AiCommentator: A Multimodal Conversational Agent for Embedded Visualization in Football Viewing

Peter Andrews
MediaFutures and t2i Lab, University of Bergen
Bergen, Norway
peter.andrews@uib.no

Oda Nordberg
MediaFutures, University of Bergen
Bergen, Norway
oda.nordberg@uib.no

Stephanie Zubicueta Portales
t2i Lab, University of Bergen
Bergen, Norway
stephanie.portales@student.uib.no

Njål Borch
Schibsted
Oslo, Norway
njaal.borch@gmail.com

Frode Guribye
MediaFutures, University of Bergen
Bergen, Norway
frode.guribye@uib.no

Kazuyuki Fujita
Research Institute of Electrical Communication, Tohoku University
Sendai, Miyagi,, Japan
k-fujita@riec.tohoku.ac.jp

Morten Fjeld
MediaFutures and t2i Lab, University of Bergen, Norway and Chalmers
Sweden
morten.fjeld@uib.no

Figure 1: AiCommentator, a Multimodal Conversational Agent (MCA), enhances football video content by providing both automated non-interactive and interactive commentary. AI commentary is supported with embedded visualizations to facilitate better user comprehension of player performance and in-game events. Users can monitor or interact with the commentators through a Discord bot using natural language or a menu-based system over a secondary device to trigger specific functionality on the primary display.

IUI '24, March 18–21, 2024, Greenville, SC, USA

ACM ISBN 979-8-4007-0508-3/24/03
https://doi.org/10.1145/3640543.3645197

ABSTRACT

Traditionally, sports commentators provide viewers with diverse information, encompassing in-game developments and player performances. Yet young adult football viewers increasingly use mobile devices for deeper insights during football matches. Such insights into players on the pitch and performance statistics support viewers' understanding of game stakes, creating a more engaging viewing

experience. Inspired by commentators' traditional roles and to incorporate information into a single platform, we developed AiCommentator, a Multimodal Conversational Agent (MCA) for embedded visualization and conversational interactions in football broadcast video. AiCommentator integrates embedded visualization, either with an automated non-interactive or with a responsive interactive commentary mode. Our system builds upon multimodal techniques, integrating computer vision and large language models, to demonstrate ways for designing tailored, interactive sports-viewing content. AiCommentator's event system infers game states based on a multi-object tracking algorithm and computer vision backend, facilitating automated responsive commentary. We address three key topics: evaluating young adults' satisfaction and immersion across the two viewing modes, enhancing viewer understanding of in-game events and players on the pitch, and devising methods to present this information in a usable manner. In a mixed-method evaluation (n=16) of AiCommentator, we found that the participants appreciated aspects of both system modes but preferred the interactive mode, expressing a higher degree of engagement and satisfaction. Our paper reports on our development of AiCommentator and presents the results from our user study, demonstrating the promise of interactive MCA for a more engaging sports viewing experience. Systems like AiCommentator could be pivotal in transforming the interactivity and accessibility of sports content, revolutionizing how sports viewers engage with video content.

CCS CONCEPTS

• **Human-centered computing** → **Information visualization**; **Natural language interfaces**; *Usability testing*; • **Computing methodologies** → *Tracking*; *Object detection*.

KEYWORDS

Embedded Visualization, Multimodal Conversational Agent, Conversational User Interface, Usability Testing, Human-Computer Interaction, Computer Vision, Deep Learning, Multi-Object Tracking

ACM Reference Format:
Peter Andrews, Oda Nordberg, Stephanie Zubicueta Portales, Njål Borch, Frode Guribye, Kazuyuki Fujita, and Morten Fjeld. 2024. AiCommentator: A Multimodal Conversational Agent for Embedded Visualization in Football Viewing. In *29th International Conference on Intelligent User Interfaces (IUI '24), March 18–21, 2024, Greenville, SC, USA*. ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/3640543.3645197

1 INTRODUCTION

Football boasts a global following, with over five billion fans worldwide, stretching across regions like Europe, Latin America, the Middle East, and Africa [12]. This fan base is rising and includes an increased interest in previously underrepresented leagues. For instance, FIFA Women's World Cup 2023 witnessed a viewing surge from 1.12 billion in 2019 to 2 billion in 2023 [16]. As global interest increases, so does the technological shift in how viewers consume football. FIFA's 2018 World Cup data showed that 77% of home viewers supplemented their TV match-watching experience by using smartphones or tablets [12]. Pfeffel et al. [28] underscore this behavior, noting that football enthusiasts often resort to secondary devices to seek functional information, such as game statistics, thereby enriching their contextual understanding of the ongoing match. Although searching for information across multiple platforms can make the viewing experience more engaging, it can sometimes be distracting, with the potential for viewers to miss real-time match developments. However, it's worth noting that some viewers appreciate the second-screen experience as it can provide a private information space, especially in a social setting. A potential solution could be for broadcasters to adapt to varied viewers' preferences by offering the option to integrate this functional information directly into the official experience.

Embedded visualizations offer a novel approach to augmenting video streams with relevant information, enhancing content engagement. Although sports broadcasts have predominantly used these visualizations for professional analysts, a significant research gap remains concerning their effect on the overall viewing experience. Current research regarding embedded visualization for sports has mainly focused on the visual and usability aspects of the system [5, 24]. Whether users prefer active, interactive viewing over its traditional passive counterpart remains to be seen. Furthermore, there has been limited research on how embedded visualizations can support viewers' knowledge of players on the pitch and their teams. In most sports, commentators typically assume this role, providing contextual information to heighten the viewer's enjoyment, satisfaction, and perceived quality of content [23].

Our work aims to bridge the above-mentioned research gaps by redesigning the commentary role and adding multimodal elements. We propose a novel method of supporting embedded visualization with a conversational agent that takes on the persona of two commentators. By reconceptualizing commentators as conversational agents, we retain their traditional qualities while introducing an adaptable user experience enriched by interactive engagements. Embedded visualizations under this system become a form of "italicizing", using visuals with commentary to focus the viewer's attention [22, 26, 46].

To our knowledge, our system AiCommenatator is the first to provide interactive commentary for sports media, thereby adding significant novelty to the field. We validated such technology's usability with young adults to understand its potential. Young adults are among the highest age groups to use mobile devices while watching football [28] and are more adept at multi-tasking with mobile technology to access additional game content [12]. Being digital natives, this user group's familiarity with new technology makes them suitable candidates for testing new technological prototypes [1]. To assess the usability, we conducted a user study that considers this technology's user experience and its potential for enhancing viewer understanding and immersion. We aimed to understand whether such a system enhances viewers' enjoyment and satisfaction and how it may influence their understanding of the game. In particular, within the context of two cognitive antecedents – team identification and quality of opponent – documented to increase viewer satisfaction in sports [25]. Addressing these questions is vital in order to design and optimize our system, contributing to a

broader understanding of interactive media in sports broadcasting. This paper reports our comprehensive user study to answer the following Research Questions (RQs):

RQ1) Which mode of AiCommentator, non-interactive or interactive, offers the user a higher level of engagement and satisfaction?
RQ2) How can the two alternative modes of AiCommentator, non-interactive and interactive, support young adult viewers' knowledge of players on the pitch and their performance?
RQ3) How do young adults perceive the usability of the interactive mode of AiCommentator?

To answer these research questions, we conducted a comprehensive mixed-methods evaluation of AiCommentator with sixteen participants. Our user study employed a within-group design (AB), ensuring all participants engaged with both system modes. We gathered quantitative data on users' perceptions of the system and its functions from post-function questionnaires, post-system questionnaires, and the System Usability Scale (SUS). For qualitative insights, we sourced information from pre-study questionnaires, video-recorded full-testing sessions, and post-study interviews.

In summary, we present three key contributions: 1) AiCommentator, a Multimodal Conversational Agent (MCA) that provides visual feedback of real-time and historical in-game statistics and player locations, facilitated by text-based interactions with a Discord bot; 2) Automated sports commentary to communicate real-time game developments while engaging the users conversationally; 3) A comprehensive study evaluating the usability of an MCA to modernize the sports viewing experience.

This paper is structured as follows: We start by presenting related work on embedded visualization, automated commentary, and conversational agents. We then delve into the design of AiCommentator, followed by a detailed presentation of the system. Subsequently, we describe the user study and present the results. In the discussion section, we analyze these results against our three research questions and provide design recommendations for future work. Finally, we conclude by summarizing the key takeaways, acknowledging study limitations, and proposing potential avenues for future research.

2 RELATED WORK

In this section, we present related work on embedded visualizations, automated commentary, and conversational agents for sports viewing.

2.1 Embedded Visualizations in Sports Viewing

While rooted in visual analytics, embedded visualizations integrate data-driven graphical content within diverse platforms and applications, enhancing comprehension and offering opportunities for more interactive user experiences. Bolstered by developments in computer vision and data mining techniques, embedded visualizations offer a more complex analysis of movement and trajectory, more profound insights into game dynamics and patterns, and a detailed examination of team formation analysis.

Traditionally, tools for sports game analysis supported analysts who aimed to delve deep into game developments and statistical data. However, with the advent of systems like Viz Libero [43] and Piero [33], there has been a noticeable shift in this paradigm. These professional systems now support video editors in crafting engaging visual content. Complementing this change, recent studies by Chen et al. [5] and Lin et al. [24] highlight a growing emphasis on serving a wider audience, reflecting the evolving dynamics of embedded visualizations in sports.

The foundation of embedded visualization in sports is movement and trajectory, which convey spatial-temporal patterns that offer insights into player and team performance and strategies. Shuttlespace [50] demonstrates an innovative application of this possibility and seeks to reduce the cognitive load by visualizing 2D badminton strokes in Virtual Reality (VR). Courtvision [13] enabled the measurement of basketball shot precision by analyzing distributions of positional 2D data points over a five-year period. Similarly, Snapshot [29] utilized data from the 2010-2011 hockey season, presenting it through diverse visualization methods, notably radial heatmaps. These visualizations were further refined using metadata filters, enhancing the analytical utility of the system. To visualize movement patterns within video content, Stein et al. [38] tracked football players while applying reverse perspective transformations to analyze localized performance.

Researchers recently used these cartesian coordinate systems to infer game dynamics and patterns. Both Stein et al. [36] and Xie et al. [49] implemented data mining techniques to extract spatio-temporal patterns from 2D coordinates to infer dynamic developments in football matches. In PassVisor, Xie et al. [49] utilized topic-based pattern detection to gain insights into passing patterns in football matches, whereas Stein et al. [36] classified events of interest with a feature ranker to help experts understand the relevance of each event. Stein et al. [35] later built upon classification systems to contextualize events within the scope of predefined scenarios, resulting in more intricate explanations. Beyond mining techniques, Stein et al. [37] also developed a system to find interaction spaces, free spaces, and pass options from spatial information such as player clusters, distance, direction, and a grid-based free-space algorithm. Moving from an individualized performance rating to a team-based one, Forvisor [48] used a clustering algorithm with the Hungarian algorithm to assign football player identities and map them to team formations.

While many existing methods function as interactive tools for analysts, a smaller subset delves into interactive embedded visualizations explicitly tailored for video content targeting the general viewer. Chen et al. [5] introduce gaze-moderated interactions designed to guide viewers through the intricacies of basketball strategies. Meanwhile, Lin et al. [24] reformulate simulated basketball content to allow users to interact with statistical and tactical data. Both studies demonstrate that embedded visualizations engage viewers, aiding in the communication of insights from statistical data in the case of Lin et al. and insights from computer vision data in the case of Chen et al.

2.2 Automated Commentary

In this section, our primary focus revolves around two sports commentary types: play-by-play and color commentary. While play-by-play commentary offers continuous feedback on in-game developments and events, color commentary provides game context via

statistical information, popular news stories, and player information, typically filling uneventful spaces during the game when the action is minimal.

Original research in automated commentary dates back to the 1990s and the RoboCup simulated sports datasets via SoccerServer. Systems like ROCCO [44] and MIKE [39, 40] output natural language by building templates filled with specific event attributes and use a pooling or selection processes to determine output timing and priority. MIKE overcame commentary pacing issues with a pooling system that abbreviated or interrupted templates. The Byrne system [2] built upon the advancements in automated live sports television commentary to introduce speech synthesis and facial animation, mapping templates to animations to infer emotions. Zheng et al. [51] overcame the limits of rule-based event systems by combining C4.5 decision tree algorithm [32], Naïve Bayes, and K-Nearest Neighbor to classify more complex events. However, the system was limited in the range of events, failing to encapsulate the whole dynamics of the game. Moreover, defining natural language by templates provided limited dynamic language for the commentary [2, 39, 40, 44].

Lee et al. [22], Chitrakala et al. [7], and Prathibha [30] all framed the color commentary task as an information retrieval problem. Lee et al.'s system, SCoRes, used feature vectors to represent attributes such as the score and teams playing to employ a machine learning system with the information retrieval system AdaRank to rank suitable articles. Chitrakala et al. used a similar methodology but with different implementations, evaluating the top stories based on metrics such as winner-takes-all and Normalized Discounted Cumulative Gain (NDCG). Both SCoRes and Chitrakala et al.'s systems increased enjoyment by recommending contextually relevant stories. Prathibha further enriched color commentary context by incorporating a video processing module for play-by-play information. This system provided keywords to assist commentators in developing detailed and engaging explanations easily, enhancing the overall quality and depth of sports commentary.

More recently, NHK Science and Technology Research Laboratories have pioneered recent research in audio descriptions for the visually impaired [15, 18, 19]. Utilizing metadata from live events, Kurihara et al. [19] created a prototype using a Speech to Text (STT) model to generate audio descriptions, testing it at the 2016 Olympic and Paralympic Games. The system maintained the latest data, detected facts, composed sentences, and updated past facts. Ichiki et al. [15] compared these descriptors with live broadcasts and found equal effectiveness, with 80% of the participants reporting an improvement in understanding after adjusting overlapping audio tracks. Kumano et al. [18] further improved understanding for sighted and visually impaired participants by reinterpreting metadata through a 'belief' system, which dynamically adjusted information based on the user's current knowledge and was updated as needed. Other efforts in tennis automated audio descriptions mapped ball tracking to 3D binaural audio to represent ball placement [14]. While research showed the need to improve audio descriptions for the visually impaired, the results were inconclusive with respect to how 3D binaural audio could improve this experience.

2.3 Conversational User Interface in Sports Viewing

Conversational User Interfaces (CUIs) encompass a range of technologies designed to provide users with access to data and services via natural language dialogues [**?**]. The primary goal of these interfaces is to mimic some degree of human-like conversational ability, making interactions with technology more straightforward and intuitive [**?**]. Conversational interactions can be written (such as with a chatbot), spoken (such as a voice-based digital assistant), or a combination of several modalities (such as when typing a question to an interactive podcast (e.g., [21]), and getting both a textual and auditory answer). We often distinguish between rule-based and generative conversational interfaces. The first provides predefined answers and the second generates answers from deep learning models [20], such as large language models.

In the field of human–computer interaction (HCI), multimodality refers to interactive technologies where the user receives stimuli from several senses, such as sight, hearing and touch, and where systems use several output channels, such as text, sounds and video [42]. When interactions with CUIs and the information they can provide become increasingly complex, the integration of additional modalities becomes more common [8]. MCAs normally combine visual content with the natural language interface [8].

Several attempts have been made to use a second screen, such as a phone or a tablet [12], during the viewing of sports matches, for example by using conversational interfaces, such as a chatbot. In addition to sports chatbots developed by the industry, chatbots have also been explored by several researchers. Segura et al. [34] developed Chatbol, a social chatbot where users can interact with it through text to ask general questions about the Spanish football league La Liga. Users can ask questions such as who is playing and in which stadium they are playing. Zhi et al. [52] developed GameBot, a visual-augmented sports chatbot, as a means of providing users with statistical data during a match; GameBot also includes data visualizations as a supporting means. Even though Sporthesia by Chen et al. [6] was developed for sports analysts, the results of their study suggested a potential for sports viewers to type commands in order to receive embedded visualizations during sports viewing, which, they speculate, could improve the viewing experience for regular sports audiences.

Several terms are used to describe CUI technologies, however, in this paper, we'll use conversational agent and MCA to emphasize our focus on the commentators and the capabilities of the multimodal interface. The foundation of our MCA is adapting automated commentary to become interactive and incorporating embedded visualizations. While our work does not directly contribute to the field of automated commentary, we draw inspiration from it to design interactive commentary. Building upon the foundational research of Chen et al. [5] and Lin et al. [24], we integrate play-by-play and color commentary with embedded visualizations, providing a multimodal feedback experience. Infusing these interactive elements, we aim to elevate the traditional linear viewing experience into a dynamic and engaging experience for viewers by developing an MCA.

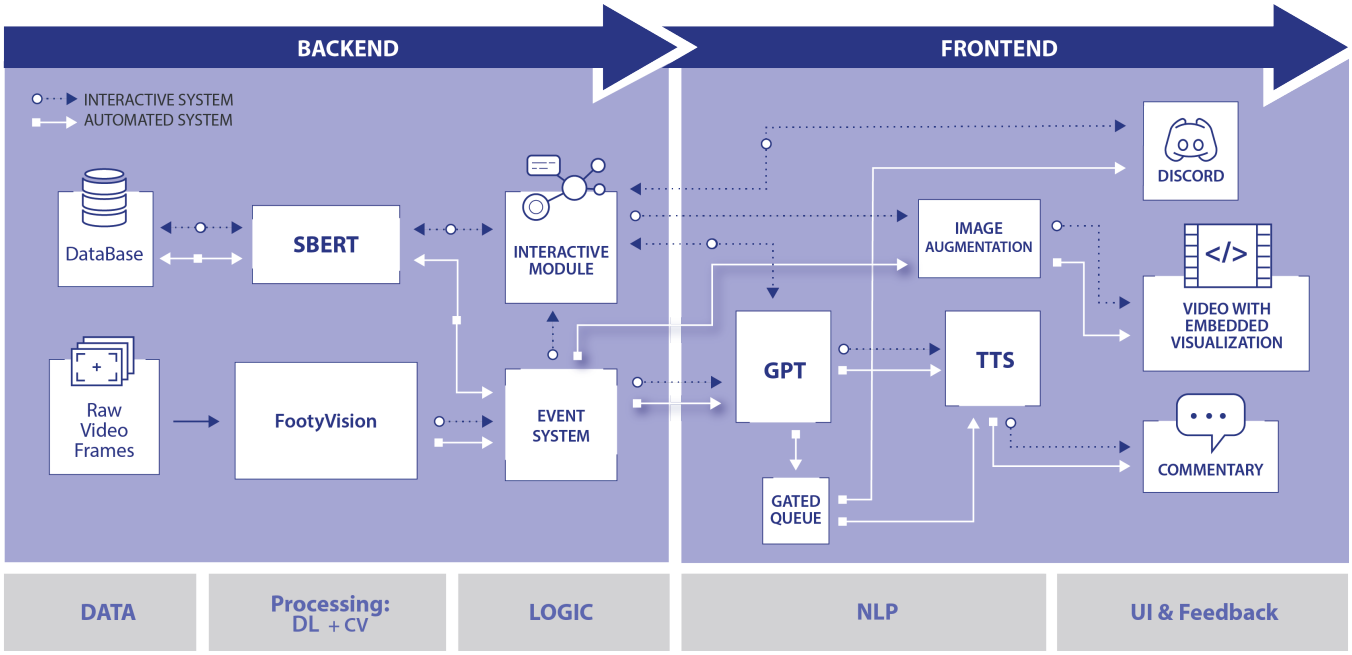


Figure 2: The automated commentary system consists of a backend and frontend system which leverages data from Multi-Object Tracking (MOT) and perspective transformation matrices preprocessed by the FootyVision system. An event system infers events based on the output from FootyVision and retrieves relevant data by cross-referencing tracking identities with a database containing information regarding players and their respective teams. We employ feature embeddings from Sentence Bidirectional Encoder Representations from Transformers (SBERT) to yield the most applicable context data for the GPT module. As well as informing the GPT module, the event system notifies the image augmentation module for embedded visualizations. Output from the GPT module is added to a gated queue which releases commentatary to the Text to Speech (TTS) and discord module based on priority and lifespan. The result is video with embedded visualizations, auditory commentary, and commentary updates to a mobile device through Discord.

3 AICOMMENTATOR SYSTEM DESIGN

The AiCommentator system was designed to offer embedded visualizations and automated commentary in sports viewing while operating in one of two alternative modes. Figure 2 outlines the AiCommentator system while showing the pathway of both modes. We now describe each mode in more detail.

- **Non-interactive Mode**: In this mode, AiCommentator autonomously generates commentary, emulating the feel of a traditional, non-interactive viewing experience. Output from our CV and DL model, FootyVision (refer to Section 3.1), is processed within the event system. This gives context to the commentary generated by Generative Pre-trained Transformer (GPT) version 3.5 Turbo 0613. Commentary dynamically adjusts based on the in-game events, with embedded visualizations mainly focused on the player currently in possession.
- **Interactive Mode**: This mode fosters user engagement. Here, viewers can interact with the system via a Discord bot on a mobile device. Based on user input, the system searches the database to find the closest match with Sentence Bidirectional Encoder Representations from Transformers (SBERT) encodings. Upon identifying the relevant context, the query, now augmented with additional context from a pre-defined gallery, is directed to the GPT-based class. Depending on user input, specific functions are activated, leading to the generation of events that are then managed within the interactive framework.

3.1 FootyVision

Comprehending the video's context is paramount for seamless embedded visualization in video content and appropriate feedback from a conversational agent. To derive context and information from uncalibrated video data, we implemented our model FootyVision (in press, 2024). FootyVison (see Figure 3) serves as an all-inclusive model for identifying, tracking, and localizing the players and the ball during a football match. It was built upon a YOLOv7 [45] backbone and trained on the SoccerNet [9] and ISSIA [10] datasets to achieve state-of-the-art player and ball detection. On top of the YOLOv7 backbone, there are two supplementary modules:

- **Multi-Object Tracking (MOT) Module**: This module's primary function is to maintain a consistent ID for each player throughout video segments. It does so by identifying the task as an assignment problem and using the Hungarian algorithm, an optimization method, to match identities over consecutive frames. The Hungarian algorithm is designed to find the optimal assignment that minimizes the

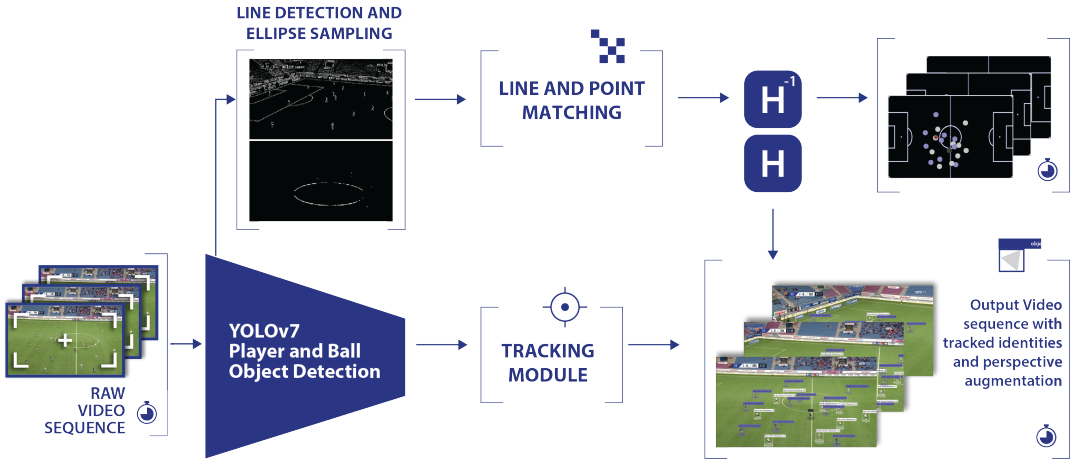


Figure 3: FootyVision is an all-in-one model for player and ball Multi-Object Tracking (MOT) and localization. YOLOv7 [45] serves as a backend object detection network and intermediate layers provide visual context to compute homographies. The tracking module assigns and retains identities to players detected on the pitch.

total cost from a cost matrix C. In our case $C = \lambda_{\text{feat}}(1 - J) + \lambda_{\text{iou}} \cos(\theta) + \lambda_{\text{dist}}(|c_x - c_y|)^2 + \lambda_{\text{vel}} V$, where J is the Jaccard Index or Intersection-over-Union (IoU) of the detected bounding boxes, $\cos(\theta)$ is the cosine similarity of feature embeddings derived from Wieczorek et al. [47], $(|c_x - c_y|)^2)$ is the Euclidean distance among multiple bounding box centroids, and V is velocity. The lambda coefficients represent a weight for each attribute, which each contributes to the sum of one.

- **Perspective Transformation Module**: This module computes the homography matrix H using lines and ellipses from uncalibrated camera data. The transformation matrix H projects graphics into the camera perspective space, while the inverse H^{-1} maps the player locations onto a standardized football pitch template. These mapped template-space locations play a crucial role in our event detection system.

FootyVision outputs several key pieces of data that serve as the foundation for our event and embedded visualization systems. The dataset includes player and ball bounding boxes, localized points within the football pitch template space, a list of track identities, color assignments for team identification, and the homography transformation matrix. Currently, FootyVision does not achieve real-time inference speeds and does not assign player names to the track identities. Therefore, we postprocess and clean the data from FootyVision before streaming it into the real-time AiCommentator modules. We employ a Kalman filter to smooth the trajectories of the cartesian coordinates, enhancing the data's precision. Additionally, track switches were rectified, identities were merged, and each track was manually labelled to ensure accuracy. The cleaned data was then streamed into AiCommentator for real-time inference.

3.2 Event System

The event detection system aims to infer in-game events from the data derived from FootyVision described in Section 3.1. Using spatial-temporal data from FootyVision, we developed rule-based logic to trigger events within the system. Our event system processes data from the main camera view, which consists of a wide-angle game viewpoint suitable for play-by-play commentary. When no perspective transformation matrix accompanies the streamed frame, we classify the second viewpoint as a cutscene. The cutscene view actively captures closeups of players between gameplay, acting as a trigger for color commentary. The event system serves as context for the conversational agent used for automated and interactive play-by-play commentary. The core infrastructure of the event-system framework can be condensed into two core components we cover next.

3.2.1 Detecting collisions. Detecting collisions has been widely used in game development and computer graphics, with many applications using the Separating Axis Theorem (SAT) to determine whether two convex shapes intersect. Inspired by SAT, we implemented a more direct version which calculates the minimum and maximum x and y coordinates for each bounding box and then checks for overlaps along the x-axis and y-axis individually. Mathematically, the overlaps along x and y are determined using logical AND operations as follows:

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

If both x_overlap and y_overlap are `True`, then a collision is confirmed. This method of detecting collisions serves as a way to determine which player has possession of the ball at any one time. To overcome false positive possession events, we set a time threshold parameter which disregards possession below the threshold. To handle instances where the ball collides with multiple player bounding boxes, we prioritize the bounding box of the player currently possessing the ball. Otherwise, we measure the ball's distance with the center base of the other colliding boxes to determine the closest player.

3.2.2 Spatial and Trajectory Analysis. To enhance understanding of in-game dynamics and contribute to richer contextualization,

spatial data is acquired from projecting bounding box locations into the template view via the inverse of the homography matrix. This spatial data supplements visual cues and allows for nuanced event detection. Specific events derived from the spatial and trajectory analysis were identified for the practical implementation of this concept. Based on predetermined logical conditions, these events serve as vital markers during the match, shedding light on pivotal gameplay moments. Table 1 elucidates these specific events, their descriptions, and the underlying logic employed for their detection. Each event is recognized based on distinct conditions, which rely on the spatial relationships and trajectories of both the players and the ball. Figure 4 displays the trajectory analysis from the ball's perspective while accounting for distance attenuation. Here, the green line represents the ball's trajectory, whereas the red lines show the ball's visual field. Players highlighted within the visual field are within the path of the ball's trajectory.

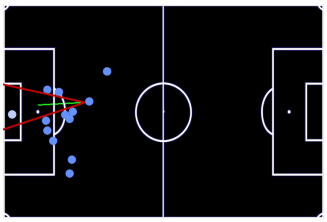


Figure 4: Players and the ball are localized on a football pitch template using an inverse transformation derived from the homography matrix provided by FootyVision. Players appear as blue dots, while a green line illustrates the ball's trajectory. Red lines depict the ball's visual field; players highlighted in light blue are within the trajectory's range.

3.2.3 GPT3.5-Turbo 0613. We utilized the GPT3.5-Turbo 0613 model for our natural language processing tasks. We opted for GPT-3.5 over GPT-4 primarily because GPT-3.5 offered faster processing speeds and was more cost-effective. We tasked GPT-3.5 with three distinct functions:

- Generating natural language responses based on predefined template prompts for automated commentary.
- Producing natural language feedback in response to user queries.
- Function calling with specified parameters.

We established commentator profiles within the system context of the GPT-3.5 model. For the purposes of our experiment, we designated two AI commentators: "Emily" and "Doug". Their profiles were crafted to generate dialogues reminiscent of sports commentators, relying on real-time information provided by the system. The GPT-3.5 6080 variant allows for the predefinition of functions that can be invoked live based on user input. We established six such functions: "Highlight Team", "Track Player", "Seasonal Statistics", "In-Game Statistics", "Recap Events", and "Query". The specifics of these functions will be elaborated upon in subsequent sections.

3.3 Automated Commentary

As Section 2.2 highlights, sports commentary consists of play-by-play and color commentary. Our automated system generates commentary by subscribing to an event system which maps each occurring event to a set of predefined template prompts. Frequently occurring events, like possession changes, are associated with multiple variations of the template, selected randomly to enhance diversity. To ensure commentary timely delivery, the automated module preprocesses the video, outputting the commentary dialogue with a timestamp to a JSON file. This ensured the automated commentary was the same for each participant in the following user study. The commentary was streamed from JSON files during real-time inference and released via a queuing system, as explained below.

3.3.1 Play-by-Play Commentary. Play-by-play commentary keeps the viewer informed of ongoing in-game events. The AiCommentator's event system notifies the language processing module upon event detection. This module then processes relevant arguments and fits them into predefined templates based on the events listed in Table 1. These templates act as input prompts for GPT, generating dialogue reminiscent of sports commentators. Any dialogue not terminating in proper punctuation or with a concluding sentence below a predetermined length threshold is trimmed or discarded.

3.3.2 Color Commentary. As established earlier, color commentary delves deeper, offering insights into player and team statistics during the match. Since color commentary often fills moments of reduced game-play activity, our system triggers it during cutscenes, which, in our dataset, usually indicates the ball is out of play. During these moments, the automated commentary module randomly selects a player involved in the tracked cutscene and enters their name into a randomized "Query" template. Examples include queries like "how well has [name] performed this season?" or "share an interesting fact about [name]". We provided GPT with additional context, including background information regarding each player and club with their respective statistics. The data is transformed into 768-dimensional feature vectors using the SBERT. Feature embeddings were extracted from incoming player names or teams and compared to the gallery using cosine similarity. The item in the gallery with the highest similarity score provided context for the GPT model.

3.3.3 Event Pooling. Before distribution to the Text to Speech (TTS) module, each event undergoes a pooling process to prioritize the dialogue. Our system uses a gated queue that releases events when the TTS module is free. Each event enters the pooling module equipped with dialogue, a timestamp, lifespan, and priority, defined together as $E = \{dialogue, timestamp, lifespan, priority\}$. Operating on a separate thread, the gated queue discards expired events and reorders the queue based on priority. We designed the system to allow certain events, specifically "Shot Event", "Cross Event", and "Box Event", to interrupt the TTS module due to their significance. Additionally, we prioritize "play-by-play" events over "Query" events, meaning "play-by-play" events will interrupt "Query" events. We chose this design because commentators generally prioritize unfolding events over color commentary.

Event Name	Description	Logic
Possession Event	Identifies player with current possession	Check if the ball bounding box collides with a player bounding box over a set threshold of frames. Possession remains if the collision ends but the ball remains within a distance threshold.
Pass Event	Detects a pass event	When a player's possession cycle ends and the next player to gain possession is a member of the same team, this constitutes a pass.
Interception Event	Detects an interception	When a player's possession cycle ends and the next player to gain possession is of the opposite team.
Free Kick Event	Detects a free-kick event	Triggered when the ball and players are static on the pitch for a period of time and a player has possession.
Throw Event	Detects a throw event	If a cutscene has occurred and the ball enters the pitch from the sideline, this is a throw in.
Deflection Event	Analyzes the ball's trajectory to find when a deflection occurred	When the ball's trajectory vector significantly changes direction and the closest player to the ball's location is from a team different than the one currently possessing the ball.
Challenge Event	Detects when two players fight for possession of the ball	Identifies two opposing players when the ball trajectory vector rotates beyond a predefined threshold.
Open Ball Event	Determines when the ball is loose on the field	The system fits a spline to the ball's historical location data. If no players are found within the predicted trajectory of this spline, the event is triggered.
Box Event	Detects when a player with possession is in the opposing team's box	The player with possession enters the opposing team's box.
Cross Event	Detects a cross from the sideline into the box	When the player with possession is located on the sideline and the ball accelerates towards the box/goal.
Shot Event	Detects a shot at the opposing team's goal	When a player with possession is within a distance threshold from the goal and the ball accelerates towards the goal.

Table 1: List of events and logic base behind detection. Events are inferred through a logic-based understanding of data derived from player and ball-localized 2D coordinates and bounding box intersections.

4 MULTIMODAL CONVERSATIONAL AGENT FOR INTERACTIVE COMMENTARY AND EMBEDDED VISUALIZATIONS

In this section, we delve into the design and execution of the interactive MCA. We discuss the architecture of the CUI, highlight our user-focused design process through pilot studies, and present a comprehensive design framework.

4.1 Conversational User Interface (CUI)

The CUI facilitates direct communication with the AI Commentators through a secondary device and the activation of specific functions for embedded visualizations. Similar to the non-interactive system, our CUI employs GPT3.5-Turbo 0613 to generate feedback in response to prompts verbalized by Google TTS. However, unlike the non-interactive system, the prompts are generated by the users. All user commands are treated as "Query" events by the system and are then routed to appropriate functions.

We developed a Discord bot named AiCommentator to enable interaction with the CUI. Discord was our platform of choice, given its popularity among our target demographic, the array of interactive options, and its ease of integration. The Discord channel logs interactions and feedback from the AI Commentators through dialogue. The dialogue is verbalized through the primary display, which also shows the video stream with embedded visualizations.

4.2 Pilot Studies

Our development process was user-centered, emphasizing iterative refinement based on user feedback. We undertook various pilot studies with two user groups: two participants with little knowledge of football and another four participants with a deep understanding of football. These studies helped us refine the prototype to prepare for the full user study.

Originally, the system had eight functions, which included live heatmaps and birdseye player tracking views. However, feedback from the pilot study revealed that neither user group found these visual aids valuable for player identification or event tracking so they were removed, leaving six core functions.

Another insight from the pilot was the users found it challenging to formulate phrases in their second language to trigger the desired response from the system. During the interactive guided walkthrough, the users learned a set of structures and keywords to

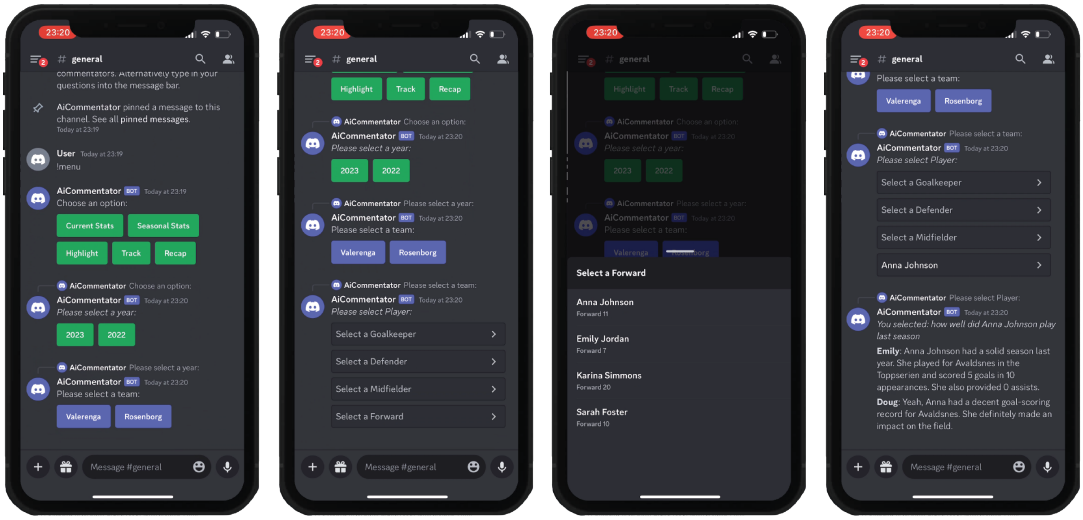


Figure 5: To trigger functions users can navigate a menu system to automate the prompt to the commentator. This provides a more accessible manner of triggering the functionality of the AiCommentator's interactive mode.

activate the system functions. However, during the interactive full-testing phase, participants struggled with remembering the structure and keywords, requiring assistance in formulating phrases. As a result, they became disengaged with the primary display, instead focusing on the mobile device. We introduced a menu system to address this problem, providing structured interaction options (Figure 5). However, to preserve the system's flexibility, we retained the option for users to pose open-ended questions using natural language.

4.3 Design Framework

Our model follows a context-driven methodology to tailor the system response to the user's request (Table 2). Considering each of the function calls, the framework follows a pattern. *Based on the* **interaction** {☰ ⌨}, *the database is queried to provide the appropriate* **context data** {⛁ ⌖ ⚑} *for a* **prompt** *that is either user-defined or developed from a template, providing responsive* **multimodal feedback** {🖵 🔊 }. Further elucidation of these core components is as follows:

- **Interaction**: As discussed previously in Section 4.2, the interactive mode of AiCommentator offers two types of interaction. The first approach is natural language ⌨, which allows the user to type specific queries to the GPT module. For example, the user can ask *"compare Rosenborgs' performance over previous seasons"* or *"tell me some interesting facts about Eleanor Rigby"*, and the system will respond according to a "Query" event. The second interaction approach automates the prompt creation by navigating the user through a series of menus ☰.
- **Context Data**: To ensure that AI commentators deliver insightful feedback, providing context that assists GPT in generating relevant responses is essential. The event system offers three key sources for this context. First, there is a database ⛁, which houses details about a player's background, individual performance, and team's seasonal statistics. Second, the tracking data ⌖ offers precise locations on the pitch. Finally, the event data ⚑ provides context, offering specific event information and on-the-fly statistics collected for each player.
- **Prompt**: Depending on the chosen mode of interaction, either natural language ⌨ or menus ☰, the input directed to the GPT module is derived directly from a user's interaction or a predefined template, respectively. Interacting with Discord menus ☰ navigates the user through a series of questions where the output is mapped to the input template's relevant sections.
- **Feedback**: The system's multimodal feedback encompasses embedded visualizations 🖵, TTS audio 🔊, and text messages sent to the Discord channel . Five of the six available functions provide embedded visualizations 🖵 relevant to user queries. In contrast, the "Query" function abstains from offering such visuals due to the myriad of potential user interactions. Future research could probe into refining visualizations to align more closely with the variety of user requests.

4.4 Interactive Functions

Concerning RQ2, our objective was to assess whether an MCA could enhance a user's comprehension of players on the field, their respective performance and in-game events. The interactive functions are instrumental in delivering the content that underpins this understanding. We now describe each function in detail.

4.4.1 Query. The "Query" function facilitates direct interaction with the AI commentators, preserving the user's agency to pose open-ended queries. Users employ natural language ⌨ to articulate their inquiries. The system will use data from both the database ⛁ and event system ⚑ to provide feedback to the user. Both auditory

Function	Interaction	Context Data	Prompt Example	Feedback	Embedded Visualization
Query	⌨	🗄	*'Which team has been performing better this season?'*	🔊 Discord	
Highlight Team	☰	⌖	*'Highlight [Team Name]'*	🖵 🔊 Discord	Tracking with labels
Track Player	☰	⌖	*'Track [Player Name]'*	🖵 🔊 Discord	Spotlight, labels, and player card
Seasonal Statistics	☰	🗄 ⌖	*'How well has [player name] been performing [this/last] season?'*	🖵 🔊 Discord	Player highlight, season statistics box, and player card
In-Game Statistics	☰	⌖ ⚑	*'How well is [player name] performing in this current match?'*	🖵 🔊 Discord	Player highlight, current statistics boxes, and player cards
Recap	☰	⚑	*'Recap Events'*	🖵 🔊 Discord	Tracking and labels

Table 2: Design Framework of AiCommentator's Interactive Mode Functions: Interaction types include Natural Language ⌨ and Menu ☰. Context Data Sources comprise Database 🗄, Tracking Data ⌖, and Event Data ⚑. Output Modalities feature Embedded Visualizations 🖵, Text to Speech (TTS) Audio 🔊, and Text Messages relayed to AiCommentator's Discord Channel Discord.

and textual feedback is relayed through the AI commentators and displayed on the Discord channel.

4.4.2 Highlight Team. This function gives users an overarching view of each team, their formations, and elementary player details. Embedded visualizations augments tracked bounding boxes for every player within a predefined team and superimposes the player's name, jersey number, and on-field position above the bounding box. Users can call the function by navigating through the Discord channel menu ☰ while selecting the desired team. The tracking data ⌖ offers context, pinpointing player locations. AI commentator feedback acknowledges the user's selection and imparts knowledge about the players and their designated positions.

4.4.3 Track Player. During high-action or obscured moments, identifying players can be challenging. The "Track Player" function tackles this by pinpointing players and showcasing a player card with their image, name, nationality, team, number, age, and height (Figure 6b). This feature aids viewers in player recognition by highlighting the targeted player with a spotlight, a technique inspired by sports analysts. Users activate it via the menu system ☰, choosing a team and player sorted by position. The AI commentators use the tracking data ⌖ to communicate the player's last known position.

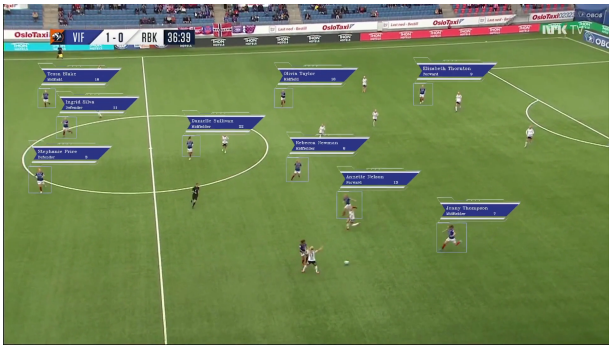
4.4.4 Seasonal Statistics. The "Seasonal Statistics" function offers a comprehensive statistical breakdown to elucidate a player's seasonal performance, detailing metrics such as goals scored, assists made, games started, and minutes played. If the player is tracked, the visualizations are augmented to the player's current position (Figure 6c). Otherwise, the visualization moves beneath the player's card. The system retrieves relevant data from the database 🗄 to generate a player performance table, which, combined with commentary, offers holistic, multimodal feedback.

4.4.5 In-Game Statistics. The "In-Game Statistics" function imparts real-time performance data of players using context from the event system ⚑. The event system collects player data such as successful pass rate, total time in possession, distance covered, and speed. When called, the function augments the data into the video stream, and the AI commentators summarize the data to provide a play-by-play performance summary.

4.4.6 Recap Events. The "Recap Events" function, devised to ensure that users remain informed of recent events, offers a synopsis of the preceding five events. Activation via the menu ☰ prompts visual highlights of players involved in these events, coupled with commentary elucidating previous events (Figure 6e). Future iterations of the prototype will allow for dynamic user-defined recap lengths. However, for our study, five events provided enough information for relatively short video sequences.

5 USER STUDY

To assess young adults' experience of AiCommentator, we conducted a mixed-method user study involving sixteen participants split into equal groups. The study followed a within-group design (AB), with all participants engaging in interactive and non-interactive sessions. The order of which session they interacted with first depended on their assigned group to counterbalance the order effect.

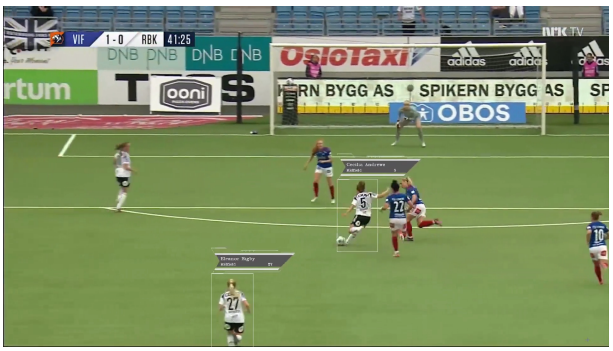


(a) Highlight Team Function

(b) Track Player Function

(c) Seasonal Stats Function

(d) Current Stats Function



(e) Recap Function

Figure 6: Five embedded visualization functions of the AiCommentator: (a) Highlight Team, (b) Track Player, (c) Seasonal Stats, (d) Current Stats, and (e) Recap. Subfigures (b, c, d) include an inset image credited to Beate Oma, courtesy of Norsk Telegrambyrå AS, © 2022

5.1 Setup

The study had four phases: Introduction, non-interactive session, interactive session, and post-study interview. The order for the introduction (first) and post-study interview (last) were set, while the order for when participants engaged in the non-interactive and interactive sessions depended on their assigned groups.

Introduction: The participants were introduced to the project and AiCommentator, followed by information about their rights and the signing of the consent form. All participants filled out a pre-study questionnaire with questions regarding their demographic and self-described football knowledge and background.

Non-interactive session: A researcher provided a guided walk-through of AiCommentator's non-interactive mode. Following, the participants got to try out the non-interactive mode by watching a one-minute video clip that could be replayed. Afterwards, participants proceeded with the full testing session. Here, 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 think-aloud method. We employed the concurrent think-aloud (CTA) method, supplemented by occasional prompts from the moderator. This approach was chosen to ensure that information from the participants' short-term memory is articulated effectively and accurately while interacting with each test condition [11]. Upon completion, 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".

Interactive session: The non-interactive and interactive sessions followed an almost identical structure. However, due to the interactive mode offering six functions, the interactive session was more comprehensive and included an additional questionnaire related to these functions. These six functions extended the guided walkthrough as participants had to interact separately with each of the functions, followed by the associated post-function questionnaire. The questionnaire included themes such as engagement, satisfaction, trust, consistency and overall preferences. As well as filling in the post-system questionnaire, the participants also complete the SUS [4] questionnaire.

Post-study interview: Participants attended a post-study interview aimed at getting more in-depth accounts of their experiences. The interviewer followed a semi-structured interview guide with topics concerning participants' experiences with each of the prototype's modes, their likes and dislikes, preferences, and such. The interviews lasted between 6.43 and 12.28 minutes and were audio-recorded.

5.2 Questionnaires

As previously mentioned, the participants completed two post-system questionnaires and six post-function questionnaires during the guided walkthrough. The initial design of our questionnaire categories drew inspiration from Chen et al. [5] and Lin et al. [24]. However, we tailored these categories to better evaluate usability aspects relevant to our system's multimodal design. During the pilot studies, we meticulously refined and reworded the questions to ensure that the intent of each statement was communicated clearly. Throughout the user study, a moderator was consistently available to clarify questions from participants, ensuring their complete understanding of the statements. Both questionnaires implemented a 7-point Likert scale to capture the nuances of user feedback more effectively. The broader range of options a 7-point scale gives enhances the sensitivity, allowing the user to express their opinions with greater precision [17]. Such a scale is beneficial in detecting subtle variations in user perceptions, providing more reliable feedback [31], essential for our study's objective of understanding user interactions with the system in depth. We now provide detailed descriptions of each questionnaire in Appendix A and Appendix B, along with justifications for their design.

5.2.1 Post-System Questionnaire. Questions were grouped and analysed under the categories of "knowledge and understanding", "engagement and immersion", "satisfaction and future use", "trust and reliability", "consistency", and "overall preference". The questionnaire (Appendix A) was designed to scale to both the interactive and non-interactive sessions, allowing for a direct comparison between the two systems. In this manner, questions were designed to reflect the usability of the system while not directly referencing interactivity, as only one system was interactive. Questions under the category of "knowledge and understanding" were designed to gain insights in RQ2, while RQ1 was covered by statements under the category "engagement and satisfaction". Statements relating to "Trust and Reliability" were designed to provide useful insights regarding how users perceived information communicated by the MCA. Finally, the remaining categories contribute to the generic usability of the systems.

5.2.2 Post-Function Questionnaire. This questionnaire was designed to retrieve a more nuanced understanding of the usability of each function of the interactive system (RQ3). The questionnaire (Appendix B) aimed to validate the usability and relevance of each function in order to derive design choices for future iterations of an MCA. The initial question sought to determine the relevance of the function; a lack of relevance could account for subsequent negative ratings. We followed by determining whether the function was engaging, lending further nuanced information for RQ1 to determine which function contributed to this factor. Further questions retrieve information related to feedback from the multimodal system before determining satisfaction and relevance of the function outside of the lab environment. For example, "I would use this function in a real-world setting" reflects the perceived intent of the participant using this function in a traditional football viewing environment outside of the laboratory.

5.3 Participants

Sixteen participants (n=16) took part in the evaluation, nine males and seven females between the ages of 20 and 28. These participants were recruited through various means, including project promotions on multiple university course websites and in-person pitches delivered during university lectures. Participants were offered cinema gift cards of 200 NOK ($19.05) as an incentive. As part of the pre-study questionnaire, participants self-reported their level of football knowledge. Participants who self-defined as knowledgeable about football (N=8, P01-P08), reported watching at least 1-4 football matches in a normal month, with most watching between 5 and 10 and one person watching over 10. These viewings were reported to usually take place at home or sometimes at a bar, either with friends or by themselves. These participants described being interested in diverse types of information during a match, such as the lineup, player statistics, game statistics, and season statistics. All the knowledgeable participants reported using supplementary apps, such as Fotmob, to get additional information during a match. Most of them played or had played football games, such as FIFA or fantasy football. On a 5-point Likert scale, where 5 meant very interested, everyone reported 4 or 5 in response to their interest in football and in learning more about it.

Participants who self-defined as less knowledgeable about football (N=8, P09-P16), reported normally watching football at home or in a bar, often because family or friends wanted to do so. None of them had played any football-related games or used football-related apps. The less knowledgeable participants reported being interested

in general information from the match and the players, and several wanted explanations about the rules and the referee's decisions.

Knowledgeable participants exclusively consisted of males, whereas among the less knowledgeable participants were seven females and one male. Because gender and self-reported knowledge appear to be confounding variables, we chose not to make further use of these variables.

5.4 Analysis

The data material used in our project consists of demographic data from the pre-study questionnaire, post-function questionnaire about the usability of each function, post-system questionnaires about the overall usability of each of the two modes, SUS questionnaire for the interactive mode, video recordings from the full testing sessions, and audio recordings from the post-study interviews.

Thematic analysis [3] was used to uncover recurring themes within the qualitative data material. This material consisted of pre-study questionnaires, video recordings, and interview transcriptions. The qualitative analysis began with one researcher familiarizing themselves with the data material by reading through all interview transcripts. Following, a collaborative effort where two researchers extensively examined the interview transcriptions from two interviews took place. This initial review led to the identification of preliminary themes of interest. Subsequently, one researcher conducted an in-depth analysis of the remaining material, incorporating the annotations from the initial review. Further analysis sessions were conducted, involving both researchers, to fine-tune the annotations and explore additional points of interest.

For the scores obtained from the post-system questionnaire, we used the mean scores of several questions obtained for each category of "Knowledge and understanding", "Engagement and immersion", "Satisfaction and future use", "Trust and reliability", and "Consistency". For overall preference, we used the scores for the statement "Overall, I liked the system". We performed Shapiro-Wilk tests on the data of the obtained scores and all of them showed normality ($p > .05$). We used parametric methods for the statistical tests reported below.

6 RESULTS

This section presents the quantitative and qualitative results of our user study.

6.1 Quantitative Results

Figure 7 shows the results of the post-system questionnaire. We conducted paired T-tests for each category of the questionnaire between the two modes (i.e., interactive mode and non-interactive mode) and found that the scores were significantly higher with the interactive mode than the non-interactive mode in "Engagement and immersion" ($p = .008$, Cohen's $d = 0.760$), "Satisfaction and future use" ($p = .011$, Cohen's $d = 0.726$), and the "Overall preference" ($p = .041$, Cohen's $d = 0.559$). The mean SUS score for the interactive mode was 70.52 ($SD = 8.455$) corresponding to *"GOOD"* in adjective rating.

Figure 8 shows the participants' ratings obtained for each interactive function using a scale that ranges from 1 (Strongly Disagree) to 7 (Strongly Agree), with a rating of 4 indicating a neutral viewpoint.

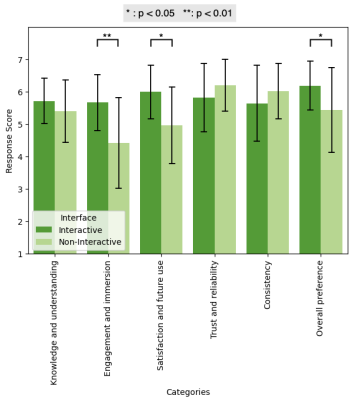


Figure 7: Results of post-system questionnaire

Note that C3 and C5 were removed from the "Query" function because this function did not augment embedded visualizations. Overall, the participants rated each interactive function positively, with a mean score of five or higher for all functions. Among the participants, the interactive function "Highlight Team" received the lowest mean score while "Track Player" emerged as the most favorably rated interactive function.

6.2 Qualitative Results

Both modes of AiCommentator were appreciated by the participants, with the interactive mode described as more engaging. Participants reported feeling well-informed and updated. Some explained that AiCommentator added a new dimension to the football viewing experience. The non-interactive mode was compared to watching a regular football match with additional information and features, while the interactive mode presented an innovative way of receiving relevant information on demand. Nonetheless, there were certain drawbacks associated with each of the modes. The non-interactive mode was reported to leave users feeling overwhelmed due to the sheer volume of information presented by the AI commentators and through the embedded visualizations. In the interactive mode, some participants observed instances of inaccurate information, though they considered these as exceptions and maintained their perception of the overall information as reliable. All participants reported that, in the future, they would like to have one or both modes of AiCommentator; the interactive or a combination of both were the preferred options.

6.2.1 The interactive and non-interactive modes. The non-interactive mode was usually described as "okay" and "fine." Participants highlighted how the viewing experience resembled that of a regular football match with additional features. When asked what they liked, participants answered that they could sit back and relax, not having to use several screens simultaneously or think about questions to ask. Several participants also highlighted the value of the two AI commentators commenting back and forth with each other, making the communication appear more natural. Some mentioned how the embedded visualizations helped them understand parts of the game. Though participants reported being satisfied with the non-interactive mode, they expressed more enjoyment regarding

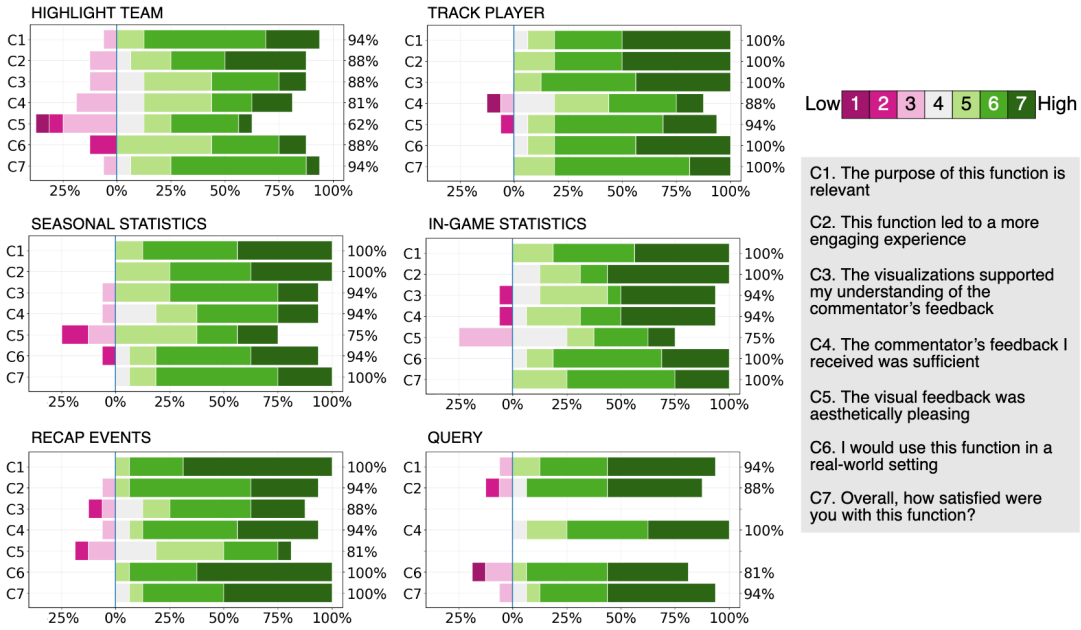


Figure 8: Results of ratings for each interactive function. Please note that the query does not trigger embedded visualizations, so C3 and C5 were omitted from the questionnaire for this function.

the interactive one. When describing the interactive mode, participants used terms such as "cool," "fun," and "engaging," and some even compared it to playing a computer game. When stating what they liked, most participants answered that they were in control of what information they got and when. Some participants highlighted that they perceived the information presented in interactive mode as more relevant compared to the non-interactive. When asked which mode they preferred, 14 of the 16 participants answered the interactive.

6.2.2 Interactions through queries and functions. During the full testing sessions, the knowledgeable participants experimented more with asking direct questions to the system, while the less knowledgeable participants relied more on functions presented in the menu. This correlated with statements from the interviews where several of the less knowledgeable participants pointed out challenges with figuring out questions to ask. However, all participants who asked questions through the chat expressed excitement and were impressed. For example, after receiving the answer to a question about a specific player's past, P11 stated that she liked the information and that it was "cool".

Several of the knowledgeable participants explained how they often had questions while watching a football match. Some even joked about the desire to ask the commentators these questions. They described using a second device, such as a phone, to search for answers on football apps or the internet. As one participant said when describing his enthusiasm for the interactive prototype:

> "It's fun, you know. It's perhaps something one desires when watching football, to be able to ask questions. It's not a real thing, but now it is. So, that was cool." (P02)

It is not only about getting the information but also how one gets the information and the nature of the presentation. There is a difference between reading the answer on a website versus hearing commentators discuss it. The participant further adds to the previous statement by addressing the value of the way one receives this information.

> "[...] you get the answer in commentator style." (P02)

Additionally, the less knowledgeable participants explained that due to limited prior knowledge about the sport, they would often ask friends or family questions during a match, which could be described as bothersome or embarrassing. For them, the interactive mode enhanced the viewing experience and their general football knowledge:

> "I find it enjoyable when I watch with someone who has knowledge, but then I keep asking that person questions all the time. So now there's a tool that allows you to do the same. It becomes a bit more entertaining right away." (P14)

During the interactive full testing, numerous users posed a variety of questions, several of which were not answerable given the prototype's current data. For example, P08 wanted information about a midfield player's defensive statistics, which AiCommentator could not provide. It could be challenging for users to understand what the system is able to provide answers to and what not. While participants had the option to interact by using functions presented in the menu, some struggled to differentiate between these functions and their actions. Additionally, many participants found it challenging to recall the function names and their respective purposes. This was apparent when they interacted with the

prototype and in the interviews when they were trying to recall the interaction.

6.2.3 Knowledge, trust, and error tolerance. Participants reported, both during the full testing sessions and the interview, that the combination of commentary and visual elements was useful in both modes and could potentially enhance their understanding of the game and help them gain more knowledge about different players. Participant P12 explained how the combination of commentary and visual elements helped her identify and learn more about specific players and their teams, which was supported by statements from P14 and P15 about the harmony between the commentary and the visual information. Additionally, the interactive mode was considered especially useful for understanding in-game events.

In terms of the prototype's functions, "Track Player" and the player card visualization were some of the most mentioned and most appreciated. When asked, P1 and P4 explained how player cards helped them get to know different players and learn more about them. Several participants mentioned the usefulness of the recap function, and P03 explained how the function was perfect for getting up to date on the match. In addition, the different statistics functions were mentioned as valuable by several participants. Participants P10 and P11 stated that the results from the "Seasonal Statistics" function met their expectations, while P14 and P15 explained that the visual elements matched the commentary. However, a key challenge with the visual elements was the duration for which they appeared, with participants mentioning a lack of time to process the information.

Participants reported finding the information shared by the AI commentators and in the visualizations to be predominately credible and explained their different manners of assessing their credibility. Some believed the information due to the presentation of data, as seen in this quote:

> "I would say that this is very fact-based. You blindly trust what they say." (P13)

The knowledgeable participants often used available data, e.g. the commentators' statements, together with what they saw taking place on the field and their own prior knowledge, to evaluate how the types of information corresponded with each other. The AI commentators sometimes gave conflicting descriptions. P07 explained how the AI commentators always started by mentioning how great a player had performed while stating the current stats, even though the player had not played a great game at all. There were several instances within the interactive mode when the AI commentators provided incorrect information to the users' requests. This was, for example, when a player was affiliated with the opposite team or stated a score different from what was presented in the visualizations. Yet, the participants who experienced such mistakes still described the information they received as credible and trustworthy and regarded these errors as exceptions, indicating an error tolerance.

6.2.4 Potential for improvement. There was some critique regarding the amount and relevance of information being presented, especially in the non-interactive mode. Many expressed that the non-interactive mode provided too much information, which the participants experienced as overwhelming. In addition, although several participants liked the embedded visualization, it was often reported to be distracting, especially when several visual elements appeared simultaneously or when both auditory and visual information inundated them while they attempted to follow on-field action.

Both during the full testing session and in the post-study interviews, participants reported that the robotic features of the AI commentators' voices lacked passion and emotion, qualities highly associated with football viewing. For some, this lack of personality together with an unbiased attitude could potentially undermine trust. This seems to have mainly concerned the non-interactive mode. Several participants reported preferring human commentators to the commentators in the non-interactive mode for this reason. P07 typified this response:

> "However, in a real football match, I sense a better dynamic in the commentary language. They manage to filter out the important information and exclude the unimportant. Also, I miss the personality compared to what a real football commentator brings." (P07)

P07 and others suggested that it would improve the system if they could specify what type of information they were interested in and their knowledge level, and the AI commentators would base their discussion and commentary accordingly. Some suggested combining modified versions of the two modes, and others suggested having the option to turn both the commentary and the embedded visualizations on and off so they could use them on-demand.

7 DISCUSSION

Structured by our original research questions (RQ1-RQ3), this section discusses the quantitative and qualitative findings while focusing on the broader implications for research and design practice.

7.1 Engagement and Satisfaction

Sports commentators seek to improve viewers' satisfaction by creating engaging and immersive experiences while communicating play-by-play events and information surrounding team and player performance. We designed two modes of the AiCommentator: one to reflect the passive, traditional commentary with the addition of embedded visualization and the other to promote active interaction from the viewer in order to direct commentary and embedded visualizations. RQ1 was intended to determine whether the interactive mode of AiCommentator would enhance engagement and satisfaction compared to the non-interactive mode.

Our T-test shows that the measures "engagement and immersion", "satisfaction and future use", and "overall preference" indicate a significant difference between the two system modes, with the interactive scoring higher in all instances (Figure 7). Participants reported that the non-interactive mode delivered an experience akin to traditional football viewing with the addition of embedded visualizations. In contrast, the interactive mode drew users into an active viewing state with on-demand information, giving more control to the users' viewing experience.

Participants reported that AiCommentator's non-interactive mode could be overwhelming due to the breadth of information delivered in quick succession by the play-by-play and color commentary. Combined with computer-generated voices, which lacked the dynamic vocal range of real commentators, the non-interactive

mode detracts from the sense of immersion. In comparison, the interactive mode limited the commentary to each interaction. It provided direct feedback from user queries and function calls with "matter of fact" characteristics more suited to the computer-generated voice. Hence, a lower cognitive load heightened the immersion, and users felt they were communicating with the AI commentators.

Future efforts should carefully consider the dynamic nature of the commentator's delivery concerning play-by-play commentary events. Typically commentators change pace, pitch, and volume to accentuate their own personality and reflect the current events on the pitch. Integrating a more sophisticated delivery method concerning these three aspects may reduce the distracting properties of the computer-generated voice.

To further reduce cognitive load in the non-interactive mode and improve immersion, our findings identified three viable options:

- Reporting only events of interest for play-by-play commentary
- Allowing users to customize the information relevant to their preferences
- Giving users the option to toggle aspects of the commentary

By balancing a combination of these three factors and adding a customizable layer over the current non-interactive mode that prioritizes events of interest, we can tailor the experience more closely to user preferences and knowledge levels.

Another consideration for the interactive mode is the variety of multimodal interactions available. Although both modes provided multimodal feedback, the interactive system naturally provided a richer variation. As reflected in Figure 8, participants reported that visualizations supported their understanding of the commentator's feedback (Q3 $\geq$ 88%). This multimodal information synchronization and versatile presentation seem to enhance immersion and engagement.

Building on this observation of enhanced immersion with multimodal synchronization, we confirm findings similar to those of Chen et al. [5], who acknowledged that participants interacting with embedded visualizations while following the commentary experienced a heightened sense of immersion. Chen et al. referenced the McGurk effect [41], a multisensory phenomenon where conflicting visual and auditory speech cues lead to a perception different from either modality alone. Our system synchronizes the visual and auditory modalities to complement each other, reinforcing the two to avoid such conflicts. Essentially, this synchronization is italicizing, defined in Section 1, commonly used to promote engagement and understanding of sports content. While both modes of AiCommentator support italicizing, the interactive mode offers more variety, which may contribute to a heightened sense of engagement. However, we must refrain from inundating viewers with multimodal content, as it can distract from the game.

7.2 Knowledge and Understanding

Both modes of AiCommentator provided multimodal feedback regarding players on the pitch, their respective performances, and in-game events. To judge the effectiveness of feedback, RQ2 was aimed at determining whether each mode could support the viewer's understanding. The T-test p-value for "Knowledge and understanding" indicated no significant effect for system mode. Figure 8 confirms

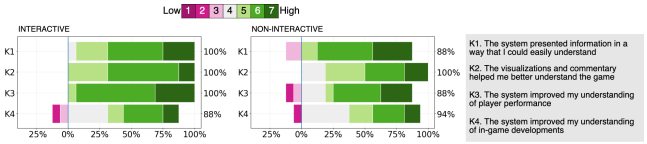


Figure 9: Results of "Knowledge and Understanding" Rating

both systems sufficiently assisted the user's knowledge and understanding of the game.

A closer examination of the independent variables constituting "knowledge and understanding" offers significant insights, as seen in Figure 9. Overall, the interactive system outperformed in presenting information, delivering multimodal content, and enhancing the understanding of player performance, as indicated by (Interactive K1 $\geq$ 100%), (Interactive K2 $\geq$ 100%), and (Interactive K3 $\geq$ 100%), respectively. P11 found the "Query" function effective for understanding a player's past. P12 explained how adding embedded visualizations with commentary helped identify and learn more about players and their respective teams. While other participants reported that the "Seasonal Stats" function improved their knowledge of individual player performances.

However, the play-by-play commentary of the non-interactive system proved more effective in conveying in-game developments, as evidenced by Non-Interactive (K1 $\geq$ 94%) in Figure 9. This disparity arises because events are automatically queued and verbalized in a timely manner with the non-interactive mode. In contrast, in the interactive mode, users must specifically request play-by-play information using the "In-Game Statistics" and "Recap" functions. While functions like "Recap" were reported to give satisfactory results (P03, P12, P15, P12), the play-by-play commentary was sparse in comparison to the non-interactive mode.

7.3 Usability

The results present in Section 6.2 revealed a "Good" SUS score for the interactive system usability. As seen in Figure 7, the response score for the usability measures received satisfactory results, with mean values above five on the Likert scale for all categories. The interactive mode scores lower than the non-interactive mode for "consistency" and "trust and reliability", likely due to reported hallucinations from the GPT model [27]. Further quantitative data displayed in Figure 8 shows system functions were perceived as relevant and received a high score in satisfaction (C7 $\geq$ 94%) and typically led to a more engaging experience (C2 $\geq$ 88%).

We now examine our findings to consider the usability of the MCA and provide design recommendations for future work. These findings are based on our post-function questionnaire visualized in Figure 8 and SUS.

7.3.1 Synchronizing Multimodal Feedback. Participants overwhelmingly felt that the visualizations enhanced their comprehension of the commentator's feedback (C3 $\geq$ 88%). In particular, the "Seasonal Statistics" function stood out with a high affirmation rate (C3 $\geq$ 94%). P10 and P11 reported that "Seasonal Statistics" met their expectations, while P14 and P15 found the visual information harmonious with the commentary. This shows the importance of combining audio with visual content, as the combination can foster

and enrich a more holistic understanding, enhancing the viewer's experience

7.3.2 Designing for Non-Intrusive Embedded Visualizations. Subtlety is crucial when crafting embedded visualizations for MCA. As discussed in section 7.1, users consistently found commentary in the non-interactive mode overwhelming. Similarly, participants felt specific visualizations were distracting and obstructed the ongoing gameplay. Overall, visual aesthetics received comparatively low ratings, with "Highlight Team" (C5 $\geq$ 62%), "Season Statistics" (C5 $\geq$ 75%), "In-Game Statistics" (C5 $\geq$ 75%), and "Recap Event" (C5 $\geq$ 81%). The "Highlight Team" function likely scored lower because tracking multiple players simultaneously can obscure the gameplay. In contrast, the "Track Player" feature (C5 $\geq$ 94%) received a higher rating, likely because it was less visually intrusive, and was considered more aesthetically pleasing. For future endeavors, we advise minimizing the complexity of embedded visualizations to maintain clarity and viewer focus.

7.3.3 Bridging Traditional Content Features with Multimodal Conversational Agents. Part of the appealing nature of AiCommentator's interactive mode was the feedback of on-demand information in the style of commentators. Participant P02 echoed this sentiment, noting the heightened experience derived from posing questions during a football match and receiving answers imbued with the distinct flair of a commentator. To seamlessly integrate an MCA with video content, researchers should draw inspiration from the content to repurpose properties that resonate with the original experience. In doing so, the content's authenticity improves, ensuring the user's engagement with the content. We achieved this by modernising commentary while borrowing from the classical elements of traditional commentary.

7.3.4 Personalization. We observed participants desired different information from the system. For example, P08 missed more complex defensive stats for a midfield player, while P11 was keen on player background. Others felt the commentary was overly optimistic, which risks undermining system trust, as noted by P07, and diminishes the authenticity. Genuine commentators often have biases that add to the entertainment. Personalization would help tailor feedback better to individual preferences. However, the depth of the MCA's insights reflects the database's breadth. Enriching this database with external match-related stories and using methods like those by Chitrakala et al. [7] and Lee et al. [22] can offer a broader range of multimedia content.

7.3.5 Interface Design. The realm of communicating with linear media is still emerging. While Chen et al. [5] and Lin et al. [24] have designed systems for interacting with basketball visualizations via gaze and voice, our approach prioritizes interactions using natural language textual input and a menu-based system. Our results show MCA was a novel method of interacting with embedded visualizations. However, some participants, especially those less acquainted with football, faced difficulties framing questions to AI commentators. Interacting with MCA is an unfamiliar experience for many people, which can cause uncertainty in interactions. As users become more acquainted with MCA this hesitation will likely diminish. In our work to address this current unfamiliarity, we integrated a menu system. However, this resulted in less dynamic feedback since it relied on repetitive templated prompts. Future research should consider striking a balance between the two approaches we have developed.

8 LIMITATIONS

In Section 5.3, we note that gender is a confounding variable within our dataset. Therefore, we couldn't directly study its impact on user knowledge concerning our system. However, the focus of our work was not on this aspect. Instead, we derived insights from qualitative data based on individual users' self-described knowledge levels. Future research could investigate how a user's knowledge affects such MCA, providing richer design insights for personalization. Another limitation lies in the duration of the video content and the controlled setting of the study. We had only five minutes of footage available, which offers limited time for users to explore a multifaceted system. Yet, during the guided walkthrough sessions, facilitators ensured participants felt comfortable with the system before transitioning to the full testing sessions. It's important to note that sports viewing is often a social experience, and conducting the study in a controlled environment might not fully capture viewers' natural dynamics and reactions compared to a relaxed group setting. While AiCommentator modules perform in real-time, FootyVision cannot achieve real-time inference speeds. Therefore, it is first necessary to preprocess and clean the data prior to inference with AiCommentator. For this reason, currently AiCommentator cannot be integrated into broadcasting workflows. Future works should consider improving inference speeds of MOT and automated identity recognition of players on the pitch. In doing so, systems like AiCommentator could revolutionize sports content interaction and accessibility.

9 CONCLUSION AND FUTURE WORK

Our work serves as a foundational study in non-interactive and interactive MCA for sports commentary. It establishes a benchmark in the field and reveals areas for improvement, thereby guiding future research in embedded visualizations and MCA for sports content. Our prototype, AiCommentator, employs state-of-the-art CV, DL and NLP to introduce two novel commentary modes: interactive and non-interactive. This advancement redefines the traditional commentary landscape, enabling users to interact directly via natural language or menu-driven interfaces facilitated by a Discord bot. By merging AI commentators with embedded visualizations, we present a contemporary take on commentary, integrating synchronized multimodal feedback—a method known as "italicizing". In our evaluation, sixteen participants underwent a thorough mixed-methods examination of AiCommentator. While both modes rated high on satisfaction and engagement, the within-group (AB) design revealed a distinct preference for AiCommentator's interactive mode. The findings indicated that the interactive mode gave participants a more captivating and immersive experience, leading to high satisfaction levels. The usability of the interactive mode received a "Good" rating on the SUS scale, and data from post-function questionnaires further endorsed our function design framework. Qualitative insights echoed the quantitative data, with some users vocalizing their aspiration for such a system. Meanwhile, other participants highlighted elements they appreciated from both modes, suggesting

the potential for a hybrid solution. Future research should optimize the balance between interactive and non-interactive embedded visualizations. The expressed interest in diverse features and content underscores the potential for tailoring the viewing experience to individual preferences. However, achieving this personalized experience will necessitate further exploration into making embedded visualizations more adaptable to a broad spectrum of user requests. Developers and designers can directly utilize our findings in sports broadcasts, sports analysis, and academic projects.

ACKNOWLEDGMENTS

This work was supported in part by the Norwegian Research Council (MediaFutures, 309339), JSPS Grants KAKENHI (19KK0258, 21H03473), and the Wallenberg Foundations (MMW/MAW). We thank Ayça Ünlüer for artwork and illustrations, Philippa Beckman for proofreading, Barbara Stuckey, for proofreading and editing, Mariana Heggholmen at Sindella Productions for video production. Izabela Krejtz for advice on data analysis. We further acknowledge NRK for the women's football dataset and TV2 for the inspiring conversations.

REFERENCES

[1] Lee Anderson. 2012. Main findings: Teens, technology, and human potential in 2020. https://www.pewresearch.org/internet/2012/02/29/main-findings-teens-technology-and-human-potential-in-2020/. Accessed: 2024-01-03.

[2] Kim Binsted and Sean Luke. 1999. Character design for soccer commentary. In *RoboCup-98: Robot Soccer World Cup II 2*. Springer, 22–33.

[3] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[4] John Brooke. 2020. SUS: A 'Quick and Dirty' Usability Scale. *Usability Evaluation In Industry* July (2020), 207–212. https://doi.org/10.1201/9781498710411-35

[5] 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.

[6] 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.

[7] S Chitrakala, K Akshaya, and S Nisha. [n. d.]. STORY SELECTION AND RECOMMENDATION SYSTEM FOR COLOUR COMMENTARY IN CRICKET. ([n. d.]).

[8] Pietro Crovari, Sara Pidó, Franca Garzotto, and Stefano Ceri. 2021. Show, Don't Tell. Reflections on the Design of Multi-modal Conversational Interfaces. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 12604 LNCS (2021), 64–77. https://doi.org/10.1007/978-3-030-68288-0_5

[9] 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

[10] 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.

[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] FIFA. 2021. The football landscape – the vision 2020-2023. https://publications.fifa.com/en/vision-report-2021/the-football-landscape/

[13] K Goldsberry. 2012. CourtVision: New Visual and Spatial Analytics for the NBA. Conference Paper. MIT Sloan Analytics Conference.

[14] Cagatay Goncu and Daniel J Finnegan. 2021. 'Did You See That!?'Enhancing the Experience of Sports Media Broadcast for Blind People. In *Human-Computer Interaction–INTERACT 2021: 18th IFIP TC 13 International Conference, Bari, Italy, August 30–September 3, 2021, Proceedings, Part I 18*. Springer, 396–417.

[15] Manon Ichiki, Toshihiro Shimizu, Atsushi Imai, Tohru Takagi, Mamoru Iwabuchi, Kiyoshi Kurihara, Taro Miyazaki, Tadashi Kumano, Hiroyuki Kaneko, Shoei Sato, et al. 2018. Study on automated audio descriptions overlapping live television commentary. In *Computers Helping People with Special Needs: 16th International Conference, ICCHP 2018, Linz, Austria, July 11-13, 2018, Proceedings, Part I 16*. Springer, 220–224.

[16] Euromonitor International. 2023. Women's World Cup 2023 viewership to cross 2 billion, double from 2019: Euromonitor International. https://www.euromonitor.com/press/press-releases/july-20232/

[17] Jon A Krosnick and Stanley Presser. 2010. Question and Questionnaire Design. In *Handbook of Survey Research* (2 ed.), Peter V. Marsden and James D. Wright (Eds.). Emerald Group Publishing Limited, [268–274].

[18] Tadashi Kumano, Manon Ichiki, Kiyoshi Kurihara, Hiroyuki Kaneko, Tomoyasu Komori, Toshihiro Shimizu, Nobumasa Seiyama, Atsushi Imai, Hideki Sumiyoshi, and Tohru Takagi. 2019. Generation of automated sports commentary from live sports data. In *2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. IEEE, 1–4.

[19] Kiyoshi Kurihara, Atsushi Imai, Nobumasa Seiyama, Toshihiro Shimizu, Shoei Sato, Ichiro Yamada, Tadashi Kumano, Reiko Tako, Taro Miyazaki, Manon Ichiki, et al. 2019. Automatic generation of audio descriptions for sports programs. *SMPTE Motion Imaging Journal* 128, 1 (2019), 41–47.

[20] Knut Kvale, Olav Alexander Sell, Stig Hodnebrog, and Asbjørn Følstad. 2020. Improving Conversations: Lessons Learnt from Manual Analysis of Chatbot Dialogues. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11970 LNCS (2020), 187–200. https://doi.org/10.1007/978-3-030-39540-7_13

[21] Philippe Laban, Elicia Ye, Srujay Korlakunta, John Canny, and Marti Hearst. 2022. NewsPod: Automatic and Interactive News Podcasts. *International Conference on Intelligent User Interfaces, Proceedings IUI* (2022), 691–706. https://doi.org/10.1145/3490099.3511147

[22] Greg Lee, Vadim Bulitko, and Elliot A Ludvig. 2013. Automated story selection for color commentary in sports. *IEEE transactions on computational intelligence and ai in games* 6, 2 (2013), 144–155.

[23] Minkyo Lee, Daeyeon Kim, Antonio S Williams, and Paul M Pedersen. 2016. Investigating the role of sports commentary: an analysis of media-consumption behavior and programmatic quality and satisfaction. *Journal of Sports Media* 11, 1 (2016), 145–167.

[24] 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.

[25] Robert Madrigal. 1995. Cognitive and affective determinants of fan satisfaction with sporting event attendance. *Journal of leisure research* 27, 3 (1995), 205–227.

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

[27] Baolin Peng, Michel Galley, Pengcheng He, Hao Cheng, Yujia Xie, Yu Hu, Qiuyuan Huang, Lars Liden, Zhou Yu, Weizhu Chen, et al. 2023. Check your facts and try again: Improving large language models with external knowledge and automated feedback. *arXiv preprint arXiv:2302.12813* (2023).

[28] Florian Pfeffel, Peter Kexel, Christoph A Kexel, and Maria Ratz. 2016. Second screen: User behaviour of spectators while watching football. *Athens Journal of Sports* 3, 2 (2016), 119–128.

[29] 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.

[30] Sheena Anees Prathibha.M. 2015. Computational Intelligence Based Color Commentary System in Sports. *International Journal of Scientific Engineering and Applied Science* 1, 7 (2015), 472–475.

[31] Carolyn C Preston and Andrew M Colman. 2000. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta psychologica* 104, 1 (2000), 1–15.

[32] J Ross Quinlan. 2014. *C4. 5: programs for machine learning*. Elsevier.

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

[34] Carlos Segura, Àlex Palau, Jordi Luque, Marta R. Costa-Jussà, and Rafael E. Banchs. 2019. Chatbol, a chatbot for the Spanish "La Liga". *Lecture Notes in Electrical Engineering* 579, May (2019), 319–330. https://doi.org/10.1007/978-981-13-9443-0_28

[35] 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.

[36] Manuel Stein, Johannes Häußler, Dominik Jäckle, Halldór Janetzko, Tobias Schreck, and Daniel A Keim. 2015. Visual soccer analytics: Understanding the characteristics of collective team movement based on feature-driven analysis and abstraction. *ISPRS International Journal of Geo-Information* 4, 4 (2015), 2159–2184.

[37] 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.

[38] Manuel Stein, Halldor Janetzko, Andreas Lamprecht, Thorsten Breitkreutz, Philipp Zimmermann, Bastian Goldlücke, Tobias Schreck, Gennady Andrienko, Michael Grossniklaus, and Daniel A Keim. 2017. Bring it to the pitch: Combining video and movement data to enhance team sport analysis. *IEEE transactions on visualization and computer graphics* 24, 1 (2017), 13–22.

[39] Kumiko Tanaka, Hideyuki Nakashima, Itsuki Noda, Kôiti Hasida, Ian Frank, and Hitoshi Matsubara. 1998. MIKE: An automatic commentary system for soccer. In *Proceedings International Conference on Multi Agent Systems (Cat. No. 98EX160)*. IEEE, 285–292.

[40] Kumiko Tanaka-Ishii, Kôiti Hasida, and Itsuki Noda. 1998. Reactive content selection in the generation of real-time soccer commentary. In *COLING 1998 Volume 2: The 17th International Conference on Computational Linguistics*.

[41] Kaisa Tiippana. 2014. What is the McGurk effect? *Frontiers in psychology* 5 (2014), 725.

[42] Matthew Turk. 2014. Multimodal interaction: A review. *Pattern Recognition Letters* 36, 1 (2014), 189–195. https://doi.org/10.1016/j.patrec.2013.07.003

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

[44] Dirk Voelz, Elisabeth André, Gerd Herzog, and Thomas Rist. 1999. Rocco: A RoboCup soccer commentator system. In *RoboCup-98: Robot Soccer World Cup II 2*. Springer, 50–60.

[45] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. 2023. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7464–7475.

[46] Lawrence A Wenner. 1989. *Media, sports, and society*. sage.

[47] Mikołaj Wieczorek, Barbara Rychalska, and Jacek Dąbrowski. 2021. On the unreasonable effectiveness of centroids in image retrieval. In *Neural Information Processing: 28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, December 8–12, 2021, Proceedings, Part IV 28*. Springer, 212–223. https://doi.org/10.48550/arXiv.2104.13643

[48] 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.

[49] 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.

[50] 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.

[51] 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.

[52] Qiyu Zhi and Ronald Metoyer. 2020. GameBot: A visualization-augmented chatbot for sports game. *Conference on Human Factors in Computing Systems - Proceedings* (2020), 1–7. https://doi.org/10.1145/3334480.3382794

A 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	○	○	○	○	○	○	○
Overall, I liked the system	○	○	○	○	○	○	○

B POST-FUNCTION QUESTIONNAIRE

Date:

Participant:

Function:

	Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree
The purpose of this function is relevant	○	○	○	○	○	○	○
This function led to a more engaging experience	○	○	○	○	○	○	○
The visualisations supported my understanding of the commentator's feedback	○	○	○	○	○	○	○
The commentator's feedback I received was sufficient	○	○	○	○	○	○	○
The visual feedback was aesthetically pleasing	○	○	○	○	○	○	○
I would use this function in a real-world setting	○	○	○	○	○	○	○
Overall, how satisfied were you with this function?	○	○	○	○	○	○	○

Explain with a couple of words:

Function:

	Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree
The purpose of this function is relevant	○	○	○	○	○	○	○
This function led to a more engaging experience	○	○	○	○	○	○	○
The visualisations supported my understanding of the commentator's feedback	○	○	○	○	○	○	○
The commentator's feedback I received was sufficient	○	○	○	○	○	○	○
The visual feedback was aesthetically pleasing	○	○	○	○	○	○	○
I would use this function in a real-world setting	○	○	○	○	○	○	○
Overall, how satisfied were you with this function?	○	○	○	○	○	○	○

Explain with a couple of words:

Function:

	Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree
The purpose of this function is relevant	○	○	○	○	○	○	○
This function led to a more engaging experience	○	○	○	○	○	○	○
The visualisations supported my understanding of the commentator's feedback	○	○	○	○	○	○	○
The commentator's feedback I received was sufficient	○	○	○	○	○	○	○
The visual feedback was aesthetically pleasing	○	○	○	○	○	○	○
I would use this function in a real-world setting	○	○	○	○	○	○	○
Overall, how satisfied were you with this function?	○	○	○	○	○	○	○

Explain with a couple of words: