Designing for Automated Sports Commentary Systems

ANONYMOUS AUTHOR(S)*

Advancements in Natural Language Processing (NLP) and Computer Vision (CV) are revolutionizing how we experience sports broadcasting. Traditionally, sports commentary has played a crucial role in enhancing viewer understanding and engagement with live games. Yet, the prospects of automated commentary, especially in light of these technological advancements and their impact on viewers' experience, remain largely unexplored. This paper elaborates upon an innovative automated commentary system that integrates NLP and CV to provide a multimodal experience, combining auditory feedback through text-to-speech and visual cues, known as italicizing, for real-time in-game commentary. The system supports color commentary, which aims to inform the viewer of information surrounding the game by pulling additional content from a database. Moreover, it also supports play-by-play commentary covering in-game developments derived from an event system based on CV. As the system reinvents the role of commentary in sports video, we must consider the design and implications of multimodal artificial commentators. A focused user study with eight participants aimed at understanding the design implications of such multimodal artificial commentators reveals critical insights. Key findings emphasize the importance of language precision, content relevance, and delivery style in automated commentary, underscoring the necessity for personalization to meet diverse viewer preferences. Our results validate the potential value and effectiveness of multimodal feedback and derive design considerations, particularly in personalizing content to revolutionize the role of commentary in sports broadcasts.

CCS Concepts: • **Information systems** → **Multimedia information systems**; • **Human-centered computing** → **Human computer interaction (HCI)**; Information visualization; • **Applied computing** → Media arts.

Additional Key Words and Phrases: Automated Commentary, Embedded Visualizations, Computer Vision, Deep Learning, Natural Language Processing, Human-Computer Interaction

ACM Reference Format:

Anonymous Author(s). 2024. Designing for Automated Sports Commentary Systems. In . ACM, New York, NY, USA, 24 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Sports commentators are essential to immersion and satisfaction while viewing sports broadcasts. They achieve this by drawing viewer attention to current play-by-play events while informing of further color commentary which delivers contextual information regarding the game and players. Their expertise and commentary clarify the rules and strategies at play and add a layer of excitement and emotional connection, making the viewing experience more relatable and memorable. With the emerging popularity of lower-division sports in news media, automating commentary provides an innovative solution to boosting popularity. As evidenced by Lee et al.'s [20] study, commentary increases enjoyment and reviewing intention due to the perceived quality of the broadcast. Furthermore, italicizing, a method of promoting visual information with verbal cues [19, 22, 33], increases the sense of immersion. Considering these factors, there is still interest and a need for automated commentary solutions. While automated commentary saw its early peak in the 90s, the focus has since shifted towards innovation in areas such as computer game commentary and making

Manuscript submitted to ACM

sports broadcasts accessible to users with hearing disabilities. Despite these advancements, a comprehensive solution encompassing both play-by-play and color commentary remained elusive until recently. Bridging this gap, previous work [anonymous] introduced a pioneering methodology for automated non-interactive and interactive commentary. This system not only addressed the need for comprehensive coverage but also enriched the viewer experience by comparing and contrasting a baseline non-interactive commentary system with its interactive counterpart. While the study primarily focused on the interactive version of their system [anonymous], it left room for further exploration into user perception of the non-interactive system for identifying key design considerations for automated sports commentary systems. The focus of our paper is to identify these design considerations to assist future research in automated commentary.

Hence, we explore uncharted research territory by delving into the nuanced audience reception of automated commentary. By bridging this vital research gap on viewer perception of automated commentary, we contribute to the design and evolution of sophisticated automated sports broadcasting systems. In particular, we aim to understand better the innovative integration of italicizing and dual artificial commentators to deliver play-by-play and color commentary. We explore how this integration, along with other design elements, enhances the system's utility and audience engagement. Specifically, our research reveals that the automated commentators' dualistic interplay enhances the commentary's perceived enjoyment. Additionally, commentary content and delivery were significant in clearly communicating engaging play-by-play commentary. Our findings indicate the plausibility and need for personalizing commentary content and visualizations, setting a new standard for commentary systems. By considering the proposed design choices, we contribute to the technological advancement of AI-powered sports broadcasting while providing actionable insights for future developments in this rapidly evolving field.

We structure the paper by first considering the role of commentary and state-of-the-art research in automated commentary and embedded visualizations. Afterwards, we give an in-depth explanation of the automated commentary system and validate its design choices. We then present the user study with quantitative and qualitative findings that lay the basis for the design considerations outlined in our discussion.

2 LITERATURE REVIEW

This section considers the role of sports commentary before exploring previous research in automated commentary and embedded visualization while highlighting our contributions to each research area.

2.1 Sports Commentary

Traditional sports commentary served as a tool to improve the accessibility of sports content while creating a more engaging and immersive experience. Sports commentators traditionally assume two roles: play-by-play and color commentators [22]. The play-by-play commentator reports on in-game developments and strategic elements of the match, whereas the color commentator fills inactive moments of the game with information regarding player performance and popular news stories. Sports commentators add polish to the already engaging viewing experience. Their professional feedback and characteristics can add an additional element of engagement and immersion. The narrative commentators produce over the video stream often aims to provoke an emotional response from the viewer to deepen the sense of engagement [20]. Hence, when commentator biases resonate with those of the viewer, they can intensify emotional engagement, thereby elevating the excitement and perceived stakes of the match. Research has documented how commentary can drastically alter the viewer's perception of the match [12]. For example, Zhou et al.'s [38] research contributes to understanding sports commentary's impact on viewer enjoyment. While they found that neither conflicting nor complimentary commentary significantly enhanced the overall enjoyment of the viewing experience, there was a noted preference for conflicting commentary. This preference was attributed to the entertaining nature of the conflicting commentary, which was characterized as more risky and argumentative, thus engaging viewers more effectively than the complimentary commentary. While language plays a significant role in rendering the narrative, the delivery is equally important in provoking an emotional response from the viewer. Sports commentators often emphasize language by adapting verbal cues such as pace, intonation, and pitch to reflect the current game state. To further assist the viewer's understanding of content, sports analysts often support commentary with visual aids during replays. Supporting visual features with commentary is known as italicizing [19]. We now transition into understanding current state-of-the-art automated commentary systems and embedded visualizations for sports.

2.2 Automated Commentary

Despite previous efforts in developing the field of automated commentary, recent research has been sparse, particularly in modern applications. For this reason, we extend the concept of commentary beyond its traditional scope to encompass audio descriptions and gaming commentary. Automated commentary evolved from play-by-play commentary in the late nineties to more nuanced color commentary in the mid twenty-tens. The latest research trends are steering towards leveraging natural language processing for creating accessible broadcast content and enhancing the gaming experience with dynamic commentary. All research fields aim to solve a common problem, deriving natural language output from a data source that is descriptive of the current medium's context. While our research contributes to contemporary automated football commentary, we tackle the overarching challenge of generating contextually relevant, natural language output from diverse data sources, an objective shared across the research highlighted in this section.

The origins of automated commentary are founded in utilizing attributes from the RoboCup dataset. This dataset provided crucial data like player and ball locations for robotic football, as well as key events such as goal kicks and throw-ins. From this data, three groundbreaking systems emerged, each advancing the domain of play-by-play commentary. MIKE, as detailed by Tanaka-Ishii et al. [27, 28], utilized an event-based analyzer for identifying play-by-play events, complemented by a state-based analyzer grounded in statistical analysis and observations of game dynamics. It generated commentary by filling language templates with event attributes, using a pooling system to manage the pacing of commentary. In a similar vein, ROCCO, developed by Voelz et al. [30], employed a method to populate templates with spatial-temporal data and synthesized speech with inferred emotional cues based on pitch and speed variations. Lastly, Bryne, conceptualized by Binsted et al. [2], distinguished itself with the synthesis of speech and facial animations, adding an emotional layer that varied based on team allegiance.

While these early systems laid the groundwork for automated commentary, the field continued to evolve, embracing more sophisticated technologies and data sources. This progression is exemplified by the work Zheng & Kudenko [37], who proposed using trace data from Championship Manager with a machine learning solution to extract more complex information and events such as assigning roles based on player attributes, decipher possible paths of a pass, and identifying kick intention.

These systems significantly contributed to automated play-by-play commentary but shared a common limitation: their inability to provide in-depth color commentary and their restricted applicability to robotic data and simulated environments, limiting their generalization to broader broadcast video contexts.

While the limitation of providing automated color commentary has not been addressed, other researchers have employed methods of assisting real commentators with relevant color information. For example, Lee et al. [19] introduced a sports commentary recommendation system (SCoReS), which suggested news stories to commentators during inactive periods in baseball. Similarly, Chitrakala et al. [6] recommended articles by ranking relevance to support cricket commentators. To highlight the relevance of an article, Anees [1] extracts play-by-play data via a video processing module. This delivers stories which take into account the context of the current game state, ensuring context-relevant stories.

Other research has considered using real-time broadcast competition data, which are mapped to their respective templates before being synthesized with TTS and delivered as audio descriptions for the visually impaired [15, 17, 18]. Kumano et al. [17] found utterances that reaffirmed previous game states helped increase viewer understanding. In other work, Ichiki et al. [15] sought to overcome overlapping play-by-play commentary with audio descriptions by adjusting sound levels. They found that in doing so, 80% of participants found the description easy to understand. Kurihara et al. [18] implemented such a system in the 2016 Olympic and Paralympic Games, showing users found the system effective. For tennis Goncu et al. [14] developed a binaural audio system to augment 3D audio inferred by ball tracking. However, results from a qualitative and quantitative study were inconclusive regarding the benefit of 3D auditory augmentation over traditional radio broadcast coverage.

Whilst advancements in audio descriptions and binaural systems have marked significant progress in traditional sports broadcasting, the field of automated commentary is also expanding into the realm of eSports. For instance, Wang & Yoshinaga [32] trained an encoder-decoder network on subtitle data to transform data from League of Legends API to natural language. They found a hierarchical encoder overcame data loss from key-value pairs and outperformed the baseline model but suffered from hallucinations and could not replicate humour. Karouzaki & Savidis [16] aimed to generate a social avatar personalized to player profiles to facilitate user understanding of board games. By adapting the sense-think-react strategy to sense-react-think-adapt-react, the model facilitated game-state adjustments to the avatar's emotions and personalized comments depending on the user's progress and social profile.

To summarize, previous research has aimed to adapt a data source into natural language to convey play-by-play narratives, recommend articles to assist color commentators, create audio descriptions for the visually impaired, and convey personalized emotional responses to user input. However, no current system is able to create dynamic play-by-play and color commentary. Furthermore, studies have not considered commentators as a duo, conversing together to deliver more natural-sounding content, nor have they explored users' impressions of automated commentary systems and their contribution to the user experience. These insights are imperative to draw design considerations to inform future research in the field.

2.3 Embedded Visualization

While closely related to visual analytics, embedded visualizations in sports content enrich the viewer's experience by providing contextual and easily digestible data insights directly within the broadcast or digital platform. Current trends in the field have shown the potential of embedded visualizations to enhance data representation, enabling sports analysts to comprehend complex patterns easily.

For example, spatio-temporal analysis of team sport players allows for a better understanding of player behavior [35], team tactics [34, 35], and individual player performance [13, 23, 35]. PassVisor, developed by Xie et al. [35], supported visualization of individual player and collaborative pass patterns in football to gain deeper insights into strategic game elements. On the other hand, Wu et al.'s ForVisor [34] utilized spatio-temporal data to assign roles to football player distributions and map to the corresponding formations. In doing so, ForVisor helped facilitate a deeper understanding of strategic changes in team formations. While Forvisor and PassVisor were concerned with individual matches, SnapShot [23] and CourtVision [13] facilitated the visualization of team sport shots over multiple games. Snapshot plotted cartesian coordinates from 2010-2011 and enabled multiple visualizations, including radial heatmaps filtered by metadata. Similarly, CourtVision visualized basketball shots over five years and determined the two highest-performing players based on spread and range metrics.

Other research has considered motion analysis to assess player performance [26, 36] and assist in visualizations of events [8]. For example, Ye et al.'s ShuttleSpace [36] and Dietrich et al.'s Baseball4D [8] utilize 3D visualization to track badminton and baseball trajectories and track events of interest, respectively. Using Virtual Reality (VR), ShuttleSpace improved the cognitive load of visualizing badminton trajectories by allowing the user to view from the player's perspective, using an extended viewport with supportive 2D data, and implementing an efficient trajectory selection system by mimicking stroke movements with the controller. Baseball4D provided 3D reconstructions of discrete events through time, allowing for in-depth analysis with visual and statistical data. While both Baseball4D and Shuttlespace considered visual exploration of data, Directors Cut [26] developed a rule-based annotation system for football to assist analysts in identifying interaction spaces, free spaces, and pass options.

Further research looks to support sports analysts by integrating tools that automate workflows [4, 5, 25]. Stein et al. [25] proposed a conceptual framework focusing on automatic view selection and explanatory storytelling to enhance understanding of complex football game situations, whereas Chen et al.'s Sportsthesia [4] facilitated the identification of keywords in the commentary to schedule mapped visualizations to the raw video feed. Similarly, Chen et al.'s VisCommentator [5] supported automated table tennis statistics augmentation using machine learning methods.

While supporting sports analysts helps derive relevant insights, there is limited research regarding improving end viewer experience with visual support or cues. The most notable shift happened with the development of Viz Libero [29] and Pierro [24], which instead supported video editors with visual editorial tools. Regarding the client side, Chen et al. [3] and Lin et al. [21] developed automated interactive embedded visualizations to enhance the end viewer's user experience. Chen et al.'s work utilized gaze-moderated embedded visualizations to assist less knowledgeable basketball viewers while providing an unobtrusive method of interaction, while Lin et al.'s Omniculars [21] used a simulated basketball match powered by voice interactions to adapt visualizations.

To conclude, embedded visualizations provide enhanced tools for sports analysts, but only a few limited research projects consider adapting content with embedded visualizations on the viewer's side. Our research contributes to research in automated commentary by deriving design considerations that provide insights into how to use embedded visualizations to italicize video content.

3 SYSTEM DESIGN

In this section, we present an automated commentary framework based on inferring events from spatio-temporal data and generative AI models.

3.1 Overarching System Concept

The fundamental principle underpinning our system design is drawn from the *sense-think-react* paradigm. To develop a context-driven system, it is first necessary to understand the content before developing logical pathways to the system output. In our framework, we utilize spatio-temporal data derived from a CV system to infer events before adapting output with respect to the current system state. In this manner, similarly to Karouzaki & Savidis [16], we adapt the *sense-think-react* paradigm by extending with further relevant attributes. Unlike Karouzaki & Savidis, we reorder to *sense-think-reflect-adapt-react* and add an additional *interpret* stage for *sense-think-reflect-interpret-adapt-react*. In doing so, we reflect on the understanding before interpreting the information to synthesize an output. We ground the

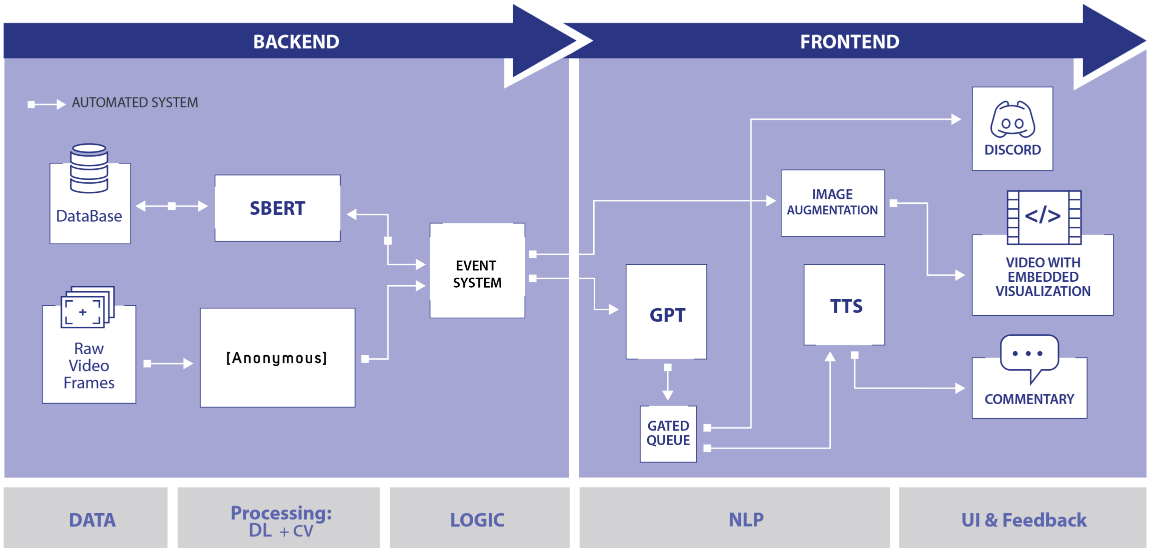


Fig. 1. The automated commentary system operates on a backend and frontend system that leverages MOT and perspective transformation matrices from the [anonymous] system. An integrated Event System detects in-game events and dispatches them throughout the system. Feature embeddings from Sentence Bidirectional Encoder Representations from Transformers (SBERT), derived from tracking identities, are matched against a database to supply context to the Generative Pre-trained Transformer (GPT) module. The automated commentary system, indicated by a white arrow, utilizes dialogue from the GPT module, regulated by a time-stamped queuing process. This gated queue prioritizes and releases events, with Google API converting Text to Speech (TTS) for audio while the text arrives at the Discord platform. The Image Augmentation system syncs with language processing to composite the embedded visualizations to complement the visual feedback. The automated commentary system is isolated from a larger model defined in [anonymous]

reorganization to align with the Cognitive Continuum Theory which states that thought processes range from intuitive to reflective judgement. However, an understanding must be reached in the decision-making process to reflect on the task's complexity and context. In doing so, we can provide both play-by-play and color commentary.

3.2 Model Overview

The automated commentary system outlined in Figure 1 can be deconstructed into two primary structures: a backend and a frontend system. The backend system consists of Deep Learning (DL) and CV algorithms that extract information from the video source. This data is utilized by the "Event System" module which determines events from the given data. These processes mimic the *sense-think* components of our architecture, whereas historical event information and a static database are the *reflect* element. The output is synthesized in the *interpret* stage before being prioritized and modified based on the current system context in the *adapt* phase. Finally, the output constitutes *react*, which is our feedback stage. We now consider the model overview with respect to the overarching system concept detailed above.

3.2.1 Sense. The automated commentary system begins operating in the *sense* stage by processing uncalibrated dynamic video footage. The system is designed to handle varying lighting conditions, diverse camera angles, and motion blur, ensuring robustness in different environments. Initial processing includes frame extraction and resizing images to 640*640 in preparation for the more advanced *think* stage.

3.2.2 *Think.* This stage serves two primary purposes:

(1) Extract information from video using DL and CV
(2) Spatio-temporal analysis to infer game states

For the first stage, we employ a model from [anonymous], an all-in-one model for player and ball MOT and localization in a top-down view. Figure 2 summarizes the model, which is built upon a YoloV7 [31] backbone trained on the ISSIA [9] and SoccerNet [7] datasets. The tracking module assigned bounding box identities with the Hungarian algorithm which is based on a cost matrix C:

$$C = \lambda_{\text{feat}}(1 - J) + \lambda_{\text{iou}} \cos(\theta) + \lambda_{\text{dist}}(|c_x - c_y|)^2 + \lambda_{\text{vel}} V \tag{1}$$

Where $1 - J$ is the inverted Jaccard Index, otherwise known as Intersection over Union (IoU), of all detections compared with the current gallery, $\cos(\theta)$ is the cosine similarity of feature embeddings extracted from each bounding box compared with the current gallery, while $|c_x - c_y|^2$ and V are the euclidean distance between bounding box centroids and velocity, respectively. Each of the lambda coefficients contributes to an overall sum of one, ensuring each attribute contributes to the final cost matric C by a predetermined amount. Output from the tracking module is corrected for any identity switches, and tracks are linearly interpolated to fill in any gaps before each track is manually labelled.

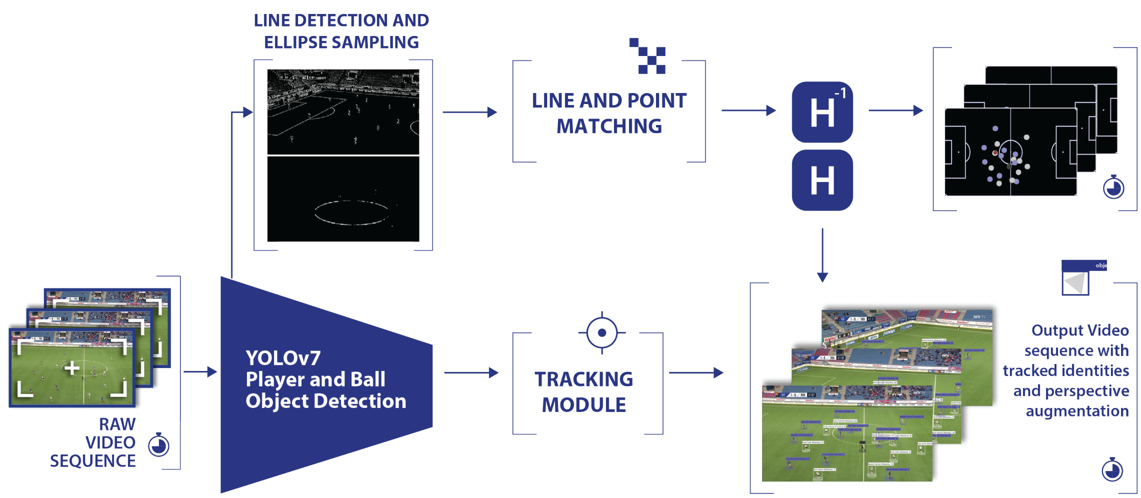


Fig. 2. [Anonymous] is an all-in-one model that detects and tracks players and the ball in football video. Furthermore, [anonymous] allows for the localization of tracked identities in a top-down view by computing the homography matrices with lines, interceptions, and ellipses extracted from intermediate conventional activation maps from the YoloV7 backbone.

[Anonymous] computes the homography matrices by extracting lines and ellipses from activation maps within the YoloV7 network, represented in Figure 2's upper processing pathway. This ensures that homographies can be computed from viewpoints with limited information, such as close-up central midfield viewpoints. Homographies are computed by matching lines and intersections with predefined templates and the extended Direct Linear Transform algorithm (DLT) [10]. We linearly interpolated the homography matrices from [anonymous] to smooth transitions before computing the top-down cartesian coordinates' trajectories, where the trajectories are smoothed with a Kalman Filter. Finally, we manually labelled tracks with player names.

Considering the second element of the *think* phase, the event system processes the MOT and top-down cartesian coordinate data to infer events via a rule-based system. The functionality of the event system can be summarized as deducing key events in football, primarily focusing on two core aspects: classification of ball possession and tracing ball trajectories. This approach is grounded in the fundamental principle of football, where possession dynamics play a critical role. We classify the player with possession by detecting collisions, implementing a more direct version of the Separating Axis Theorem (SAT), and computing the minimum and maximum x and y coordinates for each bounding box before checking for overlaps with the x-axis and y-axis individually. The overlaps along x and y are determined using logical AND operations:

$$\text{x\_overlap} = (x_{1_{\min}} \leq x_{2_{\max}}) \wedge (x_{2_{\min}} \leq x_{1_{\max}})$$

$$\text{y\_overlap} = (y_{1_{\min}} \leq y_{2_{\max}}) \wedge (y_{2_{\min}} \leq y_{1_{\max}})$$

When both x_overlap and y_overlap are `True`, a collision occurs and a player is considered as having possession if the possession time is above a certain threshold. If the ball intersects with multiple players' bounding boxes, our system prioritizes the bounding box associated with the player who currently possesses the ball. If ball possession is not clear, the system then assesses the proximity of the ball to the center base of the other overlapping bounding boxes to ascertain the player closest to the ball. In cases where the ball stops intersecting with the player with current possession, possession remains if the distance is below the threshold.

Table 1 refers to how each event is detected, specifying the datasource of the event, either based on the camera view tracking data or associated template cartesian coordinates. Methods of detection are refined to collisions, distance, acceleration, direction, rotation, trajectory, and location. Considering possession, events "Pass", "Interception", "Free Kick", "Box", "Cross", and "Shot", all rely on understanding who is in possession of the ball at any one time, whereas "Throw", "Deflection", "Challenge", and "Open Ball" are dependent on accurate tracking of the ball to determine locations and trace trajectories. Note that "Box", "Cross", and "Shot" all consider both possession and ball tracking data.

3.2.3 Reflect. This state is an essential element that derives the contextual information regarding the current game state and historical data. In this manner, the *reflect* state considers the current state on a season and game level. On a seasonal level, the system queries a database containing information and statistics surrounding player and team performance. It does so by computing feature encodings with Sentence Bidirectional Encoder Representations from Transformers (SBERT) of the players involved in the event and measuring the cosine similarity with pre-computed feature vectors of the database. Upon retrieval, this information is used for the next *interpret* stage and serves as color commentary. Regarding game-level information, the event system collects and updates:

- Completed pass count
- Intercepted pass count
- Pass completion rate
- Total possession time
- Distance covered
- Current speed

Event	Data	Method	Logic
Possession			*Measures overlap and distance of the ball with a respective player.*
Pass			*Possession changes to player of the same team.*
Interception			*Possession changes to player of the opposing team.*
Free kick			*Player has possession while ball and players are static.*
Throw			*After a cutscene the ball enters play from the sideline*
Deflection			*Significant change in trajectory while the closest player is of opposing team.*
Challenge			*Sudden change in trajectory with two opposing players in local vicinity.*
Open Ball			*Spline fitted to ball trajectory, if no players are within the vicinity, it is a free ball.*
Box			*Player with possession enters other teams penalty box.*
Cross			*Player who last had possession is located on the wing and the ball accelerates towards the opponents penalty box*
Shot			*Player who last had possession is within a distance threshold of the opponents goal and the ball accelerates toward the goal.*

Table 1. Methodology for detecting events through camera and template viewpoints. Methods include collisions, distance, acceleration, direction, rotation, trajectory, and location

Similar to the seasonal level data, the game level data is called upon when an event is detected. The game-level data is used for both color and play-by-play commentary. When initializing the system, each player profile is constructed with random, sensible values reflecting the player's team position.

3.2.4 Interpret. The seasonal and game-level data serve as a context for interpreting and synthesizing a response. We use GPT3.5-Turbo 0613 model to generate a natural language response based on predefined event templates populated by information from the event system. Data from the *reflect* phase serves as context to the GPT-3,5 model to assist the model in producing relevant, timely commentary. GPT3.5 system context was initialized with instructions for producing football commentary (more details in Section 3.3). We opted for GPT3.5 over GPT-4 because GPT3.5 provides faster inference speeds and is more cost-effective.

3.2.5 Adapt. Due to the fast-paced nature of football, the system needed to adapt quickly to ever-evolving game states and events. While some events have a higher importance than others, it is important for commentators to be selective over which ones they cover. For this reason, we implemented a pooling system consisting of a gate queue. Each set of dialogue from the *interpret* phase was given a timestamp reflecting its birth, lifespan considering its longevity, and priority based on the event type: $E = \{dialogue, timestamp, lifespan, priority\}$. The queue updates at each event iteration while releasing the next piece of dialogue when the TTS module is available. We allow "Shot Event", "Cross Event", and "Box Event" to interrupt the TTS module, as these events are considered most vital. Furthermore, play-by-play commentary takes precedence over color because commentators generally prioritize communicating relevant current events.

3.2.6 React. Finally, the *react* step provides system feedback for the user. The TTS module plays the relevant dialogue while the image augmentation module receives the information required to augment relevant embedded visualizations. The result is synchronized embedded visualizations with context-aware commentary.

3.3 Designing Commentators

Unlike previous research in automated sports commentary, our methodology facilitates feedback from two AI commentators. Inspired by traditional sports commentary, we strive to replicate the dynamic interplay offered by dual commentators. The core characteristics of the commentators were defined in the system context of the GPT model. We provided a few sentences describing the commentator's name, gender, the team they support (if any), and any personality traits relevant to the commentator. Prior to deciding which profile to use for the user study, we ran a small experiment comparing different profiles. Before adopting the characteristics of the commentators, we first decided to have a male and a female commentator to provide balance to commentating on women's football. Traditionally, one commentator would document play-by-play commentary while the other supports it with color commentary. However, for our prototype, to promote conversational interactions between the two commentators, we did not assign individual roles.

Sports commentators profoundly impact the viewer's perception of the game by controlling the narrative with respect to how they interpret current events [20]. Therefore, it is no surprise that differing commentator characteristics will dramatically impact the viewer's game experience. Zhou et al. [38] found conflicting commentary improved the perceived enjoyment of the commentary compared to complementary commentary. To observe the effect of adapting characteristics of the AI commentators on dialogue, we compare three profiles:

- **Neutral** [*Complimentary*]: Both commentators did not support either team.
- **[Team 1] Supporters** [*Complimentary*]: Both commentators support [Team 1]
- **[Team 2] Supporters** [*Complimentary*]: Both commentators support [Team 2]
- **Support Rival Teams** [*Contradictory*]: Emily is an excitable [Team 2] supporter, whereas Doug is a grumpy [Team 1] supporter.

Often, commentators may express bias towards a certain team, which often happens during international matches. Therefore, we ran experiments adapting the bias of the commentators. For "Support Rival Teams", we constructed the profile so that one commentator had excitable characteristics, whereas the other was grumpy. We exported dialogue from the system to observe any changes in dialogue according to these characteristics.

BOX EVENT	 
Neutral	**Emily**: Simmons with a great opportunity here for [Team 1]!
Support [Team 1]	**Emily**: Come on Simmons, show them what you're made of!
Support [Team 2]	**Doug**: Oh no, not Karina Simmons again! She's been a real thorn in [Team 2]'s side in the past. We need to keep a close eye on her.
Support Rival Teams	**Emily**: Oh my goodness, Karina Simmons is making a move in [Team 2]'s box! **Doug**: What is she thinking, trying to score from there?

Table 2. Examples of commentary with varying commentator characteristics

Table 2 displays various commentaries corresponding to a "Box Event", illustrating how the commentators' characteristics influence their description of events. For neutral commentary, Emily remains unbiased, remarking on a "great opportunity" without favoring any team. In contrast, when supporting [Team 1], Emily motivates the player with encouragement. Doug's response when supporting [Team 2] is characterized by a tense tone, highlighting the threat posed by the opposition player, Simmons. The "support rival teams" category exemplifies a dynamic interplay between Emily and Doug. Emily's excitement for [Team 1] is met with Doug's concern for [Team 2], reflecting a balance of divergent loyalties that enrich the commentary with a sense of competition. Please see Appendix A for more examples of these commentator characteristics.

These experiments showcase how simple adjustments in the system context can dramatically change the intent of the commentators. For our user study, we chose to keep the commentary neutral because it allows us to establish a baseline without the influence of perceived bias. Maintaining a neutral stance prevents alienating any participants due to favoritism, providing a uniform experience across the diverse group. Considering the color commentary, a more neutral stance resulted in more friendly interactions between the two commentators during a cutscene:

> **Emily:** *Karina Simmons has been a key player for [Team 1] in previous seasons.*
>
> **Doug:** *Absolutely, Emily. Simmons has consistently performed well for [Team 1], with a strong goal-scoring record.*
>
> **Emily:** *In 2021, she scored 3 goals and provided 5 assists. That's quite impressive!*
>
> **Doug:** *And in 2022, she's already scored 5 goals and provided 5 assists. She's been a real threat in front of the goal.*

By not explicitly designating play-by-play and color commentary roles and eliminating biases, the result is less confrontational dialogue with factual information expressed by commentators bouncing off one another. The commentators effectively collaborate in conveying statistics, assisting viewers in better understanding and acknowledging the game's dynamics.

3.4 Embedded Visualizations

(a) Visualizations for the main view.

(b) Visualizations for cutscene view.

Fig. 3. Visualizations for different camera viewpoints.

We enhance our commentary with basic embedded visualizations to assess their effectiveness in augmenting user comprehension and enriching the interpretation of sports content. Our primary visualization employs a spotlight mechanism, focusing viewers' attention on the player that is the focus of the current commentary. Above the highlight is a header which informs the viewer of the player's name, position, and number. Complementing the spotlight is a player card offering detailed profile information with a portrait to give viewers a complete understanding of the player in focus. These visualizations dynamically track the player with possession of the ball within the main camera view, as depicted in Figure 3a. During cutscene transitions, the player card adapts to the commentator dialogue to bring further context to the color commentary, reinforcing the narrative (see Figure 3b). The implementation of these visualizations aims to investigate their utility in helping viewers more easily identify and connect with the players who are the subject of the commentary. By integrating these elements into the primary viewing experience, we seek to determine if such enhancements can improve the overall engagement and satisfaction of the audience with the sports broadcast.

4 USER STUDY

We utilized a subset of eight participants from a larger user study focused on automated commentary systems. In this larger study, these eight participants initially viewed the automated commentary system under investigation, so the counterpart system did not influence their opinions of the non-interactive system.

The data we have re-analyzed in this paper consists of recordings from the think-aloud session and the post-system questionnaire, which both were collected from the participants before interacting with the counterpart system. In our analysis, both think-aloud and the post-system questionnaire are used as qualitative data sources. The post-system questionnaire (Appendix B) consists of statements under the subcategories of "knowledge and understanding", "engagement and immersion", "satisfaction and future use", "trust and reliability", and "consistency". For the larger user study, we took the mean values of subcategories from the post-system questionnaire. However, in this study, we display the original attributes with their respective Likert scale responses to gain a more in-depth understanding of individual features of the non-interactive system.

4.1 Setup

Below is an overview of the segments from the larger user study that pertain to the data utilized in our analysis. All participants were first briefed about the project and introduced to the prototype. They were then informed about their rights and asked to sign a consent form. Following, all participants completed a pre-study questionnaire regarding their demographic, as well as their self-described football habits and background. Upon completing the preparatory steps outlined below, participants proceeded to undertake the comprehensive study as detailed in [anonymous].

User study: A researcher demonstrated the automated commentary system. Participants then got to explore the mode of automated commentary by watching a one-minute video clip that could be replayed. Afterwards, they proceeded with the full testing session, where they watched a four-minute video with the automated AI commentators and the embedded visualizations. Participants were asked to express their thoughts throughout the full-testing session based on the concurrent think-aloud method [11]. We used this method because it allows the participant to express more nuanced information from their short-term memory, which is valuable in assessing such a system.

Questionnaire: When finished with the user study, the participants filled out a post-system questionnaire covering "knowledge and understanding", "engagement and immersion", "satisfaction and future use", "trust and reliability", "consistency", and "overall preference".

4.2 Participants

The participants consisted of four males and four females between the ages of 20 and 28. The participants were recruited through project promotions on several university course websites and through verbal pitches during lectures. The participants received gift cards of [ANONYMIZED] as an incentive for being part of the full trial.

As part of the pre-study questionnaire, participants answered questions about their football habits and estimated how often they watch football matches during an average month. Based on their answers we distinguished between active and less active football viewers. Four participants (P01, P02, P03, P07) reported watching several matches each month and can be described as active football viewers, while the other four participants (P04, P05, P06, P08) reported not watching any football matches and can be described as less active football viewers.

5 RESULTS

In this section, we present the results of the user study.

5.1 Quantitative

Based on the post-system questionnaire, participants reported the system to be effective in conveying easily comprehensible information and providing a better understanding of the game and player performance, and users were fairly neutral with their responses regarding in-game developments. We can speculate that users perception of the play-by-play commentary was similar to that of regular commentators, therefore they did not feel negatively towards the experience but instead felt indifferent. With respect to 'knowledge and understanding', users reported color commentary provided more significant improvement in understanding (C3) when compared to play-by-play (C4). All participants reported considering the system reliable and trustworthy, delivering consistent statements and feedback without any encountered conflicts. Opinions were split regarding the system's ability to heighten engagement and enjoyment, and several participants (38%) reported finding the system to be overly complex. However, the majority expressed feeling immersed in the football match, and most (75%) felt it added value to traditional football viewing. All participants except

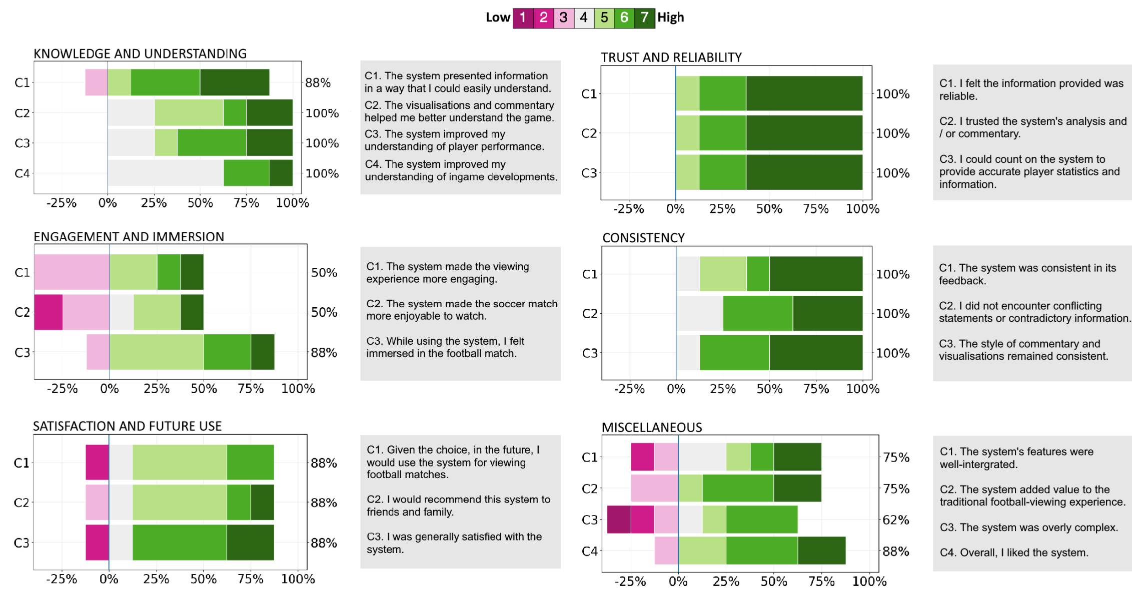


Fig. 4. Results of ratings for the post-system questionnaire.

one expressed overall satisfaction with the automated commentary, a desire to use the system in the future, and an inclination to recommend it to family and friends.

5.2 Qualitative

During the think-aloud session, participants appeared impressed with the prototype and compared the automated commentators to human commentators, with the exclusion of artificial voices and lack of emotions during big events in the game. Several participants emphasized that they liked how the two commentators interacted with each other and how the conversation bounced back and forth between them. As P08 explained:

> "[It is] actually quite similar to actual commentators. [. . .] With the exception of human emotion, of course."

P02 expressed a similar view, even though he stated not being a fan of real commentators:

> [It] sounds very FIFA-like. It doesn't mean that it is better or worse than normal commentators because normal commentators are quite bad, I think."

P02 continued with explaining that he found the information the commentators presented as interesting and that the content was good. In general, most participants praised the commentary for delivering interesting and precise information, and P01 highlighted that it provided in-game developments he had not noticed by himself.

Participants appreciated the blend of color and play-by-play commentary, highlighting how the commentators filled the gaps with interesting information, such as statistics and fun facts about the players. However, participants reported mixed feelings about sudden transitions between these elements in instances when the commentators abruptly shifted from color commentary to prioritizing play-by-play, leaving sentences unfinished when transitioning. Some participants, such as P01, liked that the commentators prioritized the play-by-play, while others, such as P06, seemed to find it too abrupt and confusing.

Even though all participants reported the information as trustworthy, some expressed disagreement with the commentator's choice of words when describing certain events during the game. P08 stated that some of the words the commentators used did not seem fitting for the specific event, and P03 explained his disagreement with the commentators when they commented on a "tackling" and stated that it was rather an "interception."

There were split opinions in regard to the combination of embedded visualization with the commentary. Some participants felt that the combination of the player card element together with the commentary helped them improve their knowledge about players on the pitch. However, some of the less active viewers experienced getting overwhelmed with the combination of the auditory information from the commentators and the visual elements. P05 explained how she had to concentrate more compared to a normal football match as it was more information to process. She further added that it could be difficult to process the different types of information and that she had to choose one to focus on (either the visual or the commentary). As she stated:

> "[It is] [t]oo much to see and too much to read."

P05 further specified that the combination of several visual elements at once was distracting, and suggested only to implement the player cards. This view is consistent with most participants' preferences. Most participants seemed to prefer the player card embedded visualization over the tracking of possession as the latter could be experienced as more distracting, especially when changing which players it highlighted. Most participants described the player cards as useful for getting to know the different players, however, they also mentioned several points of improving the design: making the design smaller and more transparent, and to include less information. Especially the active football viewers were very specific in what information they found interesting and not, but this varied within the group.

6 DISCUSSION

In this section, we begin by discussing the overall results outlined in the previous section before outlining recommended design considerations future research should contemplate while designing automated commentary systems.

The quantitative and qualitative results have provided some intriguing insights into user perception of our automated commentary system. Along with the extended system, [anonymous], this is the first to incorporate play-by-play and color commentary to improve user understanding of players on the pitch and in-game events. Our system has innovated the traditional approach to play-by-play sports commentary by integrating natural language processing techniques. This system generates real-time, play-by-play narrative responses by analyzing live data from the event tracking system. Our methodology involves extracting key game developments and translating these into coherent, natural language commentary, replicating the dynamic and informative style of traditional sports broadcasts. Quantitative results have reported a fairly neutral response to the system's ability to communicate in-game developments effectively. While the results suggest room for improvement, they indicate that the users may be indifferent to automated play-by-play commentary. P08 reported similarities between regular and AI commentators, suggesting participants might have been unimpressed or indifferent to the novel experience. This may contribute to a more neutral response in the Figure 4 "knowledge and engagement" category (c4). Additionally, other participants highlighted inaccuracies in the language output of the NLP module, which may further support the inclination towards neutral responses regarding play-by-play commentary. These observations indicate that while AI commentators can mimic traditional commentary, some areas, particularly in accuracy and novelty, may only partially meet the users' expectations or preferences.

Considering Figure 4 "knowledge and understanding" c3, the result indicates that the automated color commentary improves the viewer's understanding of knowledge of player performance. This supports the qualitative feedback reporting some users found the system improved their knowledge of players. The system was designed to construct commentary using the in-game and database data concerning the player currently in view during a "cutscene". These cutscenes often occurred during periods of inaction during the game, where the color commentary filled in these spaces. These intervals provide extended opportunities for more detailed and comprehensive commentary. During these extended intervals, the GPT module has the opportunity to fully showcase its capabilities for dynamic color commentary, effectively mimicking the dualistic nature of regular sports commentary. As indicated by the qualitative data, some participants noted the dynamic interaction between the two commentators heightened the sense of realism. Our findings suggest that color commentary was more effective, primarily due to the NLP module's ability to interpret a wide array of data with more tokens. With additional tokens and time, it was able to synthesize dynamic and nuanced AI commentary that effectively captured the dualistic characteristics of regular sports commentary.

While the system provides a combination of play-by-play and color commentary supported by dual commentators, the main limitation is its ability to synthesize commentary with emotional cues. The TTS module's artificial voices and the dynamic language's monotone expression might contribute to the more neutral ratings regarding play-by-play commentary (Figure 4 "knowledge and understanding" (c3)). Real commentators would express emotion and excitement to reflect the current events within the game. During our methodology, we have shown the range of emotions possible with our model (Appendix A). By implementing a more sophisticated TTS methodology, the dynamic range of the output would better reflect synthesized natural language.

When designing the automated commentary system, we included embedded visualizations to promote italicizing the portrayed information. We aimed to improve the system's novelty by supporting commentary with visual information, as currently, this is not possible in real-time sports viewing scenarios. We designed two simple embedded visualizations to complement the commentary and assist the users in identifying in-game events and team players. While the quantitative data indicates that the visualizations supported their understanding of the commentary, the qualitative data gives a more nuanced understanding of these insights. Interestingly, we observed inactive viewers found italicizing the content overwhelming. Their lack of familiarity with the football content might cause this reaction. Consequently, participants with limited experience viewing football matches may feel a heightened cognitive load when presented with additional visual data. Preliminarily, results indicate that while italicizing assists in understanding information, more attention is needed to ensure it supports and does not detract from the user experience.

6.1 Design considerations

Referencing both the user study results (Section 5) and our previous discussion, we now state design considerations that future research should refer to while developing future automated commentary systems.

6.1.1 The content of the commentary. Our prototype transitions between play-by-play and color commentary depending on the camera's viewpoint. As previously discussed, our color commentary received slightly higher ratings than its counterpart. Users reported they enjoyed the dynamic interplay of dialogue between commentators, mainly presented through color commentary. However, some users disliked prioritizing play-by-play commentary over color, particularly when a play-by-play event interrupted the current color dialogue. Our gated queuing system was responsible for this drawback as it prioritized some events over others. Future work should consider how to blend these two types of commentary to ensure the flow of commentary is consistent. Ensuring a consistent flow of commentary could enhance its perceived quality while maintaining viewer engagement, providing a seamless and immersive experience that mirrors the dynamic nature of live sports broadcasting.

6.1.2 The importance of wording. Sports commentators efficiently communicate in-game developments while adding additional context to the match. Therefore, the content of the commentary must be precise and clearly communicate information using the correct technical words relative to the game. In our study, two participants reported incorrect use of language for a tackle event. While this misinterpretation of information did not compromise the trust and reliability of the system (documented in Figure 4 trust and reliability), it may damage the integrity of the automated system. Interestingly, as demonstrated in Figure 4, this misinterpretation did not negatively impact c2 (Q3 = 100%), indicating these participants had a high tolerance level for minor inaccuracies in the system. In this circumstance, misrepresentation of information is a byproduct of the limitations of the event system. Using a rule-based approach, we built our event system to derive insights from spatial-temporal data, as outlined in Section 3.2.1. Therefore, deciphering a tackle from an interception was not plausible, resulting in tackles classified instead as interceptions. With a more sophisticated event system, improving event classification and interpreting a more extensive variety of events is possible. We recommend utilizing a CNN trained on the event dataset from [7], which consists of seventeen events. Training a model on this dataset would ensure timely reports of play-by-play events and increase the variety of events and commentary. Populating the GPT module with more accurate templates from an intricate event system will enhance the relevance of the commentary and ensure the correct articulation of the generated dialogue. The result would be more relevant and consistent commentary, accurately articulating the play-by-play events. Viewers would likely experience more trust in the system's ability to report events. As a result, users could experience increased engagement and satisfaction with the system. Notably, these aspects represent the areas with the lowest scores in our quantitative analysis, as illustrated in Figure 4 ("Engagement and Satisfaction").

6.1.3 The dynamic interplay of the commentators. In traditional sports broadcasting, the dynamic interplay between play-by-play and color commentators significantly enhances the viewer experience. This synergy is a crucial aspect that automated commentary systems should aim to replicate. Our approach, a pioneer in incorporating dual commentators in an automated commentary system, revealed some participants enjoyed the dynamic interplay between the commentators. While our system's commentators share the role of communicating play-by-play and colour commentary, we recommend other researchers experiment further with individual roles and their impact on the automated commentary. Enhancing the interplay between commentators could lead to a more engaging and immersive viewing experience, replicating human sports commentary's lively and varied nature.

6.1.4 The characteristics of the commentators. Research in sports commentary has reported that commentary can alter viewer perceptions [12], and commentary type can impact the viewer's liking of the commentary [38]. Therefore, the AI commentator's characteristics may significantly impact the viewer experience. Although our results did not find any specific viewer preference regarding commentator personalities, we did run some experiments adjusting the characteristics of the commentators (see Section 3.3 and Appendix A). These experiments outlined the plausibility of controlling the system's output by defining simple characteristics. We found that by adjusting the commentator personalities, we could draw biases that redefine the context of the current game state. Therefore, a key design consideration is adapting the personalities of the two commentators, which may impact viewer perception and their overall experience of the match. For example, personalities could adjust to diverse viewer preferences, which may better suit their opinions. This could be as simple as adding bias into the commentary, as shown in Appendix A, or more complex personalities could heavily stylize the commentary, altogether redefining the traditional roles of commentary.

6.1.5 Emotional depth of commentary. Sports commentators contribute to the viewer's emotional engagement by adapting the narrative and style of their commentary [19]. By expressing emotions vocally, they can enhance the viewer's emotional responses. Therefore, automated commentary systems should strive to replicate emotions that reflect the game's current state. Rocco [30] tried to replicate emotions by controlling pitch and tone. However, more complex attributes may accompany these, such as volume, speech rate, timbre, inclusion of non-lexical vocables, and variations in the generated dialogue. In our study, we only displayed variations in generated dialogue and would encourage future research further to promote the emotion with the relevant respective cues. Our qualitative study highlighted participants who felt the commentators' voices were artificial and lacked emotion. Therefore, to further engage viewers, the system must generate artificial commentary that reflects the respective emotions of the current game state. As Lee et al. [20] proposed, viewing enjoyment is not necessarily related to the outcome but instead to the emotions experienced throughout the match. We hypothesize that by adapting the emotional response of the AI commentators, the viewer will feel more emotionally engaged with the content, increasing game stakes and, in turn, enjoyment.

6.1.6 Personalized italicizing. Italicizing supports visual cues with commentary, adding a multimodal element to our system. By interpreting events through an augmentation and an NLP module, we could synchronize the output of both modalities to the current game state. We incorporated a spotlight to highlight attention to the focus of the commentators and a player card to help users identify players on the pitch. Participants in our user study reported that multimodal information was helpful in identifying players on the pitch. While users appreciated the player card, they felt the accompaniment of the spotlight was distracting. These insights could contribute to the lower score in "engagement and immersion" (Figure 4 C1 and C2) as some less active viewers reported being overwhelmed by the italicizing. Therefore, future iterations of automated commentary should carefully consider the role of italicizing. Developers should carefully create and place the visualizations to ensure they do not detract from the current events. We also recommend personalizing the visualizations and commentary to the user's preference and knowledge level to ensure their relevance to the user.

7 LIMITATIONS

Our system's main limitation is its real-time applications. The current drawback is that [anonymous], used to preprocess the video, is currently incapable of real-time inference. Furthermore, [anonymous] cannot automatically assign the corresponding player identities for the automated commentary system. Future iterations of MOT in football should aim to overcome these limitations. Another aspect impacting real-time capabilities is the latency of processing commentary through the OpenAI GPT3.5 API. It is likely that OpenAI will improve model latency with future iterations of the GPT models. Another limitation is that the automated commentary system does not support the game's strategic analysis. Adding strategic knowledge would further enhance the perceived usability of the automated commentary system. Our future work will consider deriving insights from spatial-temporal data to provide extended context for the AI commentators. Our work has shown that the play-by-play commentary is limited to the rule-based event systems' ability to detect certain events. We recommend future work to implement a DL solution to detect a more extensive variety of in-game events. As highlighted by our user study, one main drawback is the "robotic" delivery of commentary, which does not reflect the diversity in language produced by the system. We suggest future researchers carefully consider the role of delivery to better communicate the emotional stakes of the match. In doing so, viewers should feel more engaged with content.

8 CONCLUSION AND FUTURE WORK

Using CV techniques and natural language processing, our artificial commentary system draws insights from uncalibrated video sequences and provides multimodal feedback by NLP and an augmentation module. We conducted a small user study (n=8) using the concurrent think-aloud method to gain insights from the participants during the viewing experience. Results indicate that users appreciated the combination of artificial play-by-play and color commentary, though some found the transition between these styles distracting. While the dual commentary and conversational style were generally well-received, miswording in the play-by-play commentary, notably in classifying interceptions and tackles, was highlighted by two participants. The italicizing feature was seen as beneficial by most users for enhancing their understanding of players on the pitch. However, less-active viewers found it distracting, and there were varied opinions on the relevance of the information presented in the visualizations. Our work represents a significant step forward in the field of automated sports commentary while providing valuable insights into viewers' perceptions of AI commentary. It does so by presenting a system capable of translating information extracted from the video feed into commentary and analyzing findings from the user study to provide design considerations for future automated sports commentary systems. While our approach provides a novel methodology for formulating automated commentary, it does have certain limitations. In particular, the video is first preprocessed through a CV model, impacting the system's real-time capabilities. Moreover, while the AI commentators acknowledge player performance, they currently cannot understand the match's strategic components. Building on these insights and acknowledging the limitations of our current system, we have identified several critical areas for further development and research. We provide design considerations regarding the content of commentary, the importance of wording, the dynamic interplay of commentators, the characteristics of the commentators, the emotional depth of commentary, and personalized italicizing. These design considerations open future avenues of research regarding the impact of the AI commentators' personalities and emotions on viewers' perceived engagement and emotional response, personalization of embedded visualizations with supported commentary, and dynamic biases between commentators' impact on user satisfaction. Further research referencing these design considerations could revolutionize how we view sports content. The personalization of commentary tracks with visualization can improve the accessibility of content and perceived quality. This could assist lower-budget sports broadcasts in three ways. Firstly, the perceived quality of broadcast content directly influences viewing intent. Secondly, AI commentary can provide a low-cost personalized solution to commentary. Finally, we should think about more than just replicating an experience through AI but instead redefining it to enhance the user experience. Our work signifies a significant step in this direction, displaying the possibility of multimodal feedback while viewing sports matches.

9 ACKNOWLEDGMENTS

REFERENCES

[1] Sheena Anees. 2015. Computational Intelligence Based Color Commentary System in Sports. (2015).

[2] Kim Binsted and Sean Luke. 1999. Character Design for Soccer Commentary. *Lecture Notes in Computer Science* (1999). https://doi.org/10.1007/3-540-48422-1_2

[3] Zhutian Chen, Qisen Yang, Jiarui Shan, Tica Lin, Johanna Beyer, Haijun Xia, and Hanspeter Pfister. 2023. iBall: Augmenting Basketball Videos with Gaze-moderated Embedded Visualizations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.

[4] Zhutian Chen, Qisen Yang, Xiao Xie, Johanna Beyer, Haijun Xia, Yingcai Wu, and Hanspeter Pfister. 2022. Sporthesia: Augmenting sports videos using natural language. *IEEE transactions on visualization and computer graphics* 29, 1 (2022), 918–928.

[5] Zhutian Chen, Shuainan Ye, Xiangtong Chu, Haijun Xia, Hui Zhang, Huamin Qu, and Yingcai Wu. 2021. Augmenting sports videos with viscommentator. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2021), 824–834.

[6] Dr. Chitrakala. 2015. STORY SELECTION AND RECOMMENDATION SYSTEM FOR COLOUR COMMENTARY IN CRICKET. (2015).

[7] Adrien Deliege, Anthony Cioppa, Silvio Giancola, Meisam J Seikavandi, Jacob V Dueholm, Kamal Nasrollahi, Bernard Ghanem, Thomas B Moeslund, and Marc Van Droogenbroeck. 2021. Soccernet-v2: A dataset and benchmarks for holistic understanding of broadcast soccer videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4508–4519. https://doi.org/10.1109/CVPRW53098.2021.00508

[8] Carlos Dietrich, David Koop, Huy T Vo, and Cláudio T Silva. 2014. Baseball4d: A tool for baseball game reconstruction & visualization. In *2014 IEEE conference on visual analytics science and technology (VAST)*. IEEE, 23–32.

[9] Tiziana D'Orazio, Marco Leo, Nicola Mosca, Paolo Spagnolo, and Pier Luigi Mazzeo. 2009. A semi-automatic system for ground truth generation of soccer video sequences. In *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*. IEEE, 559–564. https://doi.org/10.1109/AVSS.2009.69

[10] Elan Dubrofsky and Robert J Woodham. 2008. Combining line and point correspondences for homography estimation. In *International symposium on visual computing*. Springer, 202–213. https://doi.org/10.1007/978-3-540-89646-3_20

[11] David W Eccles and Güler Arsal. 2017. The think aloud method: what is it and how do I use it? *Qualitative Research in Sport, Exercise and Health* 9, 4 (2017), 514–531. https://doi.org/10.1080/2159676X.2017.1331501

[12] Evan L Frederick, Choong Hoon Lim, Jinwook Chung, and Galen Clavio. 2013. Determining the effects of sport commentary on viewer perceptions, attitudes, beliefs, and enjoyment through violence justification. *Journal of Sports Media* 8, 1 (2013), 65–86.

[13] Kirk Goldsberry. 2012. Courtvision: New visual and spatial analytics for the nba. In *2012 MIT Sloan sports analytics conference*, Vol. 9. 12–15.

[14] Cagatay Goncu and Daniel J. Finnegan. 2021. 'Did you see that!?' Enhancing the experience of sports media broadcast for blind people. *Lecture Notes in Computer Science* (2021). https://doi.org/10.1007/978-3-030-85623-6_24

[15] Manon Ichiki, Toshihiro Shimizu, Imai Atsushi, Tohru Takagi, Mamoru Iwabuchi, Kiyoshi Kurihara, Taro Miyazaki, Tadashi Kumano, Hiroyuki Kaneko, Shoei Sato, Seiyama Nobumasa, Yuko Yamanouchi, and Hideki Sumiyoshi. 2018. Study on automated audio descriptions overlapping live television commentary. *Lecture Notes in Computer Science* (2018). https://doi.org/10.1007/978-3-319-94277-3_36

[16] Effie Karouzaki and Anthony Savidis. 2012. A framework for adaptive game presenters with emotions and social comments. *International Journal of Computer Games Technology* (2012). https://doi.org/10.1155/2012/929814

[17] Tadashi Kumano, Tohru Takagi, Manon Ichiki, Kiyoshi Kurihara, Hiroyuki Kaneko, Tomoyasu Komori, Toshihiro Shimizu, Seiyama Nobumasa, Imai Atsushi, and Hideki Sumiyoshi. 2019. Generation of Automated Sports Commentary from Live Sports Data. (2019). https://doi.org/10.1109/BMSB47279.2019.8971879

[18] Kiyoshi Kurihara, A. Imai, N. Seiyama, Toshihiro Shimizu, Shoei Sato, Ichiro Yamada, T. Kumano, Reiko Tako, Taro Miyazaki, Manon Ichiki, T. Takagi, and H. Sumiyoshi. 2019. Automatic Generation of Audio Descriptions for Sports Programs. *SMPTE Motion Imaging Journal* (2019). https://doi.org/10.5594/JMI.2018.2879261

[19] Greg Lee, V. Bulitko, and Elliot A. Ludvig. 2014. Automated Story Selection for Color Commentary in Sports. *IEEE Transactions on Computational Intelligence and AI in Games* (2014). https://doi.org/10.1109/TCIAIG.2013.2275199

[20] Minkyo Lee, D. Kim, Antonio S. Williams, and P. Pedersen. 2016. Investigating the Role of Sports Commentary: An Analysis of Media-Consumption Behavior and Programmatic Quality and Satisfaction. (2016). https://doi.org/10.1353/JSM.2016.0001

[21] Tica Lin, Zhutian Chen, Yalong Yang, Daniele Chiappalupi, Johanna Beyer, and Hanspeter Pfister. 2022. The quest for: Embedded visualization for augmenting basketball game viewing experiences. *IEEE transactions on visualization and computer graphics* 29, 1 (2022), 962–971.

[22] Tania Modleski. 1986. *"Television/Sound" in Studies in entertainment: critical approaches to mass culture*. Vol. 7. Indiana University Press.

[23] Hannah Pileggi, Charles D Stolper, J Michael Boyle, and John T Stasko. 2012. Snapshot: Visualization to propel ice hockey analytics. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2819–2828.

[24] ROSS. [n. d.]. PIERO Sports Graphics Analysis. https://www.rossvideo.com/live-production/graphics/piero/. (Accessed on 10/08/2023).

[25] Manuel Stein, Thorsten Breitkreutz, Johannes Haussler, Daniel Seebacher, Christoph Niederberger, Tobias Schreck, Michael Grossniklaus, Daniel Keim, and Halldor Janetzko. 2018. Revealing the invisible: Visual analytics and explanatory storytelling for advanced team sport analysis. In *2018 International Symposium on Big Data Visual and Immersive Analytics (BDVA)*. IEEE, 1–9.

[26] Manuel Stein, Halldór Janetzko, Thorsten Breitkreutz, Daniel Seebacher, Tobias Schreck, Michael Grossniklaus, Iain D Couzin, and Daniel A Keim. 2016. Director's cut: Analysis and annotation of soccer matches. *IEEE computer graphics and applications* 36, 5 (2016), 50–60.

[27] Kumiko Tanaka-Ishii, K. Hasida, and I. Noda. 1998. Reactive Content Selection in the Generation of Real-time Soccer Commentary. *ACL* (1998). https://doi.org/10.3115/980691.980778

[28] Kumiko Tanaka-Ishii, I. Noda, I. Frank, H. Nakashima, K. Hasida, and H. Matsubara. 1998. MIKE: an automatic commentary system for soccer. *Proceedings International Conference on Multi Agent Systems (Cat. No.98EX160)* (1998). https://doi.org/10.1109/ICMAS.1998.699067

[29] VizRT. [n. d.]. Viz Libero - VizRT. https://www.vizrt.com/products/viz-libero/. (Accessed on 10/08/2023).

[30] Dirk Voelz, Elisabeth André, Gerd Herzog, and Thomas Rist. 1999. Rocco: A RoboCup Soccer Commentator System. *Lecture Notes in Computer Science* (1999). https://doi.org/10.1007/3-540-48422-1_4

[31] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. 2022. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696* (2022). https://doi.org/10.48550/arXiv.2207.02696

[32] Zihan Wang and Naoki Yoshinaga. 2021. From eSports Data to Game Commentary: Datasets, Models, and Evaluation Metrics. (2021).

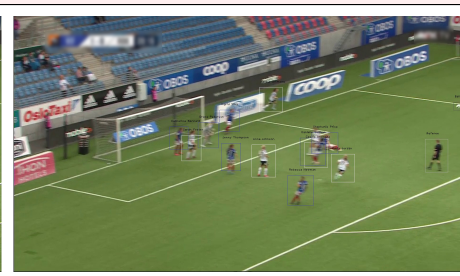
[33] Lawrence A Wenner. 1989. *Media, sports, and society.* sage.
[34] Yingcai Wu, Xiao Xie, Jiachen Wang, Dazhen Deng, Hongye Liang, Hui Zhang, Shoubin Cheng, and Wei Chen. 2018. Forvizor: Visualizing spatio-temporal team formations in soccer. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 65–75.
[35] Xiao Xie, Jiachen Wang, Hongye Liang, Dazhen Deng, Shoubin Cheng, Hui Zhang, Wei Chen, and Yingcai Wu. 2020. PassVizor: Toward better understanding of the dynamics of soccer passes. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2020), 1322–1331.
[36] Shuainan Ye, Zhutian Chen, Xiangtong Chu, Yifan Wang, Siwei Fu, Lejun Shen, Kun Zhou, and Yingcai Wu. 2020. Shuttlespace: Exploring and analyzing movement trajectory in immersive visualization. *IEEE transactions on visualization and computer graphics* 27, 2 (2020), 860–869.
[37] Maliang Zheng and Daniel Kudenko. 2010. Automated event recognition for football commentary generation. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)* 2, 4 (2010), 67–84.
[38] Shuhua Zhou, Jie Xu, and Yinjiao Ye. 2013. Effect of complimentary and conflicting commentary on sports enjoyment. *International Journal of Sport Communication* 6, 2 (2013), 173–184.

A COMMENTARY EXAMPLES

INTERCEPTION	
Neutral	**Emily**: Oh, that was a poor pass by Jenny Thompson. Kayla Larkin with a crucial interception there for [Team 1]!
Support [Team 1]	**Emily**: Larkin intercepts Thompson's pass! **Doug**: That's what we like to see!
Support [Team 2]	**Emily**: Ugh, Jenny Thompson just had the ball intercepted by Kayla Larkin. Can't believe she let that happen!
Support Rival Teams	**Emily**: Oh no! **Doug**: Typical sloppy play from Thompson
FREE KICK	
Neutral	**Emily**: Looks like [Team 1] has a chance to capitalize on this free kick. **Doug**: [Team 2] will need to defend this well and prevent [Team 1] from getting a scoring opportunity. **Emily**: Jenny Thompson is lining up to take the free kick. She's been deadly from this position before. **Doug**: The [Team 2] defense needs to stay alert and mark their players closely

Continued on next page

Table 3 – continued from previous page

Support [Team 1]	**Emily**: Oh great, another chance for [Team 2] to extend their lead. I can't stand this team. **Doug**: Ugh, [Team 2] always gets lucky. I hope Jenny Thompson messes this up.
Support [Team 2]	**Emily**: Oh, this is a great opportunity for [Team 2]! Jenny Thompson has a powerful shot. Let's see if she can put it to good use! **Doug**: [Team 2] needs to take advantage of this set piece. It could be a chance to extend their lead over [Team 1].
Support Rival Teams	**Emily**: Thompson lines up for the free kick, this could be dangerous for [Team 1]! **Doug**: Ugh, not another set piece opportunity for [Team 2]. Get the wall set up, [Team 1]!
SAVE	 
Neutral	**Emily**: And there's a great save by Rachel Riley! She denied [Team 2] a goal-scoring opportunity! **Doug**: Fantastic reflexes from Riley
Support [Team 1]	**Emily**: What a save by Rachel Riley! **Doug**: She's keeping [Team 1] in the game!
Support [Team 2]	**Emily**: Great save by Rachel Riley to keep [Team 2]'s lead intact! **Doug**: Can't believe [Team 2] actually got a shot on goal. Typical of them to waste the opportunity.
Support Rival Teams	**Emily**: Rachel Riley with a great save there! **Doug**: Finally. **Emily**: Looks like they're trying to turn things around. **Doug**: They better start playing better if they want to catch up. **Emily**: [Team 2]'s defense has been solid so far. **Doug**: [Team 1] needs to step up their game.
DEFLECTION	 
Neutral	**Doug**: Stephanie Price with a crucial deflection there! That could have been dangerous for [Team 2]!

Continued on next page

Table 3 – continued from previous page

Support [Team 1]	**Emily**: Did you see that deflection by Stephanie Price? [Team 2]'s defense is on point today! **Doug**: Ugh, [Team 2] getting lucky again. They can't rely on deflections all game.
Support [Team 2]	**Emily**: And there goes Stephanie Price with a crucial deflection! **Doug**: Great defensive play by Price to deny [Team 1]!
Support Rival Teams	**Emily**: Oh wow, what a deflection by Stephanie Price! That was some quick reflexes! **Doug**: Ugh, lucky deflection. [Team 2] still can't break through.
CROSS	 
Neutral	**Emily**: Anna Johnson with a dangerous cross into the box! Can [Team 1] capitalize on this opportunity?
Support [Team 1]	**Emily**: And Johnson with a dangerous cross into [Team 2]'s box! Can [Team 1] capitalize on this opportunity? **Doug**: They better! We need a goal to level the playing field here!
Support [Team 2]	**Emily**: Oh no, [Team 1] with a dangerous cross into [Team 2]'s box! **Doug**: [Team 2]'s defense needs to step up and clear that!
Support Rival Teams	**Doug**: [Team 2]'s defence needs to tighten up! They can't allow crosses like that into their box.

Table 3. Examples of commentary for specific events and style

B POST-SYSTEM QUESTIONNAIRE

Date:

Participant:

System Mode:

	Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree
The system presented information in a way that I could easily understand	○	○	○	○	○	○	○
The visualisations and commentary helped me better understand the game	○	○	○	○	○	○	○
The system improved my understanding of player performance	○	○	○	○	○	○	○
The system improved my understanding of in-game developments	○	○	○	○	○	○	○
The system made the viewing experience more engaging	○	○	○	○	○	○	○
The system made the soccer match more enjoyable to watch	○	○	○	○	○	○	○
While using the system, I felt immersed in the football match	○	○	○	○	○	○	○
Given the choice, in the future, I would use the system for viewing football matches	○	○	○	○	○	○	○
I would recommend this system to friends and family	○	○	○	○	○	○	○
I was generally satisfied with the system	○	○	○	○	○	○	○
I felt the information provided was reliable	○	○	○	○	○	○	○
I trusted the system's analysis and/or commentary	○	○	○	○	○	○	○
I could count on the system to provide accurate player statistics and information	○	○	○	○	○	○	○
The system was consistent in its feedback	○	○	○	○	○	○	○
I did not encounter conflicting statements or contradictory information	○	○	○	○	○	○	○
The style of the commentary and visualisations remained consistent	○	○	○	○	○	○	○
The system's features were well-integrated	○	○	○	○	○	○	○
The system added value to the traditional football-viewing experience	○	○	○	○	○	○	○
The system was overly complex	○	○	○	○	○	○	○
Overall, I liked the system	○	○	○	○	○	○	○