Development of a Serious Game to Improve Decision-making Skills of Martial Arts Referees

André Salmhofer, Lucian Gutica-Florescu, Dominik Hoelbling[a], Roland Breiteneder, Rene Baranyi[b] and Thomas Grechenig

Research Group for Industrial Software (INSO), Vienna University of Technology, Vienna, Austria

Keywords: Decision-making Training, Serious Game, Digital Game-based Learning, Referees, Judges, Martial Arts.

Abstract: While sports referees need to cover a wide spectrum of demands depending on the characteristics of the judged sport, the outstanding responsibility they are associated with, is the task of decision-making. The focus of martial arts referees lies in perception and cognitive processing to detect, categorize and evaluate fast-moving techniques performed within a short period. To accumulate the training intensity required to reach expert level, recent research suggests complementing competitive experience with a video-based training approach. By combining the benefits of video-based training with motivational game elements, the study aimed to develop a video-based serious game to train intuitive decision-making processes of martial arts referees through immediate feedback. The training platform called *JudgED* comprises two modules: (a) a serious game to train decision-making processes and (b) a content and administration interface to manage, prepare, annotate and augment the video content used in the serious game. To evaluate the effectiveness of the serious game, a method to measure the players' decision accuracy and reaction time is proposed.

1 INTRODUCTION

While referees need to cover a wide spectrum of skills encompassing perception, physical fitness, and interaction with athletes, the characteristic responsibility referees are connoted with, is the responsibility of decision-making (MacMahon and Strauß, 2014). As athletes in martial arts can perform a sequence of fast-moving techniques within a short period, referees are required to derive appropriate decisions from memory, by combining their perception of the athletes' movement with their prior experience and the rules of the sport (Carlsson et al., 2020).

Investigating the task of decision-making in detail discloses a complex social-cognitive process influenced by various external constraints specific to the officiated sport (Kittel et al., 2021). To cope with the complexity of this task, referees need to combine declarative knowledge covering the rules of the sport and procedural knowledge acquired by practical experience (Mascarenhas et al., 2006). If referees are not trained appropriately, the complexity of this process can cause decision errors having the potential to influence the outcome of competitions or tournaments (MacMahon and Strauß, 2014).

1.1 Decision-making Process

Decisions are derived by traversing a sequence of social information processing steps comprising perception, categorization, memory processing, and information integration (Bless et al., 2004). While all four steps are essential to derive a proper decision, the emphasis of each step depends on the characteristics of the judged situation (Plessner and Haar, 2006).

Schweizer et al. outline the importance of the categorization step for judging foul/no-foul situations in soccer (Schweizer et al., 2011). By referring to Brunswik's Lens model (Brunswik, 1952), they claim that categorizations are influenced by multiple cues, where only relevant cues are contributing to the accuracy of the decision. To integrate cues and derive decisions under high time pressure, intuitive processing is applied rather than deliberate processing.

1.2 Lack of Decision-making Training

The impact on competition outcomes and its associated economic consequences led to an increased in-

[a] https://orcid.org/0000-0001-7099-2576
[b] https://orcid.org/0000-0002-0088-9140

Salmhofer, A., Gutica-Florescu, L., Hoelbling, D., Breiteneder, R., Baranyi, R. and Grechenig, T.
Development of a Serious Game to Improve Decision-making Skills of Martial Arts Referees.
DOI: 10.5220/0011382800003321
In *Proceedings of the 10th International Conference on Sport Sciences Research and Technology Support (icSPORTS 2022)*, pages 29-40
ISBN: 978-989-758-610-1; ISSN: 2184-3201

vestigation of referees' decisions (Kittel et al., 2021; Larkin et al., 2011). While the literature specifies a variety of approaches to develop decision-making skills of referees, direct participation in sports competitions is acknowledged to be an ideal method to acquire these skills (MacMahon et al., 2007). Based on skill development frameworks like the 10,000-hour rule of deliberate practice (Ericsson et al., 1993), solely relying on on-field experience might not accumulate enough training intensity to reach expert level in decision-making (Larkin et al., 2018).

1.3 Video-based Training Approach

A potential solution to compensate for the lack of training time caused by the limited number of competitive events is the application of video-based training programs (Kittel et al., 2021). These allow the accumulation of practical training intensity, which would hardly be achievable by solely judging real-life competitions (Larkin et al., 2018). The trend in research towards the development and evaluation of well-grounded video-based decision-making training programs emerged over the past 17 years (Kittel et al., 2021).

Recent research confirms the effectiveness of video-based training platforms for referees in various sports (Mascarenhas et al., 2005; Schweizer et al., 2011; Put et al., 2016; Larkin et al., 2018). Although no training platform is available to improve decision-making skills of martial arts referees, the positive effects of video-based training approaches might be transferable to martial arts refereeing as well.

1.4 Serious Games

While several definitions of the term *serious game* exist, Michael and Chen (2005) describe it as games, whose primary goal is education rather than entertainment. The serious game developed in this work can be classified in the sub-category of digital game-based learning, which aims to foster knowledge and skills by utilizing challenges and associated achievements (Qian and Clark, 2016).

1.5 Design Considerations

The design of the serious game was based on the decision-making framework described above and the implications drawn by Schweizer et al. (2011) and Brand et al. (2009) to train intuitive decision-making processes of soccer referees by the principles of Hogarth's learning approach (Hogarth, 2008). This suggests that intuitions can be trained in representative environments by providing relevant and immediate feedback. Assuming the similarities to foul/no-foul judgments in soccer, these theoretical considerations might apply to decision-making in martial arts refereeing as well.

The scope of this study is to (a) design and develop a serious game to train intuitive decision-making processes of martial arts referees by enabling the judgment of numerous representative fight videos and providing immediate feedback, (b) ensure the precise recording of user inputs in accordance with the progress of the streaming video, (c) define a procedure to measure the referees' in-game performance, and (d) propose a setup for evaluating the effectiveness of the serious game.

2 METHODS

The serious game was designed and developed according to the method of prototyping, which allowed to produce artifacts demonstrating relevant aspects of the target system in early phases of the software development life cycle (Floyd, 1984). Initially gathered requirements were refined by conducting two iterations of exploratory prototyping based on mock-ups of the user interface. Subsequently, the system was developed in multiple iterations including the activities of requirements engineering, design, implementation, test, and deployment. Throughout all iterations, feedback was gathered from domain experts comprising two former professional athletes in kickboxing and karate kumite, as well as seven officially licensed kickboxing referees. Depending on the maturity of the developed system, reviewed artifacts comprised low-fidelity or high-fidelity prototypes.

Requirements Engineering: Requirements were gathered by conducting semi-structured interviews (Adams, 2015) of domain experts in kickboxing and karate kumite. While initial requirements were retrieved at the beginning of the project, the list of requirements was gradually extended and refined based on feedback retrieved during the iterative development cycles.

Design and Implementation: The frontend of the prototype was developed by exerting component-based software engineering (Xia Cai et al., 2000). The backend was developed following a resource-oriented architecture (Overdick, 2007). Endpoints were designed according to principles of API composition and API aggregation (Baldini et al., 2017), which allowed to keep the client-side code slim by hiding the complexity in the backend. Particularly, the system's architecture was achieved by applying the MERN stack (Subramanian, 2019) comprising the technologies MongoDB, Express, React, and NodeJS. The video content platform Vimeo[1] was used to store and stream the training videos.

Test: The developed features were manually tested by applying a combination of black box and white box tests (Jamil et al., 2016). While white-box testing was used to verify the correctness of self-written code, black-box testing was used to cross-check features developed by other team members. Thus, every feature went through two internal test stages before it was deployed to gather feedback from domain experts, who accepted or rejected the developed features.

Deployment: The break-down of requirements into small tasks resulted in short lead times, which supported an incremental development process (Petersen, 2010). This allowed performing frequent deployments by using an automatized CI/CD pipeline (Shahin et al., 2017), which enabled the collection of early and recurring feedback from domain experts.

3 RESULTS

This section presents the artifacts and procedures contributing to the development and evaluation of the training platform *JudgED*. After enumerating high-level requirements and the building blocks of the system, the functionality of the training platform is discussed. Subsequently, a mechanism to precisely capture and calculate performance data is described, before an evaluation approach is proposed.

3.1 Requirements List

The requirements engineering process resulted in a set of functional and non-functional requirements. Table 1 enumerates the identified high-level requirements, which are classified in the categories of content and administration (C) and serious game (G). The sections 3.4 and 3.5 describe the functionality of the training platform by referring to the related requirements.

[1] https://vimeo.com/

Table 1: High-level requirements classified in content and administration (C_i) and serious game (G_i).

ID	Description
C_1	Upload videos
C_2	Define and annotate video scenes
C_3	Compile video scenes in playlists
C_4	Configure feedback and playback modes
C_5	Release playlists for players
C_6	Video scene status management
C_7	Performance monitoring dashboard
C_8	Statistical performance evaluation
G_1	Assessment of video scenes
G_2	Immediate feedback and slow-motion
G_3	Personal performance dashboard
G_4	User performance comparison

3.2 Main Modules

The training platform comprises two modules: (a) the serious game used by referees to improve their decision-making skills and (b) a content and administration interface enabling experienced referees to prepare, manage and annotate the training videos used in the game. While the content and administration module is only provided as a web application, the serious game is additionally accessible by an Android app.

3.3 Entity Structure

To structure and prepare the content for the players, the training platform comprises the main entities videos, video scenes, playlists, courses, and users. The training material is based upon uploaded *videos* of fight scenes. Due to the reason that many video files include an entire bout, the footage can be sliced into multiple short *video scenes* corresponding to fight sequences to be judged by the users. To arrange video scenes according to didactic requirements, they are compiled in form of *playlists*. To release specific playlists to a certain group of referees, the *course* entity combines playlists and *users*. Users assigned to a course can access all playlists included in the course for a defined period. Figure 1 visualizes the involved entities and their relationships with each other.

3.4 Content & Administration Module

The content and administration module includes functions to prepare and organize the video scenes used as training material in the serious game. It provides functions to upload videos, define video scenes, compose playlists, and create courses. The subsequent sections describe the functionalities in the context of the involved entities.

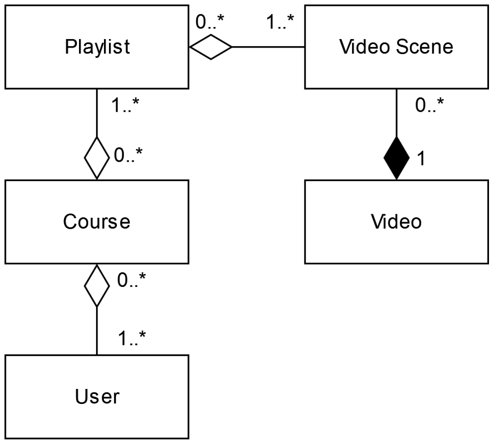

Figure 1: Relation between entities of the training platform.

3.4.1 Videos

Implemented Requirements: C_1

A video corresponds to raw footage containing fight situations involving two athletes in a certain discipline. The system provides the functionality to upload videos along with basic metadata. While most of the metadata serves for identification purposes, the fields *association* and *discipline* determine constraints applicable to the extractable video scenes. Figure 2 shows the video upload screen including the upload area and the metadata fields.

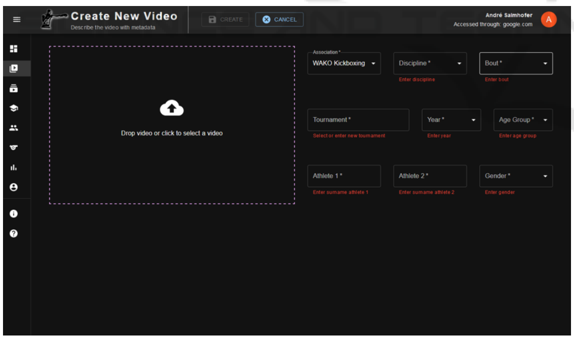

Figure 2: Upload of video including descriptive metadata.

3.4.2 Video Scenes

Implemented Requirements: C_2, C_6

Once a video is uploaded to the system, multiple video scenes can be extracted. A video scene is a partition of an already uploaded video, annotated with a list of decisions appearing in the defined time range. It corresponds to a fight situation that can be judged by the players of the serious game. Video scenes serve (i) as the basis to render the video content in the serious game and (ii) as a reference to determine the correctness of player inputs for presenting feedback and enabling statistical evaluations.

Table 2: Constraints in terms of admissible duration, decision values, and number of decisions (#) for defining video scenes in kickboxing (KB) and karate (KT) disciplines.

Discipline	Length	Decision	#
KB Point fighting	4-15 s	0-3, W/E	1, 2
KB Light contact	45-90 s	0-3, W/E	1+
KB Kick light	45-90 s	0-3, W/E	1+
KB Full contact	45-90 s	0, 1, W	1+
KB Low kick	45-90 s	0, 1, W	1+
KB K1 Style	45-90 s	0, 1, W	1+
KT Kumite	4-15 s	0-3, W/E	1, 2

Structure and Constraints: The allowed configurations for duration and decisions of a video scene are determined by the discipline inherited from the uploaded video. While video scenes of *Point Stop* disciplines can include one decision and an optional concurrent decision, video scenes of *Running Time* disciplines can contain arbitrarily many defined decisions. Table 2 summarizes the admissible duration of video scenes as well as the number of allowed decisions and their spectrum of accepted decision values by discipline. An exceptional case is posed by decisions defined with value 0, which are used to mark sensitive situations for which no referee input is expected. For simplification reasons, no distinction is made between warning and exit decisions (W and E).

Time Range & Decisions: Figure 3 depicts the screen where video scenes can be defined. The time range of the video scene is configured by defining the start and end time in the context of the uploaded video (1). The process of defining a decision is triggered by watching the video and searching for the approximate point in time of the occurring event. To easier scan the video scene, the speed can be toggled between normal mode and a 30 percent slow-motion (2). To precisely seek the exact frame of the occurring decision, the frame-by-frame function forwards/rewinds the video in steps of 0.02 seconds (3). New decisions are added by automatically adopting the point in time of the identified video frame (4). Each decision consists of the exact point in time of the decision (5), the technique determining the decision score or penalty (7), and the athlete (red or blue) to which the decision is attributed (6). The first and the last second cannot be utilized to define decisions, which allows players of the serious game a reaction time of three seconds by appending a trailing period of two seconds at the end of each video scene. The position of the red and blue athletes is configurable (8) based on the athletes' position at the start of the video scene. This setting provides the basis to assign decisions to the respective athlete and arranges the position of the red and blue scoreboard shown to the player in the serious game.

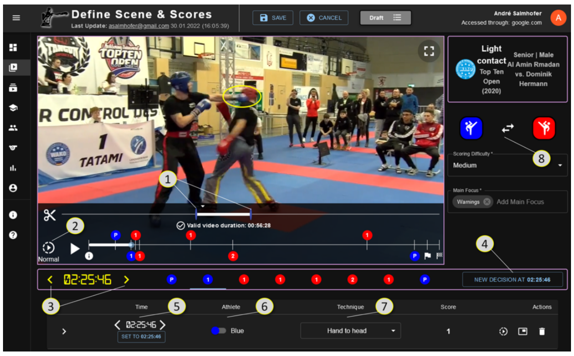

Figure 3: Definition of video scene and occurring decisions.

Highlighting: Apart from the basic decision annotations, information-rich areas can be highlighted for each defined decision separately. Figure 4 shows the highlighting drawing screen for a selected decision within the video scene. A highlighting can consist of multiple ellipses added in the context of 0.5 seconds of the defined decision (1). Within this time frame, the visibility period of the highlighting can be adjusted according to the characteristics of the situation (2). Each ellipse can be positioned and resized to emphasize important techniques (3). The highlighting is displayed as an overlay in the slow-motion feedback presented to the player in the serious game after judging the video scene, which aims to increase the player's understanding of the decision.

Figure 4: Definition of highlighting for a defined decision.

Preview: To review the correct timing as well as the configuration of the optional highlighting, a preview function is available. The preview shows a 30 percent slow-motion 0.5 s before and 0.5 s after the defined decision, which corresponds to the slow-motion feedback shown to the player.

Blurring: The usage of footage from real-world competitions poses a problem, as gesticulating referees might be visible in the video. To not influence the players in the serious game, referees can be covered by adding multiple blurring rectangles (1) for the visibility period of the gesture (2) as shown in Figure 5. Each rectangle can be positioned and resized to ensure the referee's gesture is covered (3). While the blurring rectangle is semi-transparent in the definition screen to ease configuration, it is displayed opaquely for the player of the serious game.

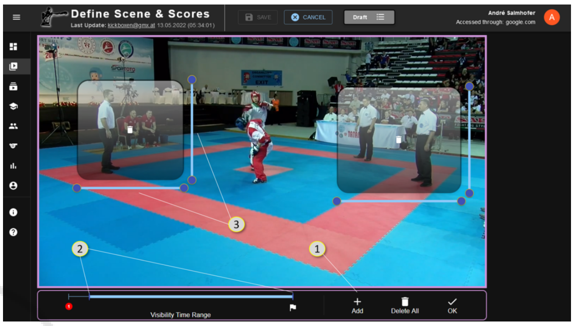

Figure 5: Covering decision-revealing gestures of referees.

Status Management: The training platform has a simple status management for keeping track of the quality of video scenes. A created video scene is initially in status *DRAFT* until it is reviewed by another user with proper permissions, who can change the status to either *APPROVED* or *REJECTED*. Changing relevant fields of a video scene automatically resets the status to *DRAFT*. Figure 6 shows the current status (1), the functions to approve (2) or reject (3) the video scene as well as the status history (4).

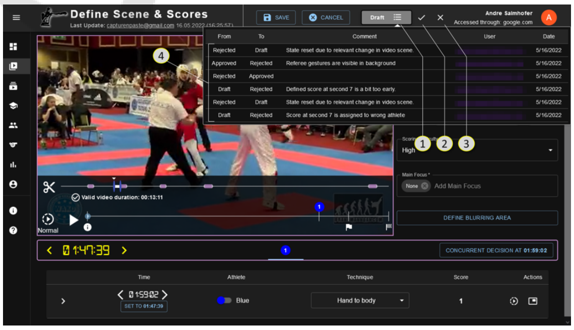

Figure 6: Status management of a video scene.

3.4.3 Playlists

Implemented Requirements: C_3, C_4

Playlists serve as a container to compile multiple video scenes according to given didactic or organizational requirements. Figure 7 shows the screen where playlists can be created by dragging and dropping video scenes. A playlist can be configured with respect to allowed repetitions, playback order, and the extent of displayed feedback. These characteristics are determined by the playlist's mode which can take the values *regular* (1), *lab* (2) or *exam* (3).

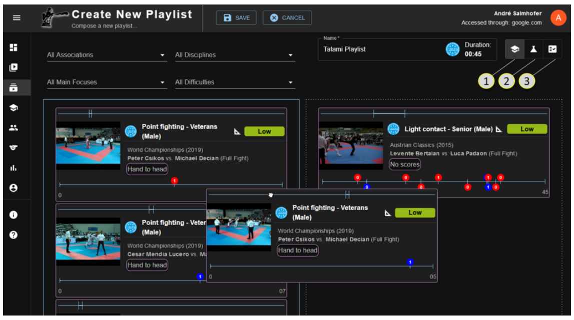

Figure 7: Playlist creation by drag and drop of video scenes.

Regular Playlists: Regular playlists can be played arbitrarily often, whereas the included video scenes appear in random order leading to a non-uniform distribution of judged video scenes. The full extent of feedback is shown after each judged video scene and the slow-motion replay can be repeatedly watched by the player. This kind of playlist is intended to be used for regular training sessions in non-scientific setups.

Lab Playlists: Similar to regular playlists, lab playlists can be played arbitrarily often, and included video scenes appear in random order. However, repetitions only appear as soon as all video scenes in the playlist were played through, leading to a uniform distribution. Feedback is shown after each judged video scene, but the slow-motion replay is not repeatable (i.e. non-repetitive). This kind of playlist is intended to be used for intervention periods in field experiments.

Exam Playlists: Video scenes included in exam playlists are presented in the defined sequence. Each video scene in the playlist can only be judged once. Neither feedback nor a slow-motion replay is provided after the judged video scene. This playlist is intended to be used for pre-, post-, and retention-tests in field experiments.

3.4.4 Courses

Implemented Requirements: C_5

Courses serve as organizational units to release selected playlists (3) to a certain audience (2) for a defined period (1). Course participants (i.e. players) can access all playlists included in the course for the defined period. Figure 8 shows the screen to configure a course.

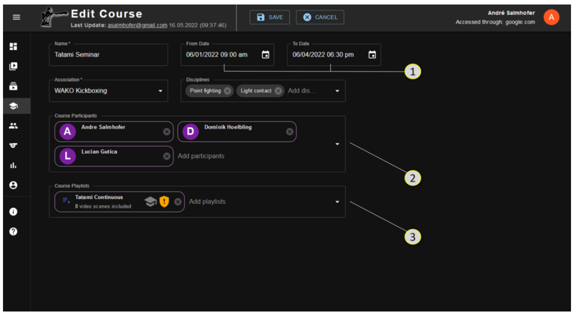

Figure 8: Course definition including playlists and players.

3.4.5 Dashboard

Implemented Requirements: C_7

Depending on the role of the administrative user, the dashboard contains slightly different widgets. While course organizers see statistics and charts restricted to their administered courses, administrators are able to see statistics of all players in the system. Figure 9 shows the dashboard of the administrator role displaying the average decision accuracy (1), reaction time (2), and training intensity (3) overall and for each discipline separately. The development of these metrics over time is visualized by a line chart (4). To detect problematic video scenes, a list of worst-rated video scenes concerning players' decision accuracy and reaction time is shown (5). In addition, the number of challenged video scenes (6) and the number of video scenes that were rejected during the quality review process (7) are displayed.

Figure 9: Administrator dashboard with performance data.

3.4.6 Statistical Performance Evaluation

Implemented Requirements: C_8

To allow the evaluation of the players' performance in scientific or course settings, the training platform provides functions to generate charts for decision accuracy and reaction time on various aggregation levels. Figure 10 shows a chart for the metric reaction time (1) aggregated by discipline (2). Statistics can be generated as bar charts (3) or as line charts (4) showing the development of the selected metric over time. To refine the charts, various filter criteria can be applied (5).

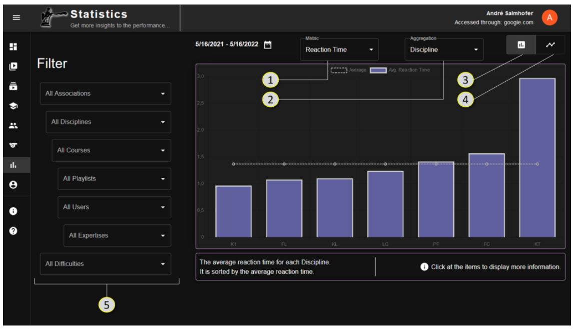

Figure 10: Chart showing the reaction time by discipline.

3.5 Serious Game Module

Based on the playlists prepared in the administration and content module, the serious game provides the actual functionalities aiming to improve the decision-making skills of martial arts referees. The serious game is additionally provided as an Android mobile application optimized for ten-inch tablets. Making judgments on a touchscreen might reduce the time between the detection of the decision and the actual user input, as the mouse cursor does not need to be moved to the respective button on the scoreboard.

The subsequent sections provide more insights into the mechanics of the serious game.

3.5.1 Serious Game Mechanics

Players in the serious game are confronted with a series of fight situations in the form of video scenes. By using a discipline-specific scoreboard, the task of the user is to judge the occurring events in real-time as accurate and fast as possible. After each video scene, the user receives feedback on the correctness of their decisions. To increase the users' motivation to train with the serious game, personal statistics, rankings, and comparisons with other players are available.

3.5.2 Judge Scene

Implemented Requirements: G_1

Training sessions are initiated by selecting an available playlist, which redirects the player to an included video scene. Figure 11 shows the progressing video scene, for which the player is requested to judge occurring events in real-time. By using a discipline-specific scoreboard (1 and 2), decisions are attributed to either the blue (3) or the red athlete (4). Depending on the configuration of the video scene, potentially revealing referee gestures are blurred.

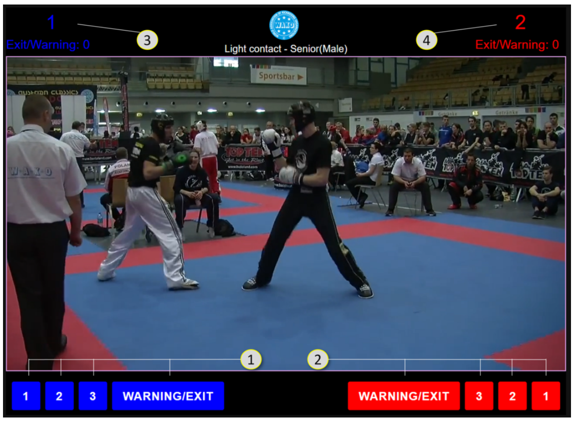

Figure 11: Video scene to be judged by the player.

3.5.3 Immediate Feedback

Implemented Requirements: G_2

At the end of each video scene, feedback about the judgment(s) is presented based on the comparison of the player's inputs with the defined decisions in the video scene (see Figure 12). For each decision defined in the video scene (1) a 30 percent slow-motion sequence starting 0.5 seconds before and ending 0.5 seconds after the time of the defined decision is shown to the player. In addition, the feedback comprises the player's decision (2), the correct decision (3), the reaction time (4), the correctness indication (5), and the applied technique (6).

To increase the player's understanding of the revealed decision, the slow-motion sequence optionally highlights information-rich areas relevant for detecting the cause of the respective decision (7). In case of disagreement with the expert-defined decisions, the player can challenge them by entering a comment (8).

The review ends with a summary of the performance of the judged video scene as depicted in Figure 13. The summary shows the overall decision accuracy for the video scene as well as the visualization of the defined decisions and player decisions on a timeline (1 and 2). In addition, it also provides feedback about redundant decisions (3), which were not presented in the detailed decision-specific feedback before.

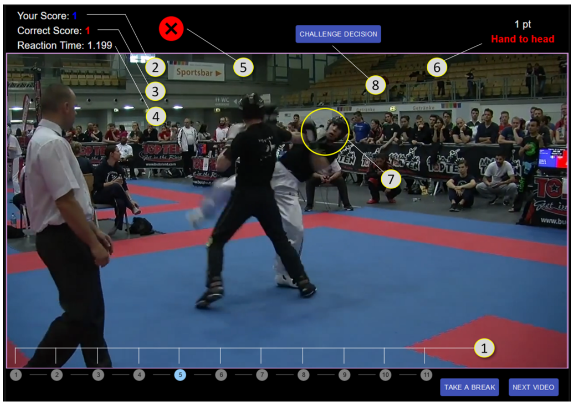

Figure 12: Feedback about the player's decision(s) after judging the video scene.

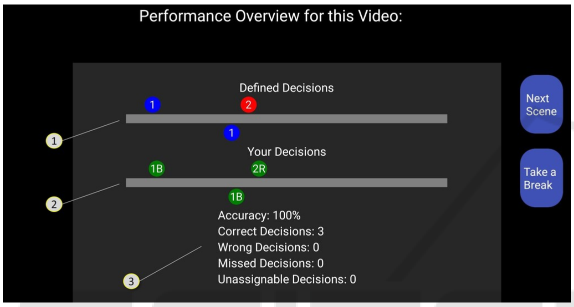

Figure 13: Holistic feedback about the judged video scene.

3.5.4 Dashboard & Game Elements

Implemented Requirements: G_3, G_4

The purpose of the dashboard shown in Figure 14 is to compactly inform the players about their judgment performance and to increase their motivation to train with the serious game.

Performance Elements: Performance data is provided in terms of average decision accuracy (1), reaction time (2), and training intensity (3) overall and for each discipline separately.

Competitive Elements: To increase the motivation of the players and to enable the comparison with other players in the serious game, the dashboard displays a leader board (5), the player's rank (6), and the average performance data of all players (7). While the leader board shows the performance data of the best performing referee, the own rank indicates the rank of the currently signed-in player. Both leader board and own rank are calculated according to the metric of decision accuracy. To avoid a high fluctuation in the elements of the leader board and rank, players with less than 50 decisions are excluded, as too little performance data is recorded to calculate a reasonable performance indication.

Replay: The dashboard enumerates all video scenes for which no single decision was judged correctly by the signed-in player (4). These video scenes are the only ones, which can be selectively replayed. They disappear from the list as soon as they reach a decision accuracy greater than zero percent.

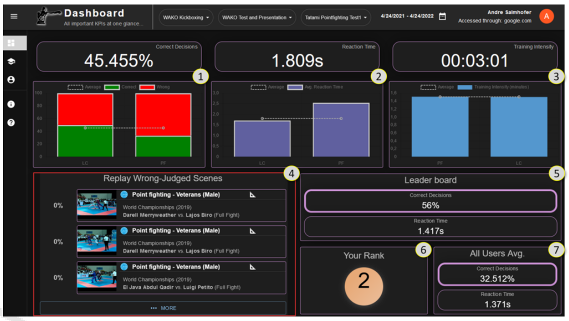

Figure 14: Game dashboard including personal performance data and comparison with other players.

3.6 Precise Data Recording

While the player judges the video scenes according to the occurring events, the inputs are logged in the system, which allows to create statistics and draw conclusions about the player's performance. Particularly, the following data is logged for each judged video scene: (i) user ID, (ii) video scene ID, (iii) playlist ID, (iv) course ID, (v) date and time of the judgment, and (vi) a list of player decisions comprising input time, decision and athlete.

This information provides the basis to calculate results concerning the correctness and reaction time of decisions. To obtain accurate results, the player decisions must be precisely recorded in accordance with the progress of the streaming video. This is ensured by the provided functionalities of the client-side player library used to render and interact with the video stream. There exist two basic approaches to retrieve the current progress of the the video stream: (i) by listening for *timeupdate* events triggered every 250 ms or (ii) by calling the function *getCurrentTime()* on demand.

Depending on the capabilities of the client-side library, either the first or the second version is used to determine the progress of the streaming video. While the web version of the serious game uses the second method to precisely determine the current progress of the video stream whenever the player makes a decision, the mobile version relies on the first version, which comes with a maximum inaccuracy of 250 ms. This allows the general statement that the serious game can record the point in time of the player's decisions in accordance with the progress of the streaming video by tolerating a maximum deviation of 250 ms.

3.7 Performance Metrics & Calculation

As a prerequisite to providing proper feedback to the user as well as to statistically evaluate the players' performance data, the correctness, as well as the reaction time of the players' decisions, is determined. While decision accuracy is defined as the ratio between correct and incorrect decisions, the reaction time of a decision is the time difference between the player decision and the defined decision in the respective video scene.

To calculate these metrics, the player decisions are compared to the defined decisions of the judged video scene. In case the video scene contains multiple decisions, this poses a complex task, as it is not always unambiguous which player input was meant for which defined decision. To perform this correlation, an algorithm was developed, which matches player decisions to defined decisions based on a defined set of rules.

3.7.1 Definitions

To describe the functionality of the correlation process, basic terms used throughout the algorithm need to be defined in advance.

Defined Decisions $\mathcal{D}$: A list consisting of all decisions included in the video scene defined by an expert referee. Each defined decision $\mathcal{D}_i$ is identified by the properties *time*, *athlete* and *value*.

Player Decisions $\mathcal{P}$: A list consisting of decisions inputted by the player of the serious game while watching the video scene. Each player decision $\mathcal{P}_j$ is identified by the properties *time*, *athlete* and *value*.

Matching $\mathcal{M}$: A matching is a tuple correlating a player decision $\mathcal{P}_i$ to a defined decision $\mathcal{D}_j$. It also contains the reaction time as well as the correctness of the decision.

Unassignable Decision $\mathcal{P}_u$: A list of elements consisting of a subset of player decisions, which were not assignable to any defined decision.

Missed Decision $\mathcal{D}_u$: A list of elements consisting of a subset of defined decisions, which were not assignable to any player decision.

Maximum Decision Time $\mathcal{T}_{max}$: The maximum allowed time difference between player decision and defined decision (defined as three seconds). Player decisions exceeding this time are not considered as matching candidates.

3.7.2 Consideration for Choosing $\mathcal{T}_{max}$

To consider a player's decision as correct, it needs to be performed within the maximum decision time of three seconds from the occurring event in the video scene. The comparable study in the sport of soccer conducted by Schweizer et al. (2011) used a time range of five seconds. Compared to the referred study, where the participants used a mouse as an input device, the current work's test setup proposes to use the mobile version of the serious game. In this case, the player decisions are indicated by taping on the respective scoreboard element on the touchscreen. Thus, a maximum decision time of three seconds was considered sufficient for the serious game in this work.

3.7.3 Correctness Evaluation

Player decisions are only considered to be correct in case (i) they are judged within the *Maximum Decision Time*, (ii) the player decisions are made in the same order as the defined decisions, and (iii) the athlete and decision value are matching.

A necessary condition to evaluate a player decision as correct is its occurrence in the list of matchings $\mathcal{M}$. Decisions included in this list already fulfill the conditions (i) and (ii) mentioned above. Thus, the first step is to generate the list of matchings $\mathcal{M}$ according to the process described in the subsequent section.

3.7.4 Decision Matching

1. Basic Matching: The list of matched decisions $\mathcal{M}$ is generated by comparing player decisions in $\mathcal{P}$ with defined decisions in $\mathcal{D}$. Player decisions are attempted to be matched with defined decisions fulfilling the correctness condition, whereas the search space for eligible player decisions is limited by the *Maximum Decision Time* constraint. In case no matching candidate fulfilling the correctness condition is found, the defined decision is correlated to the closest player decision. Already matched player decisions are not eligible candidates for further correlations.

A special situation is represented by defined decisions with the value *zero*, which demand no explicit player input to be correct. In case the defined decision is defined as zero and no assignable player inputs are found, the defined decision is correlated with a synthesized player decision constructed by adopting the time, athlete, and value properties of the defined decision.

2. Conflicts Detection: Solely performing the basic matching described above, might lead to cases where the correct order condition of decisions is violated, which poses a conflict. A conflict is reflected by a constellation in $\mathcal{M}$, in which a player decision's time parameter is smaller than the player decision's time parameter corresponding to one of its preceding defined decisions.

By illustrating the relations between defined decisions and player decisions on a time scale, a conflict can be visually imagined by an intersection of connections representing matchings. Figure 15 shows an example posing a conflict between the matching $\mathcal{M}_{1,2}$ and $\mathcal{M}_{2,1}$. Formally, two matchings $\mathcal{M}_{ab}$ and $\mathcal{M}_{xy}$ are involved in a conflict, in case the conditions Equation 1 to Equation 5 are met. Where $\mathcal{M}_{ij}$ represents the matching of a defined decision at index i in $\mathcal{D}$ with a player decision at index j in $\mathcal{P}$.

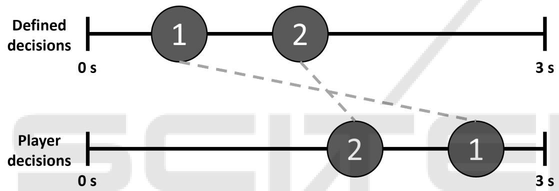

Figure 15: Conflict indicated by intersection of matchings.

As already considered by Equation 5 in the formal definition of a conflict, a special case poses the situation in which multiple defined decisions are closely spaced within a specific period (i.e. within *delta time* defined as 0.3 seconds). For a consecutive sequence of defined decisions that is within delta time, the violation of the order does not cause a conflict. This exception was introduced as the insistence on the judgment order in such cases might be too strict.

$$P[b].time - D[y].time < maxtime \tag{1}$$

$$P[y].time - D[a].time < maxtime \tag{2}$$

$$D[a].time < D[x].time \tag{3}$$

$$P[b].time > D[y].time \tag{4}$$

$$D[x].time - D[a].time > deltatime \tag{5}$$

3. Conflict Resolution: To maintain the order of decisions condition, identified conflicts in $\mathcal{M}$ need to be removed. A conflict is resolved by inspecting the matchings involved in the conflict and deciding which one to keep and which one to refuse. Besides removing the refused matching from $\mathcal{M}$, its defined decision is added to the list of missed decisions $\mathcal{D}_u$ and its player decision to the list of unassignable decisions $\mathcal{P}_u$. The process of conflict resolution is applied to all conflicts until the matching list is conflict-free.

The criteria to decide which conflicting matching to keep and which one to refuse is determined according to the level of assumed obviousness of the decisions involved in the matching. A defined decision having a higher value is considered more obvious and thus more likely to be correctly spotted by the player while assessing the video scene. For conflicts involving defined decisions of equal score values, the matching involving an earlier defined decision is kept. Equation 6 and Equation 7 showcase the hierarchy of priorities for kickboxing disciplines and karate kumite respectively. While the numbers in the expressions correspond to the score values of a defined decisions, *C1* (category 1), *C2* (category 2), and *Warning/Exit* represent penalty classes of the respective sport.

$$3 > Warning/Exit > 2 > 1 \tag{6}$$

$$3 > C2 > C1 > 2 > 1 \tag{7}$$

4. Outcome: The main outcome of the matching algorithm is the set of matched decisions $\mathcal{M}$, where each player decision is correlated to a defined decision, according to the defined time and order maintenance constraints. In addition, the sets of missed decisions $\mathcal{D}_u$ and unassignable decisions $\mathcal{P}_u$ is emerging from this algorithm. While all decisions in $\mathcal{D}_u$ and $\mathcal{P}_u$ have no reaction time and are incorrect by default, the elements in $\mathcal{M}$ contain information about their correctness and reaction time.

3.8 Proposed Effectiveness Evaluation

The serious game is proposed to be scrutinized in terms of efficacy and motivation. To achieve this, a two-tiered approach consisting of a field experiment and a questionnaire is suggested.

3.8.1 Performance Evaluation

To evaluate the effectiveness of the serious game with regard to its ability to improve the decision-making processes of martial arts referees, the conduction of a field experiment in form of a *pretest-posttest control group design* (Crano et al., 2014) is proposed. To test the development of the training effects over time, a retention test conducted three weeks after the intervention is suggested. To allow participants to familiarize themselves with the mechanics of the serious game, a short familiarization phase for both the control and intervention group before the conduction of the pre-test is recommended.

The recommendations of All et al. (2021) concerning group assignment and test design should be considered in order to increase internal validity and avoid pre-test effects. Accordingly, the assignment of participants into intervention and control groups should be performed by blocked randomization with respect to experience. Pre-, post-, and retention tests should be created in parallel versions, whereas the comparability is enabled by the same average difficulty of the tests. The average difficulty is determined by expert referees, not participating in the experiment, based on the rating of each video scene on a five-point scale covering values from *very low* to *very high*.

All phases of the field experiment can be exclusively performed in the serious game by compiling separate playlists. While pre-, post- and retention-tests are performed without feedback (i.e. playlist mode *Exam*), the intervention period is proposed to be conducted with immediate, non-repetitive feedback (i.e. playlist mode *Lab*). It is proposed to perform all phases of the experiment on a 10-inch tablet with a touch screen.

3.8.2 Motivation Evaluation

The quality of a serious game is also determined by its ability to intrinsically motivate players as a prerequisite to achieve the desired learning outcome (All et al., 2014). Therefore, the conduction of a post-experimental questionnaire is proposed by utilizing questions from the Intrinsic Motivation Inventory scale[2].

3.9 Summary of Outcomes

The presented training platform comprises a serious game module to train referees' decision-making skills as well as a content and administration module to prepare and organize training sessions. After providing a method to precisely capture the player decisions in accordance with the progress of the streaming video, a procedure to determine the accuracy and reaction time of decisions was introduced. To evaluate the effectiveness of the serious game in future studies, an evaluation setup was proposed by taking both performance and motivational aspects into account.

3.10 Limitations

The video scenes used in the serious game only partly reflect the constraints occurring in real-life competitions such as perspective, crowd noise, and sources of stress. The utilization of first-person videos or the application of virtual reality might contribute to a more representative training approach leading to a higher ecological validity by better incorporating perceptual information appearing in real-life competitions (Kittel et al., 2021).

The serious game relies on a sufficiently high internet bandwidth to stream the video scenes to be judged. Deviating internet quality might affect the rendering of videos and thus decrease the comparability of performance data among referees.

4 CONCLUSION

As evidenced by referee training programs in other sports (Schweizer et al., 2011; Mascarenhas et al., 2005; Larkin et al., 2018), the complementary application of a video-based training approach has the potential to accumulate practical experience, which would hardly be possible by solely participating in competitive events. By designing a novel training platform for martial arts refereeing according to conclusions from theoretically grounded frameworks, training with the serious game is expected to have the potential to improve the intuitive decision-making processes of martial arts referees in terms of decision accuracy and reaction time. Future studies need to be conducted to evaluate the acceptance, effectiveness, and ability to transfer the gained decision-making skills to real-world competitions.

Apart from its application in scientific studies, the serious game might be used to complement classical educational settings. Due to the integrated administrative functionality to define video scenes and make them available for certain users, training supervisors can selectively tailor the training content and provide the serious game as a practical intervention in seminars and referee education. Especially in pandemic times, the possibility of practically training decision-making skills locally independent might be beneficial.

[2]https://selfdeterminationtheory.org/intrinsic-motivation-inventory/

REFERENCES

Adams, W. C. (2015). Conducting Semi-Structured Interviews. In *Handbook of Practical Program Evaluation*, pages 492–505. John Wiley & Sons, Ltd.

All, A., Castellar, E. N. P., and Looy, J. V. (2021). Digital Game-Based Learning effectiveness assessment: Reflections on study design. *Computers & Education*, 167:104160.

All, A., Nunez Castellar, E. P., and Van Looy, J. (2014). Measuring Effectiveness in Digital Game-Based Learning: A Methodological Review. *International Journal of Serious Games*, 1(2):3–20.

Baldini, I., Castro, P., Chang, K., Cheng, P., Fink, S., Ishakian, V., Mitchell, N., Muthusamy, V., Rabbah, R., Slominski, A., and Suter, P. (2017). Serverless Computing: Current Trends and Open Problems. In *Research Advances in Cloud Computing*, pages 1–20. Springer.

Bless, H., Fiedler, K., and Strack, F. (2004). *Social cognition: How individuals construct social reality.* Psychology Press. Pages: xi, 235.

Brand, R., Plessner, H., and Schweizer, G. (2009). Conceptual considerations about the development of a decision-making training method for expert soccer referees. In *Perspectives on cognition and action in sport*, pages S. 181–190. Hauppauge, NY: Nova Science.

Brunswik, E. (1952). *The conceptual framework of psychology.* Univ. Chicago Press.

Carlsson, T., Berglez, J., Koivisto Persson, S., and Carlsson, M. (2020). The impact of video review in karate kumite during a Premier League competition. *International Journal of Performance Analysis in Sport*, 20(5):846–856.

Crano, W. D., Brewer, M. B., and Lac, A. (2014). Designing Experiments - Variations on Basics. In *Principles and methods of social research*, pages 83–100. Routledge. Section: 5.

Ericsson, K. A., Krampe, R. T., and Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3):363–406.

Floyd, C. (1984). A Systematic Look at Prototyping. In Budde, R., Kuhlenkamp, K., Mathiassen, L., and Züllighoven, H., editors, *Approaches to Prototyping*, pages 1–18. Springer Berlin Heidelberg.

Hogarth, R. M. (2008). On the learning of intuition. In *Intuition in judgment and decision making.*, pages 91–105. Lawrence Erlbaum Associates Publishers.

Jamil, M. A., Arif, M., Abubakar, N. S. A., and Ahmad, A. (2016). Software Testing Techniques: A Literature Review. In *2016 6th International Conference on Information and Communication Technology for The Muslim World (ICT4M)*, pages 177–182. IEEE.

Kittel, A., Cunningham, I., Larkin, P., Hawkey, M., and Rix-Lièvre, G. (2021). Decision-making training in sporting officials: Past, present and future. *Psychology of Sport and Exercise*, 56:102003.

Larkin, P., Berry, J., Dawson, B., and Lay, B. (2011). Perceptual and decision-making skills of Australian football umpires. *International Journal of Performance Analysis in Sport*, 11(3):427–437.

Larkin, P., Mesagno, C., Berry, J., Spittle, M., and Harvey, J. (2018). Video-based training to improve perceptual-cognitive decision-making performance of Australian football umpires. *Journal of sports sciences*, 36(3):239–246.

MacMahon, C., Helsen, W. F., Starkes, J. L., and Weston, M. (2007). Decision-making skills and deliberate practice in elite association football referees. *Journal of Sports Sciences*, 25(1):65–78.

MacMahon, C. and Strauß, B. (2014). The psychology of decision making in sport officals. In *An introduction to sport and exercise psychology*, pages 223–235. London: Routledge.

Mascarenhas, D., O'Hare, D., and Plessner, H. (2006). The psychological and performance demands of association football refereeing. *International Journal of Sport Psychology*, 37:99–120.

Mascarenhas, D. R., Collins, D., Mortimer, P. W., and Morris, B. (2005). Training accurate and coherent decision making in rugby union referees. *The Sport Psychologist*, 19(2):131–147.

Michael, D. R. and Chen, S. L. (2005). *Serious games: Games that educate, train, and inform.* Muska & Lipman/Premier-Trade.

Overdick, H. (2007). The Resource-Oriented Architecture. In *2007 IEEE Congress on Services (Services 2007)*, pages 340–347. IEEE.

Petersen, K. (2010). An Empirical Study of Lead-Times in Incremental and Agile Software Development. In *New Modeling Concepts for Today's Software Processes*, pages 345–356. Springer Berlin Heidelberg.

Plessner, H. and Haar, T. (2006). Sports performance judgments from a social cognitive perspective. *Psychology of Sport and Exercise*, 7(6):555–575.

Put, K., Wagemans, J., Spitz, J., Williams, A. M., and Helsen, W. F. (2016). Using web-based training to enhance perceptual-cognitive skills in complex dynamic offside events. *Journal of Sports Sciences*, 34(2):181–189.

Qian, M. and Clark, K. R. (2016). Game-based Learning and 21st century skills: A review of recent research. *Computers in Human Behavior*, 63:50–58.

Schweizer, G., Plessner, H., Kahlert, D., and Brand, R. (2011). A video-based training method for improving soccer referees' intuitive decision-making skills. *Journal of Applied Sport Psychology*, 23(4):429–442.

Shahin, M., Ali Babar, M., and Zhu, L. (2017). Continuous Integration, Delivery and Deployment: A Systematic Review on Approaches, Tools, Challenges and Practices. *IEEE Access*, 5:3909–3943.

Subramanian, V. (2019). *Pro MERN Stack: Full Stack Web App Development with Mongo, Express, React, and Node*. Apress.

Xia Cai, Lyu, M., Kam-Fai Wong, and Roy Ko (2000). Component-based software engineering: technologies, development frameworks, and quality assurance schemes. In *Proceedings Seventh Asia-Pacific Software Engeering Conference. APSEC 2000*, pages 372–379. IEEE Comput. Soc.