in Chapter 6. It used the acceleration magnitude $|a|$, defined in (4.1), with a low threshold of $T_{|a|} = 3\text{ g}$ and a grouping gap of $R = 0.05\text{ s}$. Each proposed candidate timestamp was placed at

the local maximum of $|a|$.

A custom annotation tool was implemented to display the synchronized video, the $|a|$ signal, and the candidate peaks proposed by the annotation detector. The proposed peaks were used as markers to help the annotator locate likely impact events in the signal. They did not define the reference labels by themselves. Figure 4.5 shows the annotation workflow.

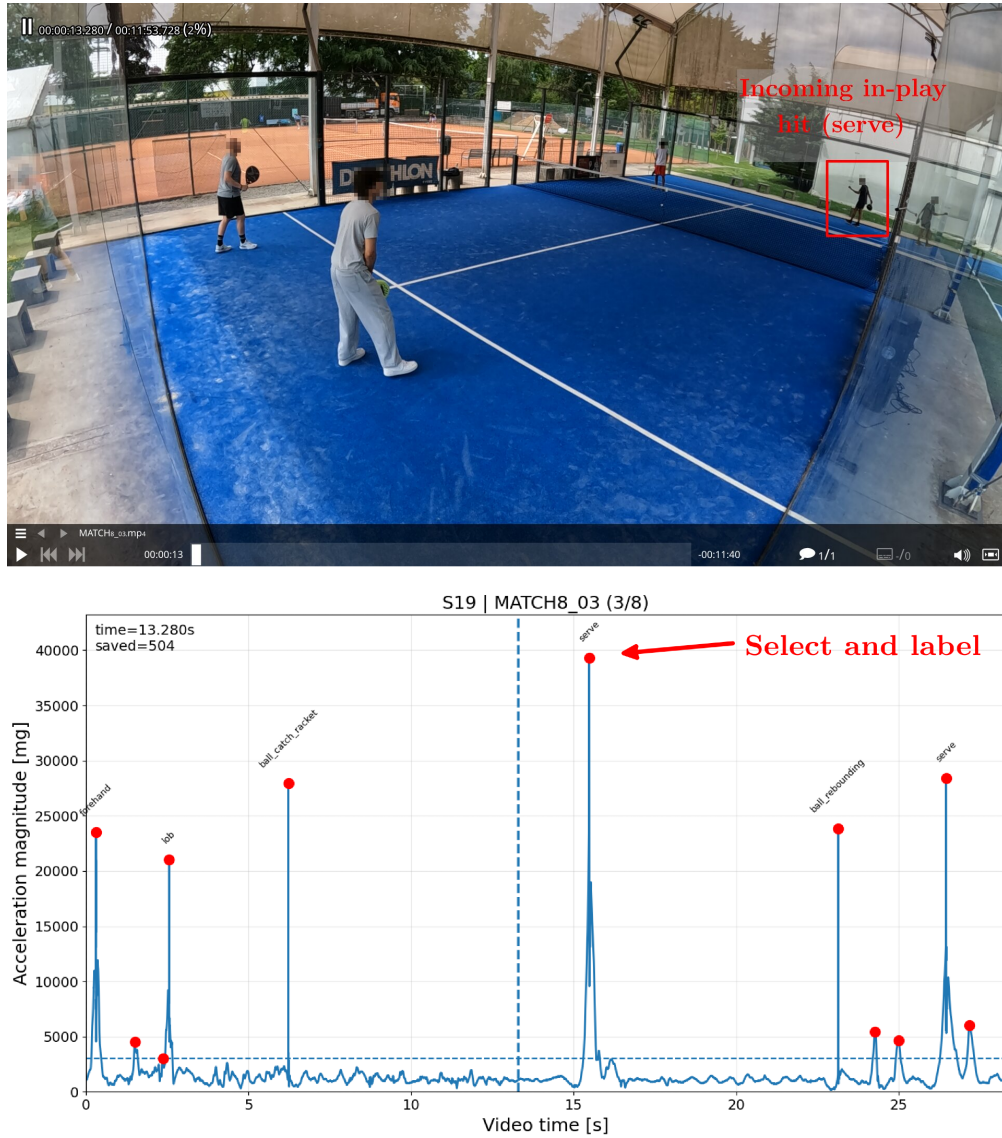


Figure 4.5: Hit reference annotation tool: synchronized video and plotted $|a|$ signal

Each video was reviewed from beginning to end. When a visible ball-racket contact was observed, the annotator selected and labeled the corresponding peak in the plotted IMU signal. In most cases, this peak corresponded to one of the candidates proposed by the annotation detector. When a visible hit did not have a corresponding proposed candidate, a manual event was added by selecting the nearest relevant local acceleration maximum. In total, 51 hit references were manually added, corresponding to 1% of all hit reference events.

Ambiguous events were inspected frame by frame. If the video did not allow a reliable decision, for example because the event was poorly visible, the event was labeled **EXCLUDE** rather than forced as **HIT**. Between-points **ball_retrieval** events were sometimes difficult because the ball-racket contact could occur close to a racket-ground contact. When the ball-racket contact produced a visible peak, that peak was selected as the timestamp of the hit. When both contacts were visually merged, the dominant merged peak was selected and the case was treated as a limitation of the annotation protocol.

This workflow produced one hit reference table per session. Each row contains the session identifier, hit identifier, IMU frame index, IMU timestamp, video identifier, estimated video timestamp, main label, sublabel, phase label, and the source of annotation. The source field indicates whether the event was added manually or came from the annotation detector candidate. The resulting hit reference table was used to evaluate whether each first-stage detector configuration recovered visible hits within a given time tolerance. Table 4.4 summarizes the hit reference dataset by match. The full distribution per session can be found in Appendix A.2.2. It contains **HIT** events and selected **EXCLUDE** events. It does not contain **NO_HIT** labels because non-hit candidates from the permissive annotation detector were not the target of this annotation step. In the evaluation protocol in Chapter 6, this annotation process is illustrated as step 1 and 8 in Figure 6.1.

4.4.3 Detector-candidate annotation procedure

After the first-stage detector configuration was selected, it is applied to the recordings to produce the candidate events available to the classifier stages. These candidates are then labeled by combining matching against the hit reference table with manual video review. These candidates represent the events available to the classifier stages of the evaluated pipeline.

Each detector candidate is first matched against the hit reference table. A candidate is labeled **HIT** if its sample index matched a hit reference event within two IMU samples. This tolerance allowed small differences in peak location between

Match ID	HIT			EXCLUDE	
	in-play	between-points	total	ambiguous	outside play
1	261	247	508	5	124
2	428	391	819	15	15
3	600	420	1020	8	198
4	101	84	185	2	3
5	439	263	702	13	202
6	597	409	1006	11	17
7	439	375	814	11	83
total	2865	2189	5054	65	642

Table 4.4: Summary of hit reference dataset by match and session. Total hits are the sum of in-play and between-points hits (excluded events are not included).

the hit reference annotation and the frozen detector output. Matched candidates inherited the label, sublabel, and phase from the corresponding hit reference row. **EXCLUDE** labels from the reference table were also transferred. Unmatched candidates are manually reviewed using the video, and labeled **NO_HIT** or **EXCLUDE**, as previously defined. Candidates labeled **EXCLUDE** are removed before classifier training and evaluation.

The resulting detector-candidate dataset is a candidate-level reference dataset, not a sample-level ground truth for the full IMU stream. Unlabeled time intervals should not be interpreted as **NO_HIT**, because they could also correspond to unreviewed non-target periods. The **NO_HIT** class only contains reviewed candidates produced by the frozen detector. This definition matches the evaluated pipeline because the classifiers operate only on detector candidates, not on every background sample. The concrete application of this procedure to the development and test partition is described in Chapter 6, after detector selection protocol is introduced. This annotation process is illustrated as step 5 and 10 in Figure 6.1.

4.5 Three-stage hit detection pipeline

This section describes the final IMU pipeline evaluated in Chapter 6. The pipeline operates on a continuous stream and produces a timestamped list of candidate events with predicted labels. The first stage detects candidate events. The second stage classifies each candidate as **HIT** or **NO_HIT**. The third stage is applied only to candidates predicted as **HIT** and classifies them as either **IN_PLAY** or **BETWEEN_POINTS**. The main target output is a set of predicted in-play ball-racket contacts.

4.5.1 Candidate event detection

The objective of the first stage was to preserve as many in-play ball-racket contacts as possible while reducing the stream to a smaller set of candidates. Candidate detection used acceleration magnitude and gyroscope magnitude, computed from the three-axis accelerometer and gyroscope signals as

$$|a| = \sqrt{a_x^2 + a_y^2 + a_z^2} \quad (4.1)$$

$$|g| = \sqrt{g_x^2 + g_y^2 + g_z^2} \quad (4.2)$$

where a_x, a_y, a_z , and g_x, g_y, g_z are respectively the raw acceleration and gyroscope values along the three-axis.

Candidate peaks were first proposed from the acceleration magnitude by using a threshold on $|a|$ such that $|a| > T_{|a|}$. Samples above this threshold were grouped into a single candidate event when they were separated by less than a fixed merge gap $R = 0.1$ s. This gap was used as a physical grouping constraint because two independent in-play hits from the same instrumented racket are unlikely to occur within such short interval. It prevents multiple threshold crossings from the same physical event from being counted as separate candidates.

For each group, the candidate timestamp was defined as the local maximum of $|a|$. A gyroscope gate was then applied at this acceleration peak sample using the threshold $T_{|g|}$ such that the candidate was accepted if $|g|_{\text{at } |a| \text{ peak}} > T_{|g|}$. This extra gate was used to reduce non-relevant candidates passed to the classifiers even further. The peak-based gate was selected because it keeps the detector simple and preserves a single timestamp for each candidate. It also makes the detector easier to transfer to a future live implementation. A window-based gyroscope gate was also compared during detector selection, and the acceleration and gyroscope thresholds were selected on the development partition, as described in Chapter 6.

4.5.2 Second and third-stage classifications

In the second stage, each candidate window was classified as either `HIT` or `NO_HIT`. In the third stage, candidates identified as `HIT` were further classified as either `IN_PLAY` or `BETWEEN_POINTS`. This stage provided context evidence from the IMU signal, but not a final scoring decision. The final interpretation of play context is left to a future match analysis layer.

For each detected candidate, a fixed time window was extracted around the $|a|$ peak. Two window sizes were evaluated: 200 ms and 400 ms. The selected window

size for each classifier is reported in Chapter 6. The window was centered on the candidate peak so that it included the close pre-impact motion, the impact peak, and the direct post-impact response. Candidates whose window was interrupted by the recording start or end were discarded.

Two feature representations were compared. The first was a handcrafted representation based on axis-independent features to account for the asymmetric placement of the sensor. It included six features: peak acceleration magnitude $\max |a|$, peak gyroscope magnitude $\max |g|$, peak acceleration jerk magnitude $\max |\Delta|a||$, where jerk represents the discrete derivative of $|a|$, $\text{RMS}_{|a|}$, $\text{RMS}_{|g|}$, and the duration for which $|a|$ stayed above the first-stage threshold. The second representation was a temporal representation. It used the fixed IMU window directly by flattening the six IMU axis into a feature vector. This preserved the local waveform around the candidate peak.

The evaluated classifiers were logistic regression, linear SVM, RBF SVM, and random forest. These models were chosen to compare lightweight classifiers and nonlinear ones, while keeping the long-term goal of embedded ESP32 deployment in mind. More complex temporal models were not considered because the dataset was limited and the objective was to build a robust IMU pipeline rather than to optimize the architecture of a larger model. The model selection and hyperparameters tuning are described in Chapter 6.

Chapter 5

Audio Recordings and Qualitative Inspection

The audio subsystems were included in the original multimodal sensing concept to provide information that is not available from the racket-mounted IMU. In the scope of this study, this chapter documents the audio recording setup and presents a short qualitative inspection of selected annotated examples.

5.1 Role and scope of the audio subsystems

The IMU subsystem observes racket motion and dynamics from the instrumented racket. The audio subsystems were intended to provide additional acoustic information. Airborne microphones can capture acoustic events occurring on the court such as ball contacts with glass panels, the ground, metal mesh or rackets. A contact microphone mounted on the net was included to capture vibration evidence for net contacts, which were expected to be weak and difficult to localize from airborne microphones alone.

A complete audio subsystem would require sound event detection in the continuous audio stream, including relevant candidate detection, sound classification, rejection of non-target sounds, and handling of between-points activity. This would require a substantially larger annotated audio dataset and a separate evaluation protocol. In this thesis, the audio work is therefore limited to documenting the recording setup and manually inspecting whether visible in-play impact events could be located in the recorded waveforms and spectrograms under favorable conditions. This scope was also motivated by the annotation burden of the current annotation workflow and by preliminary inspection showing that neighboring courts activity could introduce substantial interference in airborne microphone recordings. The

labels used in this experiment are defined as an annotation vocabulary that could be used for future classifier targets. They are summarized in Table 5.1. Events outside the target scope, such as between-points events or ambiguous events, were not included as target annotations.

Modality	Label	Description
Airborne microphone	RACKET	In-play racket-ball impact sound
	GLASS	In-play glass-ball impact sound
	METAL	In-play metal-ball structure impact sound
	GROUND	In-play ball-ground impact sound
Contact microphone	NET	Vibration of the net produced by a ball impact

Table 5.1: Audio subsystem labels

5.2 Recording hardware and recorded sessions

The airborne audio subsystem used two accessible Behringer ECM8000 omnidirectional measurement microphones connected to a Behringer UMC404HD USB audio interface through 20 m XLR cables. One microphone was placed on each side of the court near the junction between the glass panel and the metal structure, as previously shown in Figure 3.3. A custom holder with shock absorbing suspension was built to reduce vibration transfer from the metal mesh to the microphone. The microphone, audio interface and custom holder are shown in Figure 5.1.

The contact microphone subsystem used a low-cost 27 mm piezoelectric disk attached to the upper band of the net and connected to the same audio interface through a mono TS jack cable. It was used to record vibrations caused by ball-net contacts. During early recordings, one direct ball impact damaged the sensor. A small plastic furniture protector was added to it. This reduced the risk of direct impacts, but the protection remained limited. The contact microphone is shown in Figure 5.1. The wiring details are provided in Appendix A.1.

All audio channels were recorded through the same audio interface using Reaper ¹. This ensured that the two airborne and the contact microphones shared the same sample clock. This was intended for future work on side inference, which would benefit from precise relative timing. The gain was adjusted before each recording to reduce clipping while preserving usable signal.

¹<https://www.reaper.fm/>



Figure 5.1: (*left*) ECM8000 microphone with the custom mount and audio interface (*right*) Contact microphone

The audio recordings were collected during real match sessions. Each audio session corresponds to one match recording and contains two airborne microphone channels and one contact microphone channel. Table 5.2 summarizes the recorded audio sessions.

Session ID	Court ID	Setting	Duration	Players ID	# active courts
A01	1	covered	00:48:32	2,3,4,6	1
A02	1	covered	02:03:16	1,2,4,5	1
A03	2	covered	01:48:19	8,9,10,11	0
A04	2	covered	01:39:29	8,12,13,14	0
A05	3	indoor	01:22:57	15,16,17,18	3
A06	4	covered	01:24:09	1,2,4,19	1
total	4 courts	5 covered 1 indoor	09:06:42	18 players	-

Table 5.2: Summary of recorded audio sessions. # active courts corresponds to the number of active neighboring courts.

5.3 Manual inspection of selected in-play impacts

A segment of five minutes from the start of A04 was selected for manual inspection. This segment was chosen because no neighboring court was active, making it a favorable case for checking whether visible impact events could be located and labeled in the audio streams. The inspection focused on in-play impact events during rallies. Between-points impact events resulting from ball passes, rebounds, and handling actions were not annotated in this audio inspection. The more challenging A05 indoor recording, with three neighboring active courts, was also briefly inspected and showed substantial cross-court contamination. Other recordings were inspected informally without annotation. The selected A04 should therefore not be interpreted as evidence that airborne audio annotation would be robust across all court conditions.

The multichannel audio recording was opened in Audacity ² with the two airborne microphone channels and the contact microphone channel aligned on the same audio timeline. The synchronized video was reviewed in parallel and used as the visual reference to identify visible physical events. An example of this annotation process is shown in Figure 5.2.

For airborne audio, when a target event was observed in the video, the corresponding transient was located in the waveform or spectrogram of both microphone signals. Each annotated event included an onset time, an approximate offset time, an event label, and a channel. The onset was defined as the first visible acoustic rise in the microphone signal. This timestamp corresponds to the sound arrival at the microphone and not the exact physical impact time. For the contact microphone, a time interval was labeled NET when a ball-net contact was visible in the video. The onset was placed at the first visible rise in the signal, and offset was placed when the signal returned to the background level.

The annotation was time-consuming because each visible ball impact first had to be identified in the video and then located in the waveform or spectrogram. Annotating five minutes of the segment required approximately 90 minutes. This made full manual annotation of all recorded sessions impractical within the scope of this thesis.

²<https://www.audacityteam.org/>

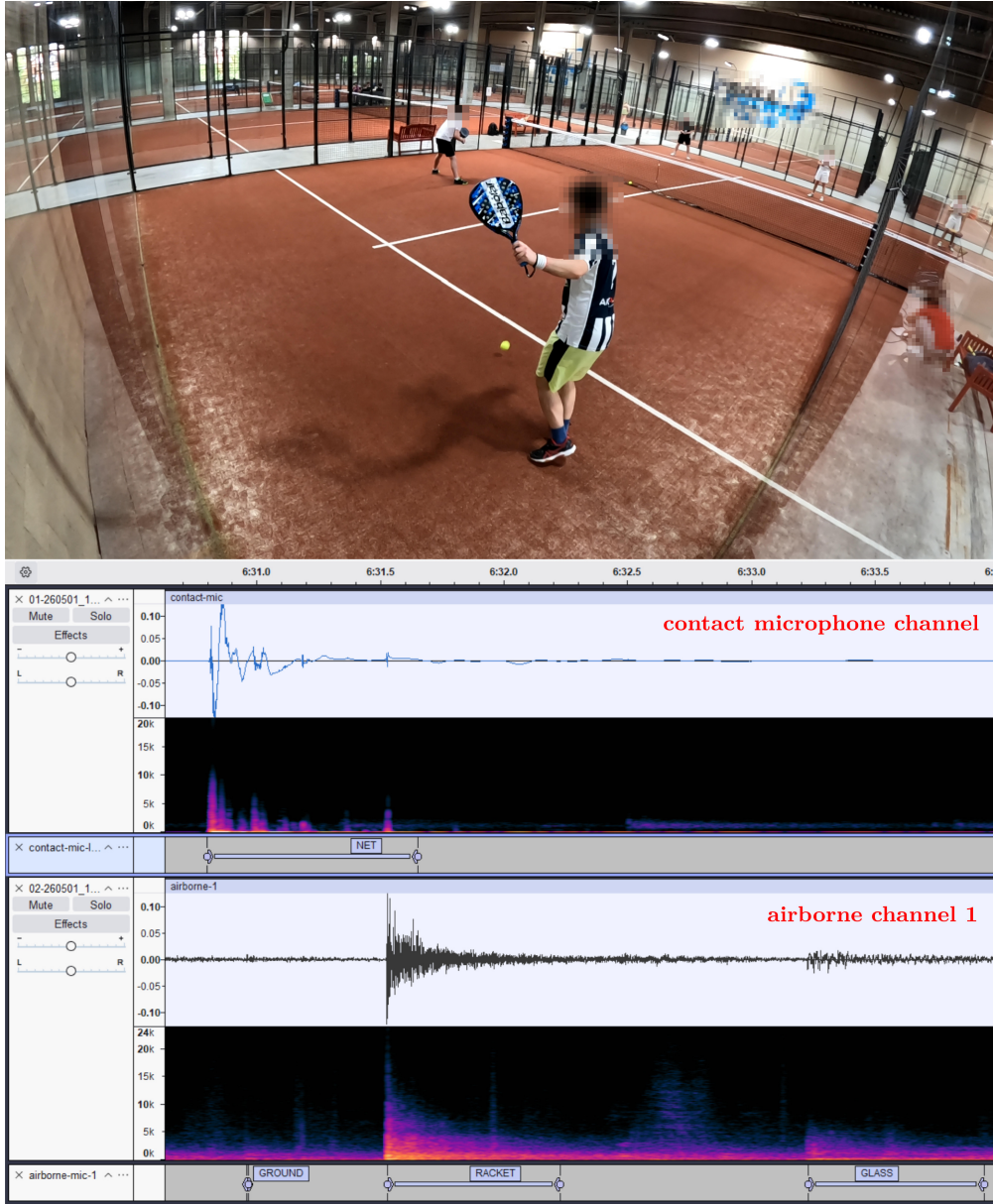


Figure 5.2: Example of the parallel audio-video inspection workflow with Audacity waveforms of each channel and their respective label rows.

Table 5.3 summarizes the visible in-play impacts that were annotated. For airborne events, the counts refer to physical events: the same physical impact was labeled in both channels, but it was counted once in the table. These counts are only descriptive and should not be interpreted as detection performance.

Class	RACKET	GROUND	GLASS	METAL	NET	total
Count	60	114	21	4	8	207

Table 5.3: Counts of visible in-play impacts annotated in the selected segment

5.4 Waveform and spectrogram examples

This section presents selected waveform and spectrogram examples from the annotated segment to illustrate the observed signal patterns and annotation difficulties. Figure 5.3 shows airborne microphone responses for annotated in-play impacts. Racket, glass and metal impacts generally produced visible transients in both airborne microphone channels. Their relative amplitude and clarity differed between channels depending on the position of the event. Some events were harder to annotate reliably. In padel, several impacts can occur within a short time interval, for example a racket hit followed by a glass or ground rebound. In such cases, the acoustic responses could overlap, making the onset and offset placement difficult, as illustrated in Appendix A.4.2. Ground impacts were the least clear airborne events inspected because they produced weaker responses and were often masked by louder impacts, footsteps, or speech.

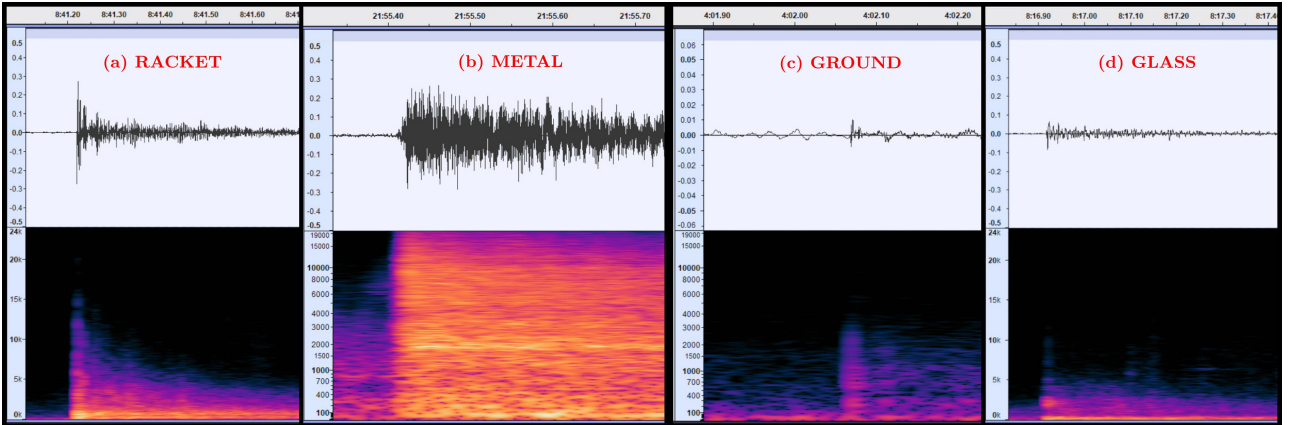


Figure 5.3: Example of isolated airborne microphone waveforms from the annotation test. The metal example was chosen from a later segment because it provide a more isolated response.

Figure 5.4 shows contact-microphone responses for visible ball-net contacts. In the inspected segment, net contacts produced localized vibration responses in the signal. Medium and strong contacts were clearly visible, while weaker contacts could often also be identified. The response depended on the contact location,

impact strength, and mechanical coupling between the piezoelectric disk and the net band.



Figure 5.4: Example contact microphone waveform and spectrogram from the annotation test

5.5 Observations

The audio inspection remained exploratory. Only five minutes were manually annotated in detail, and no automatic detector or classifier was evaluated. The results are therefore not used as evidence of automatic detection performance or class separability.

Even in the favorable A04 segment, where no neighboring court was active, manual annotation was still difficult. Racket, glass, and metal impacts were usually easier to locate than ground impacts, but overlapping events and weak ground responses made precise onset placement unreliable in some cases. The brief inspection of the A05 indoor recording showed an additional limitation for airborne audio. With three neighboring courts active, many transients from other courts were visible in the airborne channels, and some overlapped with impacts from the recorded court, as illustrated in Appendix A.4.1. This indicates that cross-court contamination may need to be addressed in future airborne work for amateur multi-court environments. The contact microphone was less affected by neighboring-courts sounds and showed visible responses for annotated net contacts. However, it was prone to clipping for medium and strong net contacts, even with the channel gain set low. In another inspected covered-court recording, net motions

caused by wind also created large contact microphone responses.

These observations support treating the audio recordings as documented exploratory material in this thesis rather than as a quantitatively evaluated subsystem. A quantitative audio evaluation is left for future work.

Chapter 6

Evaluation Protocol and Results

This chapter presents the evaluation protocol and results obtained for the IMU subsystem. It describes the dataset split and the procedure used to select models. It then covers first-stage detector tuning and the comparison of the second and third-stage classifiers. Finally, the frozen pipeline is evaluated once on the test partition, and the error analysis is presented.

6.1 Evaluation protocol

Figure 6.1 presents an overview of the evaluation protocol for clarity.

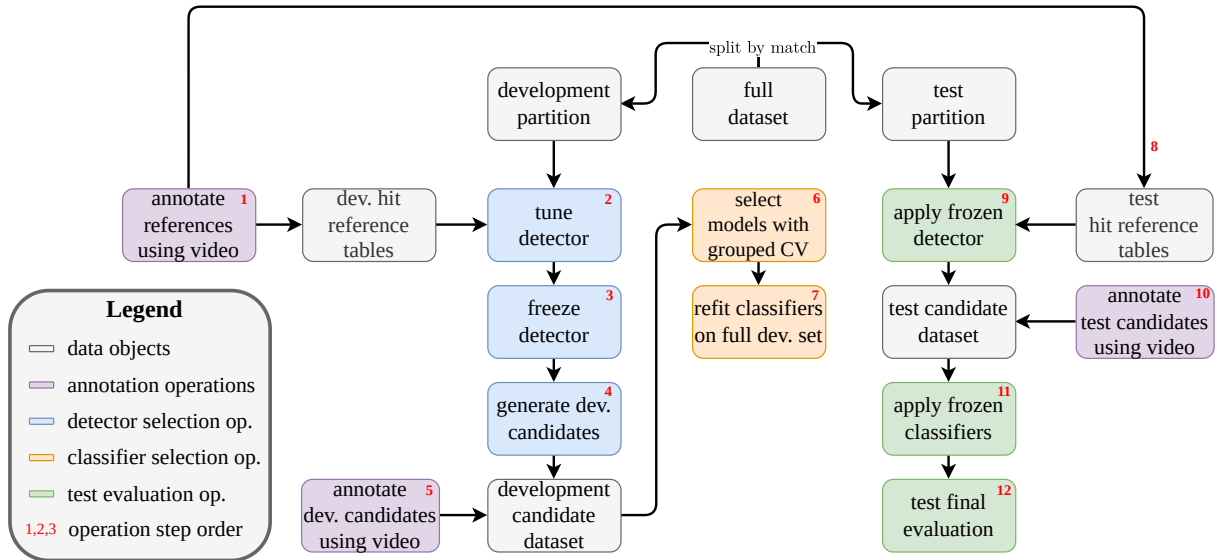


Figure 6.1: Evaluation protocol for the evaluation of the IMU subsystem

6.1.1 Data split

The IMU dataset was split by match into a development partition and a final test partition. All sessions from the same match were assigned to the same partition to reduce match-level leakage. The development partition was used for detector tuning, model selection, and hyperparameter tuning. The test partition was used only once for the final evaluation of the frozen pipeline.

The development partition contained matches 1 to 4, corresponding to sessions 1 to 14. The test partition contained matches 5 to 7, corresponding to sessions 15 to 20. The split and hit reference distribution are summarized in Table 6.1. With 7 matches available, this split was chosen to keep the test partition large enough for a meaningful evaluation. The test partition included a known racket/player crossover in a new match context, and two matches containing unseen-player/known-racket and unseen-player/unseen-racket combinations. The development partition contained 2532 hit references, and provided enough examples for selecting the detector and generating the classifier candidate dataset.

Split	Match ID	Session ID	Duration	HIT references (% of total)			EXCLUDE
				IN_PLAY	BETWEEN_POINTS	total	
dev.	1-4	1-14	10:38:54	1390 (54.9%)	1142 (45.1%)	2532	370
test	5-7	15-20	9:39:50	1475 (58.5%)	1047 (41.5%)	2522	337
total	1-7	1-20	20:18:51	2865 (56.7%)	2189 (43.3%)	5054	707

Table 6.1: Split and hit reference distribution. % are relative to the total HIT references in each split.

6.1.2 First-stage detector protocol

The objective of the first-stage detector was to keep as many in-play hits as possible and reduce the number of candidates passed to classifiers. Missing an in-play hit at this stage makes it unavailable to later stages. For this reason, recall on in-play hit was used as the main selection metric. A reference hit was considered detected when a detector configuration produced a candidate within two IMU samples. In-play recall was computed as

$$in\text{-}play\ recall = \frac{\# \text{ matched in-play reference hits}}{\# \text{ in-play reference hits}}. \quad (6.1)$$

Total hit recall, between-points recall, the number and rate per minute of unmatched candidates were also reported. At this stage, the unmatched candidates are not yet manually assigned to `NO_HIT` or `EXCLUDE`. They represent the candidate burden, not the confirmed false positives of the detector.

The grouping gap was fixed to 0.1 s as a grouping constraint before threshold selection and was not tuned in the grid search. The gyroscope gate was evaluated at the acceleration peak sample of each candidate. Detector tuning searched over combinations of the acceleration threshold $T_{|a|}$ and gyroscope threshold $T_{|g|}$ on the development partition. The detector selection rule was to choose the configuration with at least 99% in-play recall and then select the one with the lowest number of unmatched candidates. After this step, the detector configuration was frozen before generating the candidate dataset used by the classifiers.

6.1.3 Second and third-stage classifiers protocol

After the detector configuration had been frozen, it was applied to the development sessions to generate the candidate set used by the classifier stages. Candidate labeling followed the procedure described in Section 4.4.3 and is summarized in Section 6.3. Candidates labeled `EXCLUDE` were removed before classifier training, evaluation, and cascade metric computation. This implies that classifier and final cascade metrics are computed on eligible detector candidates only and not on every raw detector output. The second stage was trained to classify candidates as `HIT` or `NO_HIT`. The third stage was trained only on candidates labeled `HIT` and classified them as `IN_PLAY` or `BETWEEN_POINTS`. The test partition was labeled with the same procedure only after the complete pipeline had been selected on the development partition.

Classifiers were selected on the development partition using nested cross-validation grouped by match. For each candidate model, representation, and window size, hyperparameters were tuned using an inner `GroupKFold` split. Performance was estimated in an outer `GroupKFold`. In each outer fold, one development match was left out for validation. The best estimator selected in the inner loop was refit on the outer training matches and evaluated on the left out match. This grouping avoided sharing windows from the same match context in training and validation folds. This was important because sessions from the same match share players, rackets, and recording context. For each stage, two candidate window sizes were evaluated: 200 ms and 400 ms, using the same representations and classifiers. This tested whether one stage would benefit from longer temporal context. Feature normalization was computed only on the training data in each fold and then applied to the corresponding validation data.

For each classifier, the evaluation metrics were computed as

$$\begin{aligned}\text{recall}_c &= \frac{N_{\text{correct},c}}{N_c} \\ \text{balanced accuracy} &= \frac{\text{recall}_{\text{pos.}} + \text{recall}_{\text{neg.}}}{2} \\ F_1 &= \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}\end{aligned}$$

where N_c is the number of candidates of class c , and $N_{\text{correct},c}$ is the number of candidates of class c correctly predicted. For stage 2, the positive class was HIT. HIT retention was treated as a constraint because a missed hit cannot be recovered later in the cascade. Hyperparameters were tuned using HIT F1-score inside each inner fold. HIT recall was not used for hyperparameter tuning to avoid favoring degenerate models that predicted every candidate as HIT. Among configurations that reached at least 90% mean HIT recall on the cross-validation, balanced accuracy was used as the main selection criterion to avoid selecting a classifier that accepts too many NO_HIT. For stage 3, the positive class was IN_PLAY. Balanced accuracy was the main metric because both IN_PLAY and BETWEEN_POINTS errors were relevant. IN_PLAY recall and F1-score were also reported.

With 4 development matches, the cross-validation results are interpreted as a model-selection procedure only. The final test evaluation provides the main estimate of generalization in this thesis. The selected configurations were then refit on the full development partition, with final hyperparameters selected by grouped cross-validation on development partition.

6.1.4 Final evaluation protocol

After detector tuning and classifier selection, the complete pipeline was frozen and evaluated once on the test partition. The frozen pipeline used the first-stage detector with $T_{|a|} = 11$ g, $T_{|g|} = 200$ dps, and $R = 0.1$ s. Stage 2 used temporal SVM RBF with 200 ms window. Stage 3 used temporal SVM RBF with 400 ms window. For the test partition, detected candidates were labeled from reference video before inspecting any classifier predictions. The test partition was not used for detector tuning, model selection or hyperparameter optimization and no parameter was changed after test predictions were evaluated.

The main final metric was the percentage of reference in-play hits that survived the complete cascade as final IN_PLAY predictions. The final IN_PLAY precision

and false IN_PLAY predictions per minute were also reported. The per-minute rates were computed on the decoded test recording duration, which still includes intervals later treated as excluded. These rates may therefore be slightly diluted compared with rates only over active match play time.

6.2 First-stage detector tuning

Detector tuning was performed on the development partition using the protocol defined in Section 6.1.2. The grid search results are reported in Appendix A.3.1. Figure 6.2 shows the trade-off between recall and unmatched candidate burden across configurations.

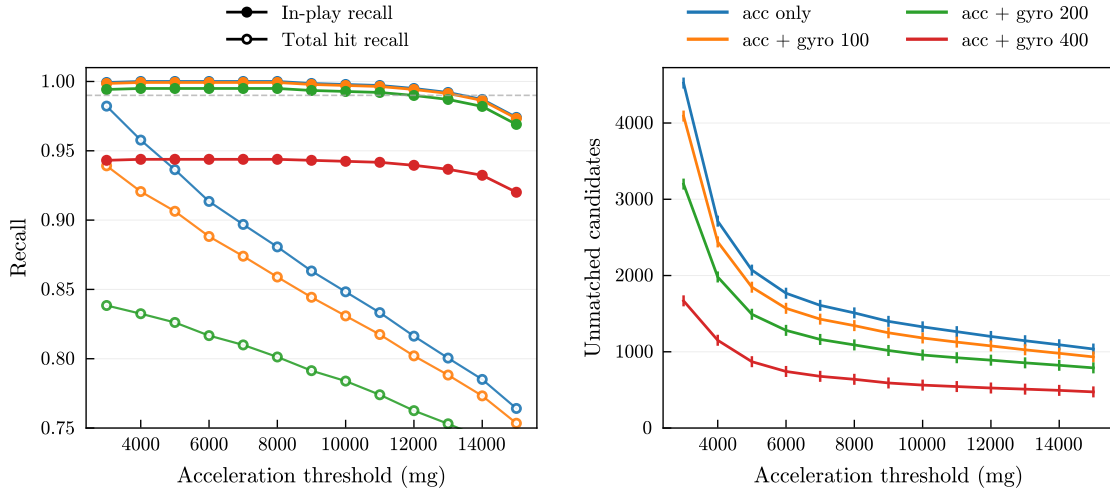


Figure 6.2: Detector tuning on the development partition. 99% recall selection threshold is marked by gray dashed line.

The fixed grouping and gyroscope gate choices were checked on the development partition. The grouping gap of 0.1 s was retained because no consecutive in-play hit references in the development partition were separated by less than this value. Additionally, a gyroscope-window gate was tested as a sensitivity check. It retained only marginally more in-play hits than the gyroscope gate at the $|a|$ peak, but allowed more between-points hits and increased the number of unmatched candidates. The gyroscope gate based on the $|a|$ peak was therefore retained because the first-stage objective was high in-play recall with minimal candidate burden. The selected detector used

$$T_{|a|} = 11000\text{ mg}, \quad T_{|g|} = 200\text{ dps}, \quad R = 0.10\text{ s}.$$

It was then inspected per match. Table 6.2 reports the detector performance for each development match and for the full development partition.

Match	HIT refs.	In-play recall	Total hit recall	Between-points recall	Unmatch. cand.	Unmatch. /min
1	508	98.5%	71.3%	42.5%	272	1.65
2	819	99.3%	76.9%	52.4%	256	1.44
3	1020	99.5%	79.9%	51.9%	364	1.50
4	185	99.0%	82.7%	63.1%	29	0.54
Total	2532	99.2%	77.4%	50.9%	921	1.44

Table 6.2: Per match and total development performance of the selected first-stage detector

The detector retained 1379 of 1390 in-play hit references, corresponding to an in-play recall of 99.2%. This meets the first stage objective, which was keeping almost all in-play ball-racket contacts before classification. The detector produced 2881 candidates, corresponding to 4.51 candidates per minute. This includes 921 unmatched candidates, or 1.44 unmatched candidates per minute.

Total hit recall was 77.4% (1960/2532), and between-points recall was 50.9% (581/1142). The detector should be interpreted primary as an in-play hit detector. Figure 6.3 shows that most missed references were between-points ball catches and retrievals.

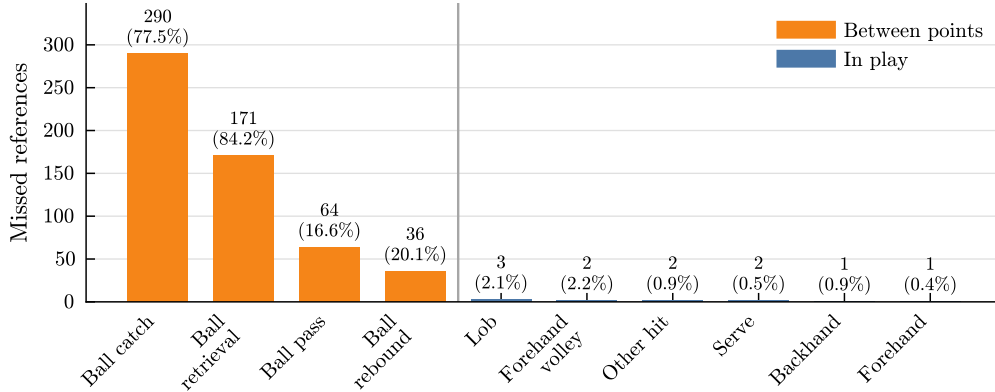


Figure 6.3: Missed hit references by phase on development partition. % are proportion missed within each sublabel.

6.3 Detector candidate dataset

The frozen detector was applied on the development partition to generate the candidate set used for classifier selection. The resulting candidates were labeled using the annotation procedure described in Section 4.4.3. This dataset is conditional on the frozen detector. The final detector development dataset contained 1379 HIT IN_PLAY, 581 HIT BETWEEN_POINTS, 408 NO_HIT, and 513 EXCLUDE. Both classifier datasets were imbalanced: the second stage class balance was 82.7% HIT/17.3% NO_HIT and the third stage was 70.4% IN_PLAY/29.6% BETWEEN_POINTS. The detailed sublabel distribution is reported in Appendix A.3.2

6.4 Classification selection

The classifiers were selected using the detector candidate dataset described above, and the protocol defined in Section 6.1.3. Full comparison results are reported in Tables A.8 and A.9. Table 6.3 summarizes the best configuration for each stage and window size. Additional development cross-validation confusion matrices for the selected configurations are reported in Appendix A.3.4. These results were used only for model selection.

Stage	Window size	Model	Repres.	Pos. recall	Neg. recall	Pos. F1	Balanced acc.
2	200	svm rbf	temporal	95.6 ± 1.5	87.1 ± 2.2	96.5 ± 0.4	91.3 ± 0.4
2	400	svm rbf	temporal	97.6 ± 0.4	80.4 ± 7.5	97.1 ± 0.7	89.0 ± 3.8
3	200	svm rbf	temporal	88.6 ± 8.9	84.9 ± 3.2	90.5 ± 4.9	86.7 ± 4.5
3	400	svm rbf	temporal	92.9 ± 7.3	83.1 ± 5.2	92.5 ± 3.6	88.0 ± 3.4

Table 6.3: Best configuration results (mean $\pm$ std across cross-validation folds) on the development partition for each stage and window size. The positive classes for stage 2 and 3 are HIT and IN_PLAY respectively.

The temporal SVM with RBF kernel was selected for both stages. For stage 2, the 200 ms window was chosen because 400 ms window increased HIT recall from 95.6% to 97.6%, but reduced NO_HIT recall from 87.1% to 80.4% and lowered balanced accuracy from 91.3% to 89%. The 200 ms window provided a better trade-off between hit retention and non-hit rejection. For stage 3, the 400 ms window was chosen because it improved IN_PLAY recall from 88.6% to 92.9% and balanced accuracy from 86.7% to 88.0%. In addition, no overlap between 400 ms HIT candidates windows was observed on the development partition.

6.4.1 Second-stage classification selection

Figure 6.4 shows the grouped cross-validation results for stage 2 using the chosen 200 ms windows. The chosen temporal SVM RBF reached the second highest HIT recall and highest balanced accuracy. Temporal random forest reached slightly higher HIT recall but showed lower and more variable NO_HIT candidates rejection. In the development cross-validation, the temporal representation improved hit retention and non-hit rejection compared to the handcrafted representation for the chosen classifier.

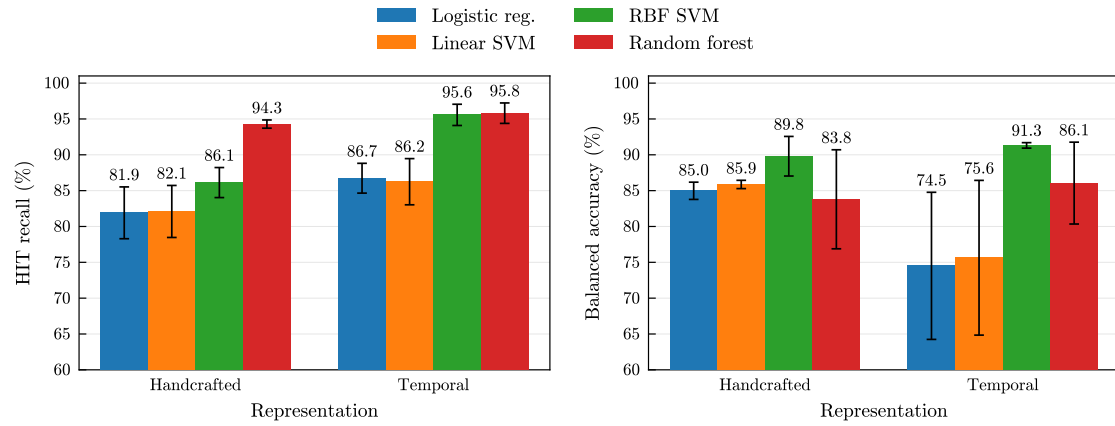


Figure 6.4: Grouped cross-validation results for second-stage classification with 200 ms candidate window: (*left*) HIT recall (*right*) balanced accuracy

6.4.2 Third-stage classification selection

Figure 6.5 shows the grouped cross-validation results for stage 3 using 400 ms window. The chosen temporal SVM RBF reached the highest balanced accuracy. Compared to stage 2, the standard deviation across folds was larger. The results suggest that, in the development cross-validation, the temporal representation and the larger candidate window may better capture information relevant to separate IN_PLAY and BETWEEN_POINTS hits.

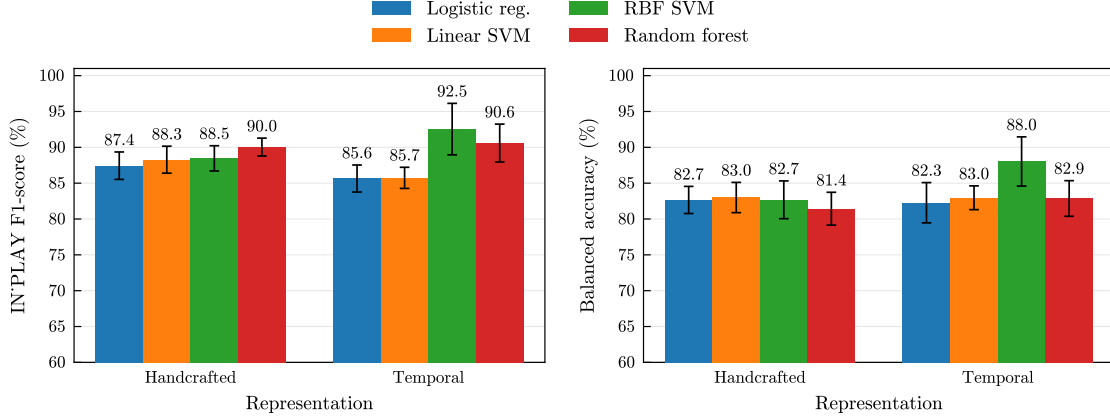


Figure 6.5: Grouped cross-validation results for third-stage classification with 400 ms candidate window: (*left*) F1-score (*right*) balanced accuracy

6.5 Final evaluation on test partition

This section reports the results of the frozen pipeline on the test partition using the protocol defined in Section 6.1.4. Table 6.1 already reported the split distribution and total count of hit references used.

6.5.1 First-stage detector performance

Table 6.4 reports the performance of the stage 1 detector on test partition. The detector retained 1461 of 1475 in-play hit references, corresponding to an in-play recall of 99.1%. This high recall was consistent across matches, ranging from 98.4% to 99.5%. Total hit recall was lower at 84.0% (2118/2522) because between-points hits were retained less consistently. This showed that the detector kept the high-recall on the test partition. The detector produced 661 unmatched candidates, corresponding to 1.14 unmatched candidates per minute.

Match	cand.	In-play ref.	In-play recall	Between-points recall	Total hit recall	Unmatch. cand.	Unmatch. /min
5	827	439	99.5%	60.8%	85.0%	230	1.20
6	1123	597	99.2%	68.9%	86.9%	249	1.19
7	829	439	98.4%	57.3%	79.5%	182	1.01
Total	2779	1475	99.1%	62.8%	84.0%	661	1.14

Table 6.4: Per-match and total final evaluation performance of the selected first-stage detector

After stage 1 on the test partition, the detector produced 2779 candidates: 1461 HIT/IN_PLAY, 657 HIT/BETWEEN_POINTS, 301 NO_HIT and 360 EXCLUDE candidates.

6.5.2 Second and third-stage classifiers performance

Table 6.5 reports the final performance of the selected stage 2 and 3 classifiers on the test partition. These metrics are conditional on the frozen detector candidate set. Stage 3 is evaluated here on detected candidates with true HIT labels. The full cascade result, where stage 3 receives candidates passed by stage 2, is reported in Section 6.5.3.

Match	Stage 2				Stage 3			
	HIT recall	NO_HIT recall	HIT F1	Balanc. acc.	IN_PLAY recall	BETWEEN_POINTS recall	IN_PLAY F1	Bal. acc.
5	95.0	52.6	96.7	73.8	90.2	96.2	94.1	93.2
6	95.7	75.4	94.9	85.5	88.2	95.7	92.7	92.0
7	97.5	78.7	97.5	88.1	95.6	87.4	94.7	91.5
Total	96.0	74.8	96.2	85.4	91.0	93.2	93.8	92.1

Table 6.5: Classifier final evaluation performance (in %) for selected stage 2 and 3 classifiers

Stage 2 retained 96.0% of true HIT candidates and rejected 74.8% of NO_HIT candidates, with a balanced accuracy of 85.4%. This suggests that the classifier remained permissive for hit retention, even though approximately one quarter of NO_HIT candidates were still accepted as HIT. Stage 3 reached 91.0% IN_PLAY recall and 93.2% BETWEEN_POINTS recall, with a balanced accuracy of 92.1%.

Figure 6.6 shows the final test confusion matrices. Stage 2 mainly acts as a permissive hit retention. It retains most HIT candidates but allows approximately a quarter of hard negative as HIT. Stage 3 shows a more balanced separation between IN_PLAY and BETWEEN_POINTS.

6.5.3 Cascaded pipeline performance

The previous sections evaluate each stage separately. The final IMU subsystem is meant to operate as a cascade: a reference in-play hit must first be detected by stage 1, then kept as HIT by stage 2, and finally classified as IN_PLAY by stage 3. The main metric is the number of reference IN_PLAY hits that survive the cascade.

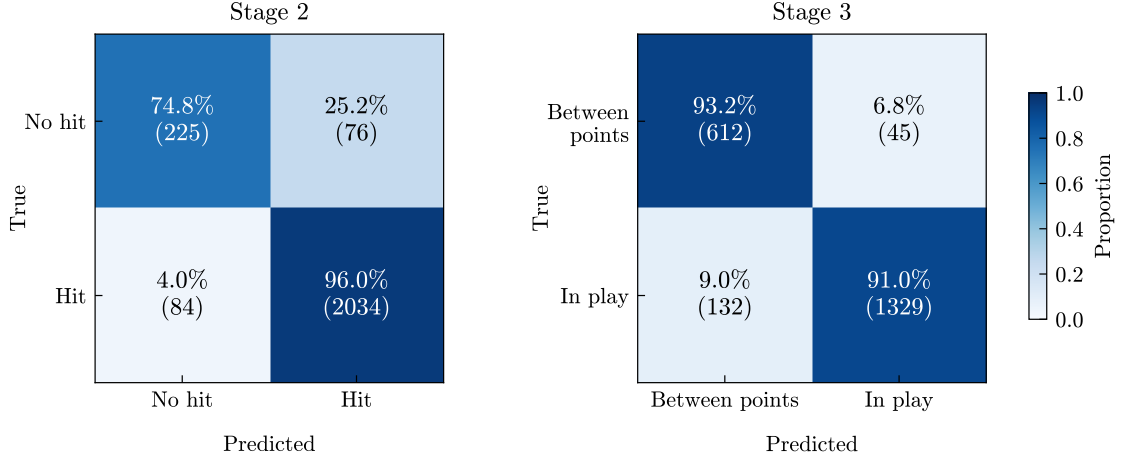


Figure 6.6: Final test confusion matrices for the stage 2 and 3 classifiers

Table 6.6 shows results of the cascade per match and across the test partition. Out of 1475 reference `IN_PLAY` hits in the full test partition, 1461 were retained by the detector, 1448 were predicted as `HIT` in stage 2, and 1320 were predicted as `IN_PLAY` in stage 3. The cumulative `IN_PLAY` recall after each stage is 99.1%, 98.2%, and 89.5%, respectively. Recall and precision varied differently across test matches. Match 7 had the best hit retention but produced the largest number of false in-play prediction.

Match	In-play ref.	Stage 1 detected	Stage 2 retained	Final in-play	Full pipeline recall	False in-play pred.	Precision	False in-play /min
5	439	437	429	388	88.4%	9	97.7%	0.047
6	597	592	590	522	87.4%	27	95.1%	0.130
7	439	432	429	410	93.4%	35	92.1%	0.195
Total	1475	1461	1448	1320	89.5%	71	94.9%	0.122

Table 6.6: Per-match and total cascaded final evaluation performance of the frozen IMU pipeline

After removing `EXCLUDE` candidates, the final output of the cascade consisted of 1391 final `IN_PLAY` predictions. 1320 were true positives and 71 were false `IN_PLAY` predictions. This corresponds to 94.9% precision. Using the test duration of 09:39:50, it equals 0.122 false `IN_PLAY` predictions per minute.

6.6 Error analysis

Table 6.7 reports the main sources of error in the final evaluation. For stage 2, false HIT predictions correspond to NO_HIT candidates accepted as hits, and false NO_HIT predictions correspond to rejected hit candidates. For the final cascade output, a false IN_PLAY prediction can come from two sources: BETWEEN_POINTS hit classifier as IN_PLAY, and NO_HIT accepted as HIT by stage 2 and classified as IN_PLAY by stage 3. Lost IN_PLAY hits are in-play hits detected and kept by stage 2 but classified as BETWEEN_POINTS by stage 3.

Stage 2				Stage 3			
Error group	Sublabel	Count	Share	Error group	Sublabel	Count	Share
false HIT predictions	racket-body	46	60.5%	cascade false IN_PLAY predictions	ball pass	38	53.5%
	air swing	12	15.8%		NO_HIT passed	26	36.6%
	racket-ground	9	11.8%		ball rebound	3	4.2%
	hand-tap	7	9.2%		ball retrieval	3	4.2%
	handling	1	1.3%		ball catch	1	1.4%
	other	1	1.3%	lost IN_PLAY hits	forehand	29	22.7%
false NO_HIT predictions	ball rebound	49	58.3%		backhand	19	14.8%
	ball catch	13	15.5%		other hit	18	14.1%
	forehand volley	6	7.1%		lob	17	13.3%
	ball pass	5	6.0%		serve	16	12.5%
	ball retrieval	4	4.8%		backhand volley	16	12.5%
	other hit	4	4.8%		forehand volley	11	8.6%
	serve	2	2.4%		smash	2	1.6%
	smash	1	1.2%				

Table 6.7: Error sources in the final evaluation of stage 2 and 3 classifiers

In stage 2, false HIT predictions were mainly dominated by racket-body contacts, corresponding to 46 of 76 errors. False NO_HIT predictions were mainly between-points hits: ball rebound, catches, pass and retrievals accounted for 71 of 84 errors, or 84.5% of missed hits. Only a few in-play hits were lost at this stage, and they were not focused on a single stroke type.

The final cascade produced 71 false final IN_PLAY predictions. The largest source was between-points ball passes, which accounted for 38 of the 71 false predictions. This is plausible because a ball pass after a point can be any stroke type, and looks like an in-play stroke in a local IMU window. The second largest source was NO_HIT candidates accepted by stage 2 then classified IN_PLAY by stage 3.

Table 6.6 already summarized the lost in-play hits in the cascade. The largest loss of in-play hit reference happened at stage 3, where 128 detected and retained by stage 2 in-play hits were classified as `BETWEEN_POINTS`. These losses were distributed across all sublabels of in-play hits. This suggests that separating active play context from a short IMU window remains difficult. The difficulty does not seem limited to one particular stroke category.

6.7 Summary

This chapter evaluated the IMU subsystem using a development and test split grouped by match. The first-stage detector selected on the development partition retained 99.1% of reference in-play hits on the test partition. The selected stage 2 and stage 3 classifiers were temporal RBF SVM using 200 ms and 400 ms candidate windows respectively. In the cascade, 1320 of 1475 reference `IN_PLAY` hits were recovered as final `IN_PLAY` predictions, corresponding to an 89.5% cascaded recall. The final output contained 71 false `IN_PLAY` predictions, corresponding to 94.9% precision and 0.122 false `IN_PLAY` predictions per minute. The main remaining errors were between-points ball passes classified as `IN_PLAY` and in-play hits assigned to `BETWEEN_POINTS` by the phase classifier.

Chapter 7

Discussion and Limitations

This chapter interprets the IMU evaluation results and discusses the main limitations of the study. The following first considers the performance of the three-stage pipeline, then discusses dataset and annotation limitations, the exploratory status of the audio subsystem, and the implications for future padel match-analysis systems.

7.1 Interpretation of the IMU pipeline

7.1.1 Main result

The final evaluation shows that a racket-mounted IMU can provide useful time-stamped evidence for in-play ball-racket contacts in continuous amateur padel recordings. Indeed, the first-stage detector retained 99.1% of these reference hits on the test partition. After the full cascade, 89.5% were recovered as final `IN_PLAY` prediction, with 94.9% precision. This corresponds to 0.266 missed in-play hits per minute and 0.122 false `IN_PLAY` predictions per minute, over the evaluated decoded duration of test recordings. However, this remains insufficient for higher-level match interpretation tasks such as rally boundaries, serve validation, and score state tracking.

7.1.2 Cascade behavior and remaining errors

The three-stage structure was useful because each stage had a distinct role. The first stage fulfilled its role by preserving nearly all in-play hits before classification. This suggests that simple acceleration magnitude threshold, combined with a gyroscope magnitude gate at acceleration magnitude peak, was sufficient to detect most in-play racket impacts in the collected recordings. The lower total hit recall reflects the design focus on in-play hits rather than coverage of all ball-racket contacts,

such as catches and retrievals. If complete tracking of all ball-racket contacts is required, the detector would need to be made more permissive or between-point contacts would need to be handled separately.

The second stage classifier mainly acted as a permissive hit-verification stage. It retained most true HIT candidates and rejected three quarters of hard negative racket events. The remaining false HIT predictions still represented one quarter of NO_HIT: they were mostly racket-body contacts, air swings, racket-ground contacts, and hand taps. These errors are plausible because such actions can produce short impact IMU pattern similar to the signal of ball-racket contacts.

When evaluated on true HIT candidates, the third stage performed well as a phase classifier. Its remaining errors however directly affected the final cascade, because this stage determined whether a retained hit appeared as an IN_PLAY prediction. The main false IN_PLAY predictions were ball passes. Unlike the previous errors, these were real ball-racket contacts, but they occurred between points. They can therefore be difficult to separate from in-play hits using only a short IMU window. Additional temporal context or complementary sensing could help resolve these cases. Some true in-play strokes were also classified as BETWEEN_POINTS, and these errors were distributed across the in-play sublabels rather than concentrated in one stroke type.

Overall, the remaining errors suggest two main directions for improvement: better rejection of the hardest negative racket actions, and better estimation of play phase from local IMU windows. The selection of a larger 400 ms window for stage 3 is also consistent with this interpretation, because phase classification appears to benefit from more temporal context compared to stage 2 hit verification.

7.1.3 Test match variation

The pipeline was relatively stable across the three test matches, including match contexts with unseen players and rackets. This suggests that the system did not depend only on exact player-racket combinations seen during training within the limited test set. The match-level split also reduced leakage between development and test data, as previously explained. Within the available test set, the results provide an initial indication of generalization to new match contexts. Match 7 produced the highest cascaded recall but also the lowest precision, even though it contained known player-racket combinations. This indicates that precision was also affected by match-specific behaviors and not only by player or racket familiarity.

7.2 Limitations of the evaluation

7.2.1 Reference labels and annotation

The reference labels were produced from practical video annotation. Ambiguous or poorly visible events were labeled **EXCLUDE** to reduce label noise. Out-of-scope periods such as warm-up or drills during pauses were not used for classifier training and evaluation because they did not correspond to the target match play task and were not relevant to the intended scoring or match analysis use case. Similar actions may also be performed differently during warm-up or drills than during normal rallies or between-points periods, which could introduce inconsistent sample into the target classes. The reported performance therefore applies to the defined scope of match play, and not all activity present in the raw recordings. Some annotation uncertainty remains. The annotations were produced by a single annotator, so there was no independent check of labeling consistency. The single camera viewpoint also made some events difficult to distinguish, especially ball retrievals and ground-racket contacts. Future reference datasets would benefit from additional annotators to check labeling consistency, a clearer taxonomy for ambiguous actions, and a second camera angle for difficult cases.

7.2.2 Dataset coverage and generalization

As previously noted, the dataset was collected during real amateur match play, which improves ecological validity compared to isolated recordings. The players were however mostly intermediate level amateurs from a similar age range. This limits population diversity and means that players, rackets, stroke types, and hard negative actions were represented unevenly. The test partition included unseen players and rackets, which provides an initial check of generalization. The results are also dependent to the selected sensor placement near the racket throat. Different mounting positions, protection layer, racket models could change the impact response measured and should be tested before transferring to other hardware setups and mounts. Additionally, the number of player, rackets and test matches remained limited. A broader dataset would be needed to evaluate whether the pipeline remains stable across new players, rackets, skills levels, sensor placement, and match conditions.

Hard negative events also depended on player habits. Actions such as racket-glass taps with the racket before serving, teammate racket checks, or cheering hand taps between plays, or other preparation routines may vary between players and were not covered exhaustively. Future datasets should introduce a wider variety of players and playing habits to test the pipeline against a wider range of racket actions.

7.2.3 Conditioned classifier dataset and offline evaluation

The classifier dataset was conditioned on the chosen detector. The `NO_HIT` class does not represent arbitrary background windows from the continuous stream, but reviewed candidates produced by the frozen detector that did not correspond to `HIT` and occurred in the target recording scope. This matches the evaluated cascade, because the classifiers are only applied after candidate detection. However, it also means that the classifier results are specific to the selected detector and its candidate distribution. Changing the detector threshold or grouping rule would change this distribution and could therefore change classifier performance.

Candidates labeled `EXCLUDE` were removed from classifier training, evaluation, and cascade metric computation. These candidates corresponded to out-of-scope events, as described previously. The reported precision and false-positive counts therefore describe performance within the defined match play evaluation scope. This also means that unlabeled background samples in the raw IMU recordings were not treated as `NO_HIT` because some unlabeled intervals may contain out-of-scope events that would have been labeled `EXCLUDE` if reviewed. In addition, per-minute rates were computed over the full decoded test duration. They should therefore be interpreted as error rates over the evaluated recordings, not only over active rallies.

Furthermore, the evaluation was performed offline from recorded data. The reported results therefore describe offline detection and classification performance rather than live deployment performance.

7.3 Exploratory status of the audio subsystems

The audio subsystem was documented as an exploratory extension. The inspected examples suggest that both airborne and contact-microphone signals can contain useful event information. However, the limitations observed in Chapter 5 prevent quantitative conclusions about audio detection performance or class separability.

For airborne audio, the main limitations were slow manual annotation, weak or overlapping ground-impact responses, and cross-court contamination in the multi-court indoor recording. These observations are specific to the tested setup, which used two low-cost omnidirectional microphones near the court structure in noisy amateur multi-court conditions. They do not rule out airborne audio under other microphone placements, microphone types, or more controlled recording

conditions, but such alternatives would need a separate quantitative evaluation.

The contact microphone appears to be worth further investigation for net contact detection because it produced localized responses for visible ball-net contacts and appeared less affected by neighboring court sounds. However, the current mounting and sensor protection were preliminary. Future work should thus improve the mechanical mounting, protect the sensor from direct impacts, reduce clipping, and test robustness to net motion caused by wind. Under the current setup, audio should therefore remain exploratory rather than a priority subsystem for quantitative fusion.

7.4 Implication for future systems

The results show progress on the event detection layer conceptualized in Chapter 3. The pipeline can provide a useful stream of timestamped racket-contact events but this stream is not sufficient for full match interpretation by itself. A larger system would still need to combine hit timing with player or side attribution, rally context, and game state tracking. The remaining errors, especially between-points ball passes, show that isolated hit classification should be complemented by temporal context and higher-level match reasoning. This would move the system from local racket event detection toward rally interpretation.

Chapter 8

Conclusion and Future Work

8.1 Conclusion

This thesis investigated racket-mounted inertial sensing for event detection in padel. The long-term motivation is automatic match analysis and scoring support, but the evaluated scope was limited to detecting ball-racket contacts during real match play and estimating whether those contacts occurred during active play or between points.

A low-cost racket-mounted inertial recording system was designed and assembled using accessible components. It was used to collect continuous amateur padel recordings. The data was synchronized with video and annotated using a semi-automatic annotation tool developed for this thesis, producing reference labels for ball-racket contacts, other racket actions, and play phase. The proposed pipeline was then evaluated on these continuous recordings. The final evaluation showed that the first stage retained 99.1% of reference in-play hits. After the full cascade, 89.5% of in-play hit were recovered, with 94.9% precision and 0.122 false in-play prediction per minute. Overall, racket-mounted inertial sensing provided a useful event source for in-play ball-racket contacts. The main remaining challenge is handling edge cases after hit detection such as ball passes.

Finally, the audio subsystems were documented as exploratory extensions, but were not quantitatively evaluated because of the limitations observed during the qualitative inspection. In conclusion, the thesis supports racket-mounted IMU sensing as a first step toward an inertial sensing layer for padel match analysis.

8.2 Future work

Future work should first validate the IMU pipeline on a broader dataset. The current test set included new match contexts and unseen players or racket, but the number of matches remained limited. A larger dataset should include more players, skill levels, racket models, sensor placements, and play dynamics. It should also include a wider range of player habits, because actions such as hand-taps, teammate racket checks, and air swings may produce hard negative events with different inertial signatures.

A second research direction is live deployment. The first-stage appears to be suitable for real-time triggering because it relies on simple acceleration and gyroscope thresholds. The classifier stages would require further testing on embedded hardware, including the inference latency, memory use, prediction delay, wireless synchronization, battery life, and physical robustness during play. A natural extension would be to instrument all rackets during a match. This would allow detected events to be associated with players or teams and could produce a simple inertial event timeline for player and team analysis.

Furthermore, additional work could improve the play-phase estimation. The current pipeline estimates the play phase of a detected hit using a short local IMU window. Future work could investigate longer temporal windows, recent racket activity, service related cues, background activity level, simple game-state logic, or provide further context. Alternative classifier designs, such as a flat multiclass classifier could also be compared with the three-stage cascade. Stroke type classification could be added later to investigate whether serve or stroke information helps identify rally boundaries.

The audio subsystem should remain exploratory under the current hardware, annotation workflow, and uncontrolled amateur multi-court recording conditions. The recordings showed that audio can contain targeted event responses, but the annotation cost, cross-court contamination, contact-microphone mounting, clipping, and wind sensitivity made it not ready for quantitative evaluation in this thesis.

For scoring support, the most direct next step is therefore to strengthen the IMU event layer and combine it with multiple rackets, temporal context, and higher-level match logic. Together, these directions would extend the current single-racket pipeline toward a more complete player and team analysis system.

Bibliography

- [1] J.-F. van Drooghenbroeck. “Utilisation des intelligences artificielles génératives note à l’attention des étudiantes et étudiants de l’UCLouvain,” Accessed: May 20, 2026. [Online]. Available: https://www.uclouvain.be/fr/system/files/uclouvain_assetmanager/groups/cms-editors-ia/documents/Utilisation_IAGenerative_Etudiants.pdf.
- [2] International Padel Federation. “FIP world padel report 2025,” Accessed: May 8, 2026. [Online]. Available: <https://www.padelfip.com/world-padel-report-2025/>.
- [3] European Data Protection Board. “Guidelines 3/2019 on processing of personal data through video devices,” Accessed: May 8, 2026. [Online]. Available: https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-32019-processing-personal-data-through-video_en.
- [4] PlaySight. “Padel SmartCourt - PlaySight,” Accessed: May 20, 2026. [Online]. Available: <https://playsight.com/our-sports/padel/>.
- [5] GAMECAM AB. “AI camera for padel | live streaming,” Accessed: May 20, 2026. [Online]. Available: <https://www.gamecam.io/>.
- [6] Playtomic Smart Courts. “Padel | playtomic smart courts,” Accessed: May 20, 2026. [Online]. Available: <https://www.playtomicsmartcourts.com/padel>.
- [7] Scorepadel. “Padel court scoreboard,” Accessed: May 20, 2026. [Online]. Available: <https://scorepadel.es/en>.
- [8] PadelPlay. “PadelPlay,” Accessed: May 20, 2026. [Online]. Available: <https://www.padelplay.ai/>.
- [9] V. Camomilla, E. Bergamini, S. Fantozzi, and G. Vannozzi, “Trends supporting the in-field use of wearable inertial sensors for sport performance evaluation: A systematic review,” *Sensors*, vol. 18, no. 3, p. 873, Mar. 2018. DOI: [10.3390/s18030873](https://doi.org/10.3390/s18030873).

- [10] C. Kranzinger et al., “Classification of human motion data based on inertial measurement units in sports: A scoping review,” *Applied Sciences*, vol. 13, no. 15, p. 8684, Jan. 2023. DOI: [10.3390/app13158684](https://doi.org/10.3390/app13158684).
- [11] C. Leddy, R. Bolger, P. Byrne, S. Kinsella, and L. Zambrano M., “The application of machine and deep learning for technique and skill analysis in swing and team sport-specific movement: A systematic review,” *International Journal of Computer Science in Sport*, vol. 23, pp. 110–145, Jun. 5, 2024. DOI: [10.2478/ijcss-2024-0007](https://doi.org/10.2478/ijcss-2024-0007).
- [12] C. J. Ebner and R. D. Findling, “Tennis stroke classification: Comparing wrist and racket as IMU sensor position,” in *Proceedings of the 17th International Conference on Advances in Mobile Computing & Multimedia*, Munich, Germany: ACM, Dec. 2, 2019, pp. 74–83. DOI: [10.1145/3365921.3365929](https://doi.org/10.1145/3365921.3365929).
- [13] J. Park, S. Yang, and S. Jo, “Silent impact: Tracking tennis shots from the passive arm,” in *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST ’24, New York, NY, USA: Association for Computing Machinery, Oct. 11, 2024, pp. 1–15. DOI: [10.1145/3654777.3676403](https://doi.org/10.1145/3654777.3676403).
- [14] M. Wu et al., “A real-time tennis level evaluation and strokes classification system based on the internet of things,” *Internet of Things*, vol. 17, p. 100494, Mar. 1, 2022. DOI: [10.1016/j.iot.2021.100494](https://doi.org/10.1016/j.iot.2021.100494).
- [15] T. Perri, M. Reid, A. Murphy, K. Howle, and R. Duffield, “Prototype machine learning algorithms from wearable technology to detect tennis stroke and movement actions,” *Sensors*, vol. 22, no. 22, p. 8868, Jan. 2022. DOI: [10.3390/s22228868](https://doi.org/10.3390/s22228868).
- [16] A. Ganser, B. Hollaus, and S. Stabinger, “Classification of tennis shots with a neural network approach,” *Sensors*, vol. 21, no. 17, p. 5703, Jan. 2021. DOI: [10.3390/s21175703](https://doi.org/10.3390/s21175703).
- [17] L. Benages Pardo, D. Buldain Perez, and C. Orrite Uruñuela, “Detection of tennis activities with wearable sensors,” *Sensors (Basel, Switzerland)*, vol. 19, no. 22, p. 5004, Nov. 16, 2019. DOI: [10.3390/s19225004](https://doi.org/10.3390/s19225004).
- [18] G. Jin and X. Li, “Wearable sensing for badminton stroke recognition with one-dimensional convolutional neural network,” *Scientific Reports*, vol. 15, no. 1, p. 41236, Nov. 21, 2025. DOI: [10.1038/s41598-025-25158-2](https://doi.org/10.1038/s41598-025-25158-2).
- [19] B. Van Herbruggen et al., “Strategy analysis of badminton players using deep learning from IMU and UWB wearables,” *Internet of Things*, vol. 27, p. 101260, Oct. 2024. DOI: [10.1016/j.iot.2024.101260](https://doi.org/10.1016/j.iot.2024.101260).

- [20] G. C. Domínguez, E. F. Álvarez, A. T. Córdoba, and D. G. Reina, “A comparative study of machine learning and deep learning algorithms for padel tennis shot classification,” *Soft Computing*, vol. 27, no. 17, pp. 12 367–12 385, Sep. 1, 2023. DOI: [10.1007/s00500-023-07874-x](https://doi.org/10.1007/s00500-023-07874-x).
- [21] G. Papaspyropoulos, E. Faliagka, T. Skandamis, C. P. Antonopoulos, and N. Voros, “Real-time padel strokes classification,” in *2025 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, vol. 1, Jul. 2025, pp. 1–4. DOI: [10.1109/ISVLSI65124.2025.11130340](https://doi.org/10.1109/ISVLSI65124.2025.11130340).
- [22] P. Blank, J. Hossbach, D. Schuldhaus, and B. Eskofier, “Sensor-based stroke detection and stroke type classification in table tennis,” in *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, Osaka, Japan: Association for Computing Machinery, Sep. 7, 2015, pp. 93–100. DOI: [10.1145/2802083.2802087](https://doi.org/10.1145/2802083.2802087).
- [23] D. Whiteside, O. Cant, M. Connolly, and M. Reid, “Monitoring hitting load in tennis using inertial sensors and machine learning,” *International Journal of Sports Physiology and Performance*, vol. 12, no. 9, pp. 1212–1217, Oct. 2017. DOI: [10.1123/ijsp.2016-0683](https://doi.org/10.1123/ijsp.2016-0683).
- [24] J. Gao and G. Zhang, “Tennis action recognition and evaluation with inertial measurement unit and SVM,” *Systems and Soft Computing*, vol. 6, p. 200 154, Dec. 1, 2024. DOI: [10.1016/j.sasc.2024.200154](https://doi.org/10.1016/j.sasc.2024.200154).
- [25] A. Bulling, U. Blanke, and B. Schiele, “A tutorial on human activity recognition using body-worn inertial sensors,” *ACM Computing Surveys (CSUR)*, vol. 46, no. 3, 33:1–33:33, Jan. 1, 2014. DOI: [10.1145/2499621](https://doi.org/10.1145/2499621).
- [26] E. Chaibub Neto et al., “Detecting the impact of subject characteristics on machine learning-based diagnostic applications,” *npj Digital Medicine*, vol. 2, no. 1, p. 99, Oct. 11, 2019. DOI: [10.1038/s41746-019-0178-x](https://doi.org/10.1038/s41746-019-0178-x).
- [27] A. Baughman, E. Morales, G. Reiss, N. Greco, S. Hammer, and S. Wang, “Detection of tennis events from acoustic data,” in *Proceedings of the 2nd International Workshop on Multimedia Content Analysis in Sports*, Nice, France: ACM, Oct. 15, 2019, pp. 91–99. DOI: [10.1145/3347318.3355520](https://doi.org/10.1145/3347318.3355520).
- [28] Q. Huang and S. Cox, “Using high-level information to detect key audio events in a tennis game,” in *Interspeech 2010*, ISCA, Sep. 26, 2010, pp. 1409–1412. DOI: [10.21437/Interspeech.2010-428](https://doi.org/10.21437/Interspeech.2010-428).
- [29] Q. Huang and S. Cox, “Hierarchical language modeling for audio events detection in a sports game,” in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, Dallas, TX, USA: IEEE, 2010, pp. 2286–2289. DOI: [10.1109/ICASSP.2010.5495935](https://doi.org/10.1109/ICASSP.2010.5495935).

- [30] R. Decorte, M. Paré, J. Vanhaeverbeke, J. Taelman, M. Slembrouck, and S. Verstockt, “Multi-modal hit detection and positional analysis in padel competitions,” in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA: IEEE, Jun. 17, 2024, pp. 3306–3314. DOI: [10.1109/CVPRW63382.2024.00335](https://doi.org/10.1109/CVPRW63382.2024.00335).
- [31] L. Caprioli et al., “Tennis timing assessment by a machine learning-based acoustic detection system: A pilot study,” *Journal of Functional Morphology and Kinesiology*, vol. 10, no. 1, p. 47, Jan. 27, 2025. DOI: [10.3390/jfmk10010047](https://doi.org/10.3390/jfmk10010047).

Appendix A

Supplementary material

A.1 Hardware details

A.1.1 Hardware costs

Table A.1 reports the approximate cost of the main components used in each subsystem. The table is included to support the claim that the sensing setup was built from accessible and relatively low-cost components.

Subsystem	Component	Approx. cost (EUR)
IMU subsystem (1x)	1x Adafruit LSM6DSO32	13.01
	1x Seeed Studio XIAO ESP32S3 Sense	14.56
	1x STEMMA QT / Qwiic JST SH 4-pin cable	0.99
	1x lithium-polymer battery, 3.7 V, 620 mAh	6.29
	Hama MicroSD card, 32 GB	5.59
total		40.44
Airborne audio	2x Behringer ECM8000 microphones	40.80
	1x Behringer UMC404HD audio interface	95.00
	2x Pro Snake TPM 20 XLR cables, 20 m	43.70
total		179.50
Contact microphone	1x Stag SMC10XP XLR cable, 10 m	11.25
	1x piezoelectric disk	0.50
total		11.75

Table A.1: Approximate cost of the main components required for the prototype sensing setup. Costs are subtotals for the listed quantities.

A.1.2 Wiring of the contact microphone

Figure A.1 shows the wiring used for the passive contact microphone prototype. Other configurations with a 21 mm disk and an additional resistor were also tested, but this configuration was chosen ultimately.

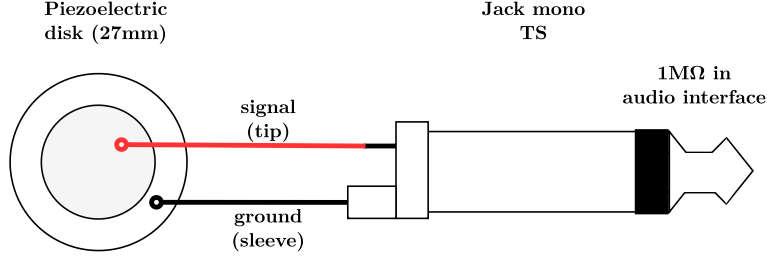


Figure A.1: Wiring diagram of the built passive contact microphone

A.2 Dataset and annotation details

A.2.1 Mapping of players and rackets

Tables A.2 and A.3 report players and rackets included in the IMU dataset. Players who participated only in audio recordings are not included in this mapping.

Player ID	Level	Sex	Age	Dominant hand	Backhand type
1	Intermediate	Male	[25-30]	Right	1-handed
2	Intermediate	Male	[25-30]	Right	1-handed
3	Beginner	Female	[25-30]	Right	2-handed
4	Intermediate	Male	[20-24]	Right	1-handed
5	Intermediate	Male	[20-24]	Right	1-handed
8	Intermediate	Male	[20-24]	Right	1-handed
9	Intermediate	Male	[20-24]	Right	1-handed
12	Intermediate	Male	[20-24]	Right	1-handed

Table A.2: Mapping of the participants included in the IMU data collection

Racket ID	Brand/Model	Shape	Weight (g)	Balance
1	Kuikma 990 Hybrid Hard	Teardrop	367 – 377	Medium
2	Head Extreme Elite 2024	Diamond	365 $\pm$ 10	High
3	Kuikma 990 Hybrid Soft	Teardrop	367 – 377	Medium

Table A.3: Mapping of the rackets used during data collection

A.2.2 Hit reference table per session

Match ID	Session ID	HIT			EXCLUDE	
		in-play	between-points	total	ambi-guous	outside play
1	S01	88	46	134	2	41
	S02	25	13	38	0	0
	S06	105	120	225	2	45
	S07	43	68	111	1	38
	total	261	247	508	5	124
2	S03	89	57	146	5	0
	S04	54	57	111	2	0
	S05	54	40	94	2	8
	S08	117	123	240	1	7
	S09	114	114	228	5	0
	total	428	391	819	15	15
3	S10	306	243	549	5	151
	S12	175	127	302	2	28
	S13	119	50	169	1	19
	total	600	420	1020	8	198
4	S11	52	32	84	0	3
	S14	49	52	101	2	0
	total	101	84	185	2	3
5	S15	252	181	433	8	32
	S16	187	82	269	5	170
	total	439	263	702	13	202
6	S17	296	212	508	5	13
	S18	301	197	498	6	4
	total	597	409	1006	11	17
7	S19	229	229	458	3	43
	S20	210	146	356	8	40
	total	439	375	814	11	83
total	20 ses.	2865	2189	5054	65	642

Table A.4: Summary of hit reference dataset by match and session. Total hits is the sum of in-play and between-points hits.

A.3 Evaluation and training details

A.3.1 First-stage detector tuning results

Table A.2 reports the parameter grid used for the detector search. Table A.5 summarizes the grid search of detector configurations on the development partition. For readability, it is sorted by unmatched candidates and only display 20 configurations. The selected configuration was the one with at least 99% in-play recall and with the lowest number of unmatched candidates. It is highlighted in gray.

Stage 1

$T_{ a }$ (mg)	$T_{ g }$ (dps)	In-play recall	Total hit recall	Between -points recall	Unmatch. cand.	Unmatch. cand./min
14000	200	98.2%	74.2%	45.1%	823	1.28
13000	200	98.7%	75.3%	46.8%	855	1.33
12000	200	98.9%	76.2%	48.5%	890	1.39
11000	200	99.2%	77.4%	50.8%	921	1.44
10000	200	99.2%	78.3%	52.9%	958	1.49
14000	100	98.6%	77.3%	51.4%	980	1.53
9000	200	99.3%	79.1%	54.5%	1017	1.59
13000	100	99.1%	78.8%	54.1%	1026	1.60
12000	100	99.4%	80.2%	56.8%	1077	1.68
8000	200	99.4%	80.1%	56.5%	1090	1.70
14000	0	98.7%	78.5%	53.9%	1092	1.70
11000	100	99.6%	81.7%	59.9%	1127	1.76
13000	0	99.2%	80.0%	56.7%	1145	1.79
7000	200	99.4%	80.9%	58.4%	1162	1.81
10000	100	99.7%	83.0%	62.8%	1181	1.84
12000	0	99.4%	81.6%	59.8%	1201	1.87
9000	100	99.7%	84.4%	65.7%	1249	1.95
11000	0	99.7%	83.3%	63.3%	1264	1.97
6000	200	99.4%	81.6%	59.9%	1281	2.00
10000	0	99.7%	84.8%	66.6%	1327	2.07

Table A.5: Top 20 detector configurations on the development partition, sorted by unmatched candidates count. The selected configuration is highlighted in gray.

Parameter	Values
Acceleration threshold $T_{ a }$	$\{3000, 4000, \dots, 15000\}$ mg
Gyroscope threshold $T_{ g }$	$\{0, 100, 200, 400, 800\}$ dps
Grouping gap R	0.10 s, fixed

Table A.6: Search grid used for first-stage detector selection

A.3.2 Development candidate dataset

Figure A.2 shows the sublabel distribution of the development candidate dataset generated by the frozen detector, and used for classifier selection on the development partition.

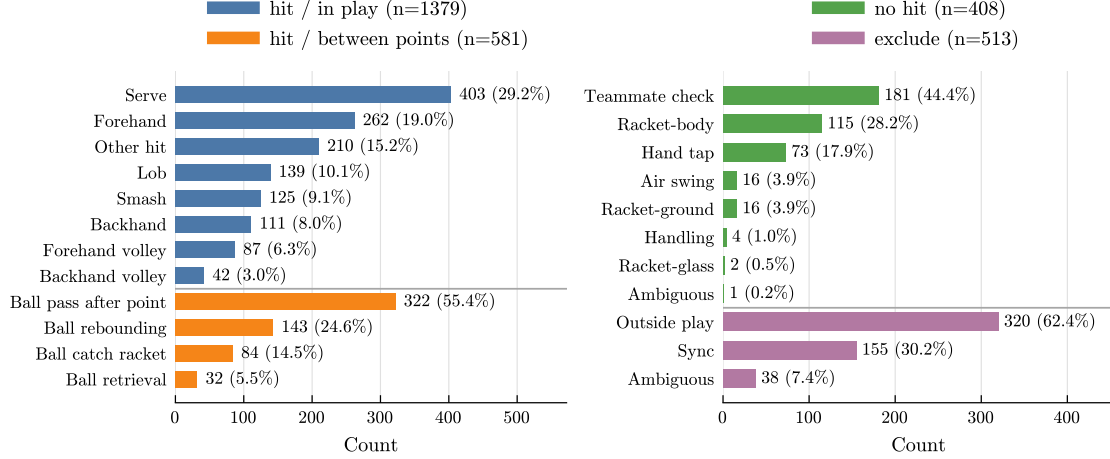


Figure A.2: Sublabel distribution of the detector candidate dataset on development partition. % are computed within each label class.

A.3.3 Classifier cross-validation results

This subsection shows the hyperparameter grid and grouped CV results used for selecting the stage 2 and stage 3 classifiers on the development partition. The hyperparameter values for each classifier were limited to a small grid. This choice was taken to reduce the risk of overfitting model selection to the small number of development matches. It also means that the reported models may not be fully optimized.

Model	Hyperparameter values
Logistic regression	$C \in \{0.1, 1, 10\}$
Linear SVM	$C \in \{0.1, 1, 10\}$
RBF SVM	$C \in \{0.1, 1, 10\}$ $\gamma \in \{\text{scale}, 10^{-3}, 10^{-2}\}$
Random forest	$n_{\text{estimators}} \in \{200\}$ $\text{max_depth} \in \{\text{None}, 5, 10\}$

Table A.7: Hyperparameter search grid used for classifier selection

Stage 2

Window	Model	Repres.	HIT recall	NO_HIT recall	HIT F1	Balanced acc.
200	RBF SVM	Temporal	95.6 $\pm$ 1.5	87.1 $\pm$ 2.2	96.5 $\pm$ 0.4	91.3 $\pm$ 0.4
200	RBF SVM	Baseline	86.1 $\pm$ 2.1	93.5 $\pm$ 5.3	91.7 $\pm$ 1.2	89.8 $\pm$ 2.8
200	Random for.	Temporal	95.8 $\pm$ 1.4	76.3 $\pm$ 11.5	96.0 $\pm$ 0.7	86.1 $\pm$ 5.7
200	Linear SVM	Baseline	82.1 $\pm$ 3.6	89.6 $\pm$ 4.7	89.1 $\pm$ 1.5	85.9 $\pm$ 0.6
200	Logistic reg.	Baseline	81.9 $\pm$ 3.6	88.1 $\pm$ 5.7	88.8 $\pm$ 1.4	85.0 $\pm$ 1.2
200	Random for.	Baseline	94.3 $\pm$ 0.6	73.3 $\pm$ 14.3	94.3 $\pm$ 2.4	83.8 $\pm$ 6.9
200	Linear SVM	Temporal	86.2 $\pm$ 3.2	65.0 $\pm$ 23.3	89.0 $\pm$ 1.5	75.6 $\pm$ 10.8
200	Logistic reg.	Temporal	86.7 $\pm$ 2.1	62.3 $\pm$ 21.7	89.1 $\pm$ 1.5	74.5 $\pm$ 10.3
400	RBF SVM	Temporal	97.6 $\pm$ 0.4	80.4 $\pm$ 7.5	97.1 $\pm$ 0.7	89.0 $\pm$ 3.8
400	RBF SVM	Baseline	85.4 $\pm$ 1.5	92.1 $\pm$ 5.8	91.2 $\pm$ 1.5	88.7 $\pm$ 3.5
400	Random for.	Temporal	95.7 $\pm$ 1.4	74.9 $\pm$ 10.9	95.8 $\pm$ 0.8	85.3 $\pm$ 5.3
400	Logistic reg.	Baseline	80.3 $\pm$ 3.7	86.3 $\pm$ 4.5	87.6 $\pm$ 1.3	83.3 $\pm$ 0.8
400	Linear SVM	Baseline	80.0 $\pm$ 4.0	84.9 $\pm$ 5.1	87.5 $\pm$ 1.6	82.5 $\pm$ 2.6
400	Random for.	Baseline	95.7 $\pm$ 1.1	65.4 $\pm$ 10.4	94.4 $\pm$ 2.3	80.6 $\pm$ 5.2
400	Linear SVM	Temporal	88.1 $\pm$ 0.9	66.0 $\pm$ 8.7	90.6 $\pm$ 1.5	77.1 $\pm$ 4.0
400	Logistic reg.	Temporal	88.3 $\pm$ 1.8	64.1 $\pm$ 4.1	90.4 $\pm$ 1.6	76.2 $\pm$ 2.1

Table A.8: Grouped cv results for HIT/NO_HIT classifier on the development partition sorted by balanced acc. The selected model is highlighted per window size.

Stage 3

Window	Model	Repres.	IN_PLAY recall	BETWEEN_POINTS recall	IN_PLAY F1	Balanced acc.
200	RBF SVM	Temporal	88.6 $\pm$ 8.9	84.9 $\pm$ 3.2	90.5 $\pm$ 4.9	86.7 $\pm$ 4.5
200	Random for.	Temporal	91.7 $\pm$ 6.8	72.5 $\pm$ 6.1	89.8 $\pm$ 2.7	82.1 $\pm$ 1.9
200	Logistic reg.	Temporal	79.6 $\pm$ 6.1	84.6 $\pm$ 4.9	85.2 $\pm$ 3.6	82.1 $\pm$ 3.6
200	Logistic reg.	Baseline	83.3 $\pm$ 5.5	80.3 $\pm$ 7.1	86.6 $\pm$ 1.4	81.8 $\pm$ 1.8
200	RBF SVM	Baseline	85.6 $\pm$ 5.5	77.8 $\pm$ 6.9	87.4 $\pm$ 1.7	81.7 $\pm$ 2.4
200	Linear SVM	Baseline	85.2 $\pm$ 5.1	77.5 $\pm$ 6.8	87.2 $\pm$ 1.9	81.3 $\pm$ 2.9
200	Linear SVM	Temporal	79.8 $\pm$ 6.0	81.9 $\pm$ 4.5	84.8 $\pm$ 3.2	80.8 $\pm$ 3.2
200	Random for.	Baseline	91.6 $\pm$ 3.6	69.5 $\pm$ 5.4	89.2 $\pm$ 1.0	80.6 $\pm$ 2.3
400	RBF SVM	Temporal	92.9 $\pm$ 7.3	83.1 $\pm$ 5.2	92.5 $\pm$ 3.6	88.0 $\pm$ 3.4
400	Linear SVM	Temporal	79.8 $\pm$ 3.1	86.2 $\pm$ 2.4	85.7 $\pm$ 1.5	83.0 $\pm$ 1.7
400	Linear SVM	Baseline	86.5 $\pm$ 6.0	79.5 $\pm$ 6.8	88.3 $\pm$ 1.9	83.0 $\pm$ 2.1
400	Random for.	Temporal	93.2 $\pm$ 6.9	72.5 $\pm$ 6.9	90.6 $\pm$ 2.6	82.9 $\pm$ 2.5
400	Logistic reg.	Baseline	84.6 $\pm$ 5.6	80.8 $\pm$ 5.1	87.4 $\pm$ 1.9	82.7 $\pm$ 1.9
400	RBF SVM	Baseline	87.4 $\pm$ 4.8	78.0 $\pm$ 6.7	88.5 $\pm$ 1.8	82.7 $\pm$ 2.6
400	Logistic reg.	Temporal	80.3 $\pm$ 1.9	84.3 $\pm$ 4.7	85.6 $\pm$ 1.9	82.3 $\pm$ 2.8
400	Random for.	Baseline	93.0 $\pm$ 2.3	69.8 $\pm$ 5.2	90.0 $\pm$ 1.2	81.4 $\pm$ 2.3

Table A.9: Grouped cv results for IN_PLAY/BETWEEN_POINTS classifier on the development partition sorted by balanced acc. The selected model is highlighted per window size.

A.3.4 Development cross-validation confusion matrices

Figure A.3 shows the aggregated cross-validation confusion matrices for the selected classifiers during grouped cross-validation on the development partition. For stage 2, 96.2% of HIT were correctly retained. 14% NO_HIT candidates were incorrectly classified as HIT. For stage 3, 93.3% IN_PLAY hits were correctly labeled. 15.3% BETWEEN_POINTS hits were labeled as IN_PLAY.

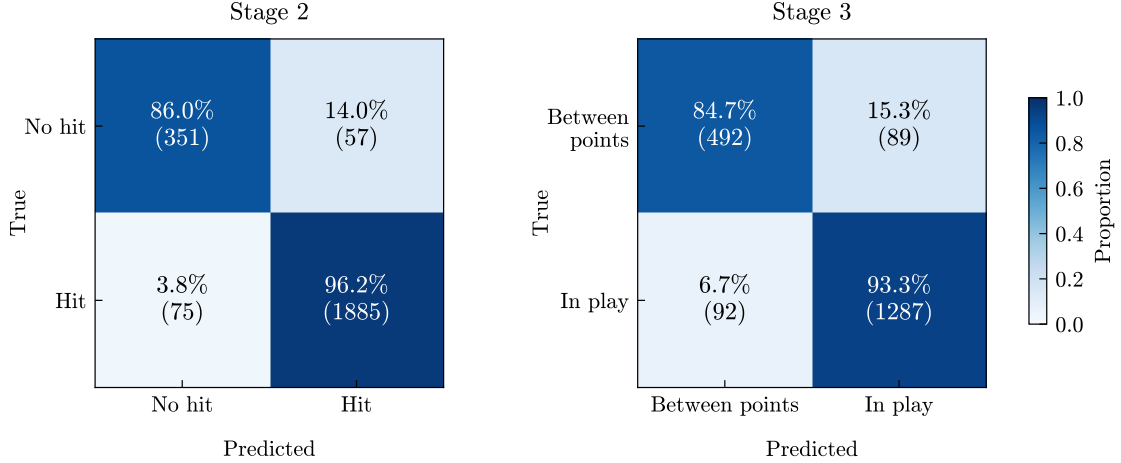


Figure A.3: Aggregated grouped CV confusion matrices for the selected second and third-stage temporal SVM RBF on the development partition.

A.4 Additional audio inspection examples

A.4.1 Neighboring-court contamination

Figure A.4 shows an example from the A05 indoor recording, where three neighboring courts were active. The example illustrates clear acoustic events in the airborne signal, highlighted in red, that are not associated with impacts on the recorded court.

A.4.2 Closely spaced rally impacts

Figure A.5 shows an example from a rally where several impacts occur within a short time interval. The sequence contains repeated racket impacts followed by ground rebounds. Later ground rebounds in the figure occur during the decay of louder racket impact responses.

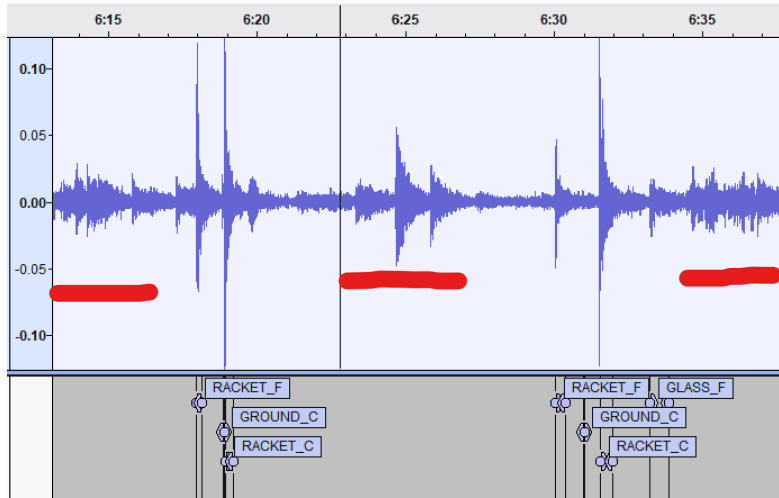


Figure A.4: Example of cross-court contamination in the airborne microphone waveform from A05 recording

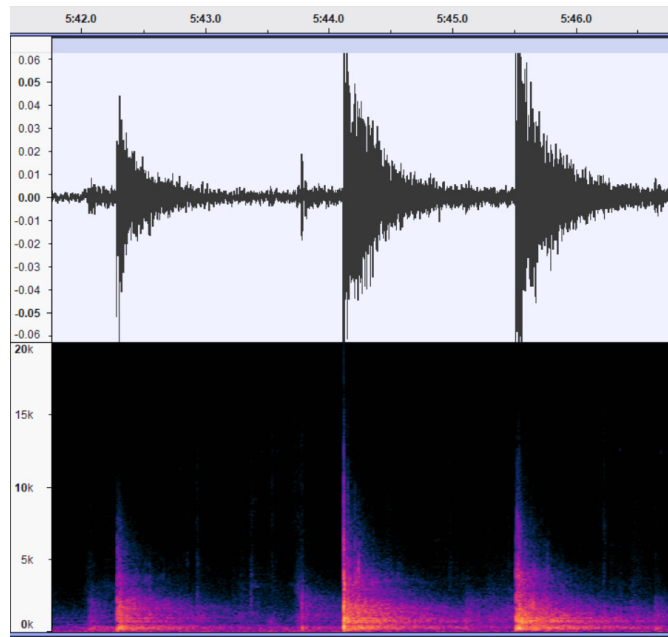


Figure A.5: Example of closely spaced airborne events during a rally, with repeated racket impacts followed by ground rebounds

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
École polytechnique de Louvain

Rue Archimède, 1 bte L6.11.01, 1348 Louvain-la-Neuve, Belgique | www.uclouvain.be/epl