A Context-aware Interface for Immersive Sports Spectating

Wei Hong Lo* University of Otago
Holger Regenbrecht† University of Otago
Barrett Ens‡ Monash University
Stefanie Zollmann§ University of Otago

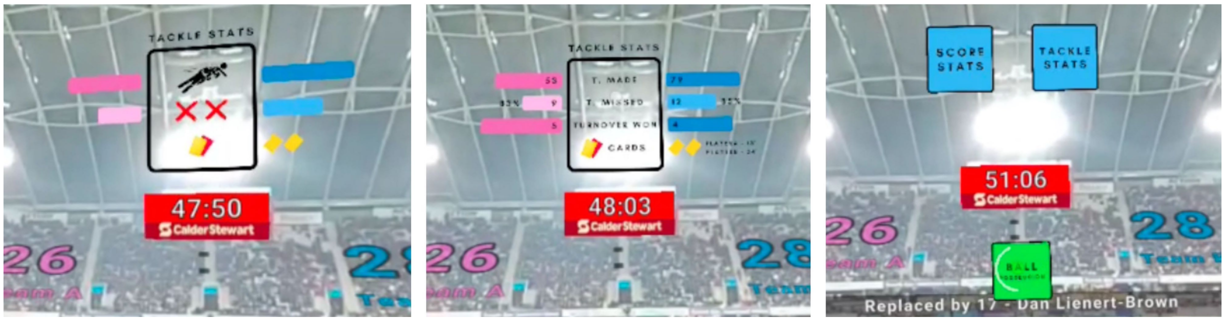

Figure 1: Our level of detail (LOD) implementation and context-aware adaptive AR interface for sports spectating: Left) A LOD-1 rugby tackle stats visualization which displays minimal information. Middle) A LOD-3 rugby tackle stats visualization with the maximum information shown if the user wants more details. Right) Part of the adaptive interface where users rely fully on context and head movement to interact. Shown is the LOD-0 icons for user to select.

Abstract

Novel Augmented Reality sports spectating interfaces allow on-site sports spectators to access game-related information by overlaying relevant digital data into their field of view. However, displaying all game-related information at once would overload the user. Therefore it is important to develop a suitable interface that is aware of the game context, the user's context, and is able to display relevant information at the right time. We developed a state inference model based on spectators' behavior and game states to provide a context-aware sports spectating interface. The interface gradually reveals information using different levels of detail that is based on the context of the game. As an implementation of our model, we created a prototype featuring a context-aware adaptive interface for a sports spectating scenario. Although our implementation is just a preliminary prototype, the goal of this research is to begin the exploration of intelligent context-aware interfaces to be used in on-site sports spectating.

Index Terms: Human-centered computing—Interaction design—Interaction design process and methods—Interface design prototyping; Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality

1 Introduction

Sports spectators often visit a stadium as this provides atmosphere and team identification [16]. However, on-site spectators are often left out of game understanding due to the lack of commentary and visualizations delivered only by advanced sports broadcast [7, 31]. Augmented Reality (AR) can address this challenge and provide a novel interface for sports spectating that allows the display of situated visualizations on-site [44]. However, a suitable, we need a context-aware interface to display all the information relevant to spectators. Ideally, this information should involve interactions between the multiple components in the 3'C's sports spectating visualization framework (canvas, content, context) [22], such as game context, user personalization, and sports content.

Therefore, our motivation for this research is to explore interface options for on-site sports spectating. We want to create a context-aware, adaptive interface to facilitate on-site sports spectating, a novel use case currently under-explored. There are rule-based context-aware mixed reality interfaces developed for day-to-day tasks [20], meanwhile in sports, most AR work revolves around training and performance [10]. As far as we know, no context-aware interface is made for on-site sports spectating. To achieve this aim, we created a state inference model, which combines information about the users' states and profiles with the context of the scene and its surrounding objects. It then provides visualizations in suitable locations, at appropriate times, and with a suitable level of detail (LOD) to support easy comprehension of relevant information. In this paper, we discuss the requirements for a context-aware sports spectating scenario and a context-aware adaptive interface (Fig. 1, right) —used alongside our LOD-based visualization method (Fig. 1, left and middle).

2 Design Space for a Context Aware Interface

Just as we need an interface to interact with computers, such as via physical buttons on a keyboard or touch through a touchscreen, we need a suitable mode of interaction with XR applications. The Window, Icon, Menu, Pointer (WIMP) interface paradigm is now part of daily life. However, it is unsuitable for XR applications due to its design limitations for two-dimensional (2D) spaces [41]. Hence, we would need to look into tangible user interfaces [15] or Spatial Analytic Interfaces [13], where the use of three-dimensional (3D) spaces introduces a closer relationship between virtual and physical objects.

2.1 XR Interaction Methods

Due to the casual nature of sports spectating, we focus primarily on interaction methods that would require the least amount of workload. Prior research has shown that interactions in XR can affect mental workload, especially for AR [17, 40], which is what we want to avoid when a spectator goes to a game. We also would like to support interaction in large-environment AR scenarios, where the data referents (object semantically associated with the data) may

*e-mail: wei.lo@postgrad.otago.ac.nz
†e-mail: holger.regenbrecht@otago.ac.nz
‡e-mail: barrett.ens@monash.edu
§e-mail: stefanie.zollmann@otago.ac.nz

be out of reach. This large environment use case is contrary to the majority of AR applications where it involves a small AR workspace [18] where the referents are close to the users.

Most XR applications use **input devices** such as controllers in combination with tailored interaction methods. The tracked controllers of a VR headset provide not only button and trigger input but also a mode of pointing. In a sports spectating scenario, it is not realistic to have spectators carry controllers such as the ones shipped with VR headsets, nor is it feasible to have spectators making large gestures with a controller during a game [2]. A potentially feasible solution is using mini Bluetooth controllers that generally contain just one or two buttons for essential interactions.

The use of **voice commands** in an interface is becoming quite common [36] but comes with its limitations, such as the comprehensibility of the language itself [11]. Although there are advancements in noise cancellation [23], voice control is not feasible in loud environments such as the stadium in our case, as the noise from other spectators is of similar frequency to the users' voice. Last but not least, voice input in any public setting could lead to awkward and embarrassing situations [6].

There are two common forms of **gesture control**, the first involving gestures on the surface of a touch screen or track-pad, and the other using motion of hands in mid-air. We focus here on mid-air gestures since introducing touch-pad components may distract users' attention from the event they are spectating [24, 42]. Some mobile phones and smart home devices use infrared sensors to recognize the waving of hands. However, a higher level of accuracy is needed in XR applications. In XR applications, the whole palm, and not only fingertips, would need to be tracked, usually by the HMD's external-facing cameras. There are also input devices such as data gloves which users could wear to track hands [34], allowing for specific hand gestures to point at and interact with virtual elements. Using such gestures in a stadium environment would be inappropriate as it quickly leads to arm fatigue for long-term use and might distract surrounding viewers, let alone potential social acceptance issues.

Head and eye tracking consider where the user is looking. Head tracking [19] is essential in every HMD, regardless if it is AR or VR. Both have been used for object selection; however, head tracking approximates the user's field of view based on their head movement, while eye tracking [9] tracks the users' eyes and tries to pinpoint where the user is looking accurately. Eye tracking is more easily implemented in VR systems as the cameras are closer to the display lens, which makes it more difficult for an AR HMD. While eye tracking might be more useful in our AR sports spectating scenario, head tracking is still the more common form of tracking in many pre-existing devices. It would be the closest alternative to ascertaining a user's viewing direction.

Physiological measurements such as electroencephalogram (EEG) [25], heart rate [37], and galvanic skin response [38] could be used as feedback mechanisms toward what the user is feeling or wanting to do. An example of this would be the use of EEG to trigger input to the computer, known as a Brain-Computer Interface (BCI). BCI is meant to be a supplementary tool with other forms of interaction. With that said, there are still many limitations to such technology where users need to be still, and it would not make sense to rig spectators in the stadium with expensive and sensitive equipment.

2.2 XR Interfaces

After discussing the different input methods for XR, the next step is to determine how these input methods could be used or combined with other XR-related technologies into one interface. Certain general design principles are implementable in AR, such as Shneiderman's design principles for direct manipulation [30]. However, with the added complexity of a 3D interface, we would need a better interaction interface. Here, we discuss some of the more common XR interfaces in other work.

2.2.1 Personal Interaction Panel

The personal interaction panel (PIP) is a private panel used for interacting with the contents of the user's surrounding environment. The original concept of PIP was by Szalavari et al. [33], where a panel held on the non-dominant hand is interacted with using a stylus on the dominant hand. This interaction method allows users to manipulate virtual objects in the environment by performing various gestures. Using an AR or VR HMD, users could have visualizations augmented onto the panel itself to provide context while providing privacy [28].

One could implement the PIP in two different approaches. As mentioned, the PIP could be a physical panel the user holds, with the user interface overlaid virtually on the panel. The physical panel would give the user something to hold on to and is suitable for conditions where the user needs to perform specific movements such as minor adjustments with the stylus. The other approach would be to have a virtual panel. Simulating a physical PIP, a virtual panel could float in mid-air, freeing one of the users' hands while still allowing users to interact with the virtual objects in the environment through the virtual PIP.

2.2.2 Image Targets

Image targets are commonly used in tangible AR applications, where a tangible user interface is bundled with an AR display to create a tangible physical interface [4]. As the name implies, it is an image that the camera of an AR system detects and identifies. The AR headset then tracks the movement of the image target and uses this as an anchor to place visualizations. Developers typically use image targets as an interaction method where images/codes printed on cards serve as user interface elements by physically moving the cards in a certain way or bringing image targets together. This method is useful for applications that involve moving visualizations around, such as city planning, where the image targets could e.g. represent buildings or chemical elements (e.g. Arloon [1]).

2.3 Context Awareness

Context is the implicit situational information that takes place when humans communicate with each other. However, this information does not transfer well to the interaction between computers and humans in traditional computing [1]. Therefore, to reduce the complexity of user input towards a system, the system needs to know about existing information regarding the users' environment, leading to the rising research of context-aware computing [21]. The increasing sophistication of hardware developed in recent times allow for better context-aware computing in sports [3]. These include the advancement of sensors, cameras, computer vision, and deep learning technologies [5, 43], which could contribute to detecting the context of the environment.

In a sports spectating context, both interests in sports and the progress of game events are the top contextual factors that influence the user experience [32], illustrating the importance of the user context and the game context. In addition, a proposed model of enjoyment for sports spectatorship that emphasized the importance of context-awareness was introduced more recently [27]. This model builds on self-determination theory [12], where competence, autonomy, and relatedness are considered intrinsic needs of people. All these show that providing relevant content at suitable timing to the right audience will significantly benefit the viewing experience.

Apart from users and external events, there is another type of context awareness relating to the connection of virtual visualizations to the real world. Our 3D stadium model made it easier to identify canvases for visual placement. However, several researchers look

[1] http://www.arloon.com/

into the placement of virtual content in the real world semantically [8, 26]. Grasset et al. [14] uses computer vision methods to identify saliency and edge map, allowing for placement of labels without obstructing anything important. Various optimization methods could establish a connection from the virtual world to the real world. Hence, there are many parameters to consider for a context-aware interface, all the way from the environment's situation to the users' preferences.

3 State Inference Model for Context-Aware Sports Spectating

To start developing an interface with context awareness, there are three components that we need to consider —object, user, and scene context [22] in a stadium environment, here for the sport of rugby. Objects refer to elements on the field, such as the players and ball, users are the spectators themselves, and the scene is the collection of events happening in the game. In this section, we will describe our state inference model to enable a context-aware sports spectating interface. A context-aware interface would mean spectators should get the right visualizations at the appropriate time and place without direct (manual) trigger events from users, which would distract from the enjoyment of the game to be followed at the same time. Our state inference model is generic in terms of its interaction methods. Different instances of the state inference model could use different interaction methods, such as eye tracking for gaze detection, head tracking, VR controllers, and EEG as input methods. Hence, in this section, we will not mention specific interaction methods and will use the terms *triggered* and *focused*.

3.1 State Inferences

The state inference model is based on related work by Tsai et al. [35], proposing a user-behaviour-driven augmented content display called iDisplay. The authors created a state inference mechanism that takes in information from the users' movement and predicts user states based on historical user information. Therefore, based on different user actions such as walking, being stationary, or looking around, different styles of visualizations are presented based on what the system *thinks* the users' state is.

The core concept described echos closely what happens in the stadium, albeit our stadium use case is slightly different where we rely more on the scene context to determine what is happening on the field. In the stadium, the sports data provider would provide the object and scene context, where player data (object context) and events happening (scene context) would be streamed to the AR sports spectating system. The system will also take in user context in the form of user localization and personalization for determining the user's position in the stadium and what the user would be interested in seeing. We illustrated an overview of the context-aware sports spectating state inference model in Fig. 2.

3.1.1 Spectator State

The user provides two main forms of context in the state inference model. First, closing in on the state perception part, the users' head movement in combination with the user's localization helps to determine the users' state (Fig. 3) and with this to start identifying which canvas is in the field of view of the user. The second is the user profile, which could be simply ascertained with a pre-game questionnaire or social media profile.

Using the onboard gyroscope sensors on the AR HMD and the assignment of "looking zones" in the stadium, we can identify where the users are looking. It is also possible to calculate the head rotation speed to determine how fast a user is looking around. These data then enter a temporary data array, where, in our case, the past ten entries taken with an interval of 0.1s are used to determine what user state the spectator is in. We have defined four user states, to begin with, although we acknowledge that there are probably more than

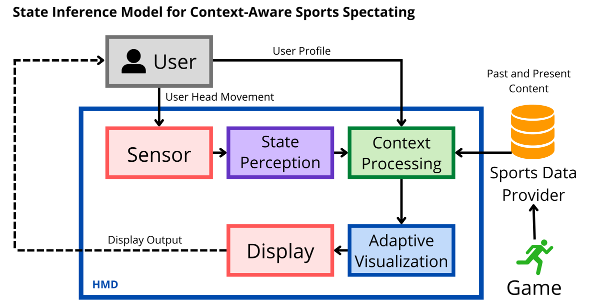

Figure 2: Overview of our state inference model based on Tsai et al. [35] which shows how the relevant visualization gets displayed to spectators based on the user, object, and scene context.

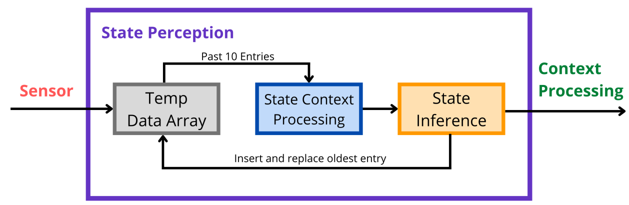

Figure 3: A closer look at the state perception part of the overall state inference model

four user states a spectator will encounter in a real game. The four states are **Focused**, **Information Gathering**, **Crowd Exploration**, and **Agnostic**.

Therefore, if the system detects that the user is only staring at one area (low head movement and lingering on a specific zone), it can infer that the user is probably focused on something; hence the focused state will be triggered. If the spectator is looking around, there will be a change in head movement speed and zones the spectators are looking at, leading to an information-gathering state. If the spectator is focused or panning slowly at the crowd, it is possible that the user is interested in what others are thinking or might want more engagement with the crowd. This movement will prompt the crowd-exploring state, displaying visualizations related to the crowd. While the state inference component then sends off the data to be processed with other contexts, a copy of it is inserted back into the temporary data array for the subsequent processing round and replaces the oldest entry.

3.1.2 Game State

While a spectator state is individualized to each user, the game state is the same for everyone attending that particular game. The game state's detection depends on an external sports data provider, which provides scene context to the system. For our prototype, we identified three general game states —**events**, **uneventful**, and **break time**. Events could then be broken down into major events, in which a scoring event occurs and minor events, such as fouls and substitutions. Characterizing all these game states allows us to display certain time-sensitive visualizations to the spectators and provide additional information to improve game understanding during uneventful times. During break time, if the spectator is still in their seat, it is possible to show full on-field situated visualization, which would not be meaningful during a game due to the overlapping of visualizations and players on the field. The sports data provider also provides the object context when needed. This is where certain player data could be displayed if, for example, the player scores or makes a foul play. If ball tracking is available, ball path visualizations could be integrated into the visualizations as part of the object context.

3.1.3 Level of Details

The level of detail (LOD) concept is integral to sports spectating visualizations. A low LOD would show only some essential information when applied to situated visualizations, while more details are provided with a higher LOD. Since sports spectators are likely to be focusing on the game happening on the field, it is essential not to overload the spectators with information when unnecessary. By introducing a LOD visualization method, spectators could get basic information about a particular game aspect and more detail if they desire to do so. Otherwise, the LOD will not increase and hence prevent visual clutter. The visualizations should disappear after a while or when they are no longer needed.

For our sports spectating scenario, we developed four levels of detail for most of our visualizations.

- LOD-0: This is simply a small icon or text to notify spectators when visualizations are available. We use this to present spectators with a choice of what they want to view. (Top right and bottom right of Fig. 1)
- LOD-1: The simplest form of visualization that involves data presentation. LOD-1 usually takes a graphical form and will not contain any form of text description. (Left of Fig. 1)
- LOD-2: This takes LOD-1 a step further and introduces some text elements to help describe some of the statistics shown. LOD-2 usually does not stray too far away from what LOD-1 looks like.
- LOD-3: This is where the full detail of the visualization is shown. It includes smaller text like numbers and percentages where the information gets specific. It could also contain additional data to supplement the LOD-2 visualization and, therefore, could be quite visually distinct from the previous LODs. (Middle of Fig. 1)

3.1.4 Context Processing

Here, we describe the relationship between the spectator and game states. The context processing step involves considering the spectator state, the game state, and the user profile to determine what visualization we should present. Besides the visualization itself, it is also here where the system manages the LOD of each visualization, deciding if it needs more details for the spectator.

We illustrated how all the components work together with a state transition diagram (Fig. 4). There are two methods to trigger the visualizations. The first method is if the system detects that the user is trying to gather information or is exploring the crowd, which is triggered by the user surveying the environment or looking at the crowd. At this stage, the LOD-0 visualizations would appear, and spectators could trigger specific visualizations if they wish to do so. By triggering a visualization, the LOD-1 visualization will be shown. A crowd-related visualization such as a poll will appear if the user is exploring the crowd.

The second method for triggering visualizations is identifying the game state, where the context processing starts. While during major events, it is pretty clear what we should show to the spectators, during an uneventful time, the system should decide if a particular visualization the spectator has not seen and might be interested in should be shown. This step is rather crucial as showing irrelevant content that spectators might not be interested in might distract them

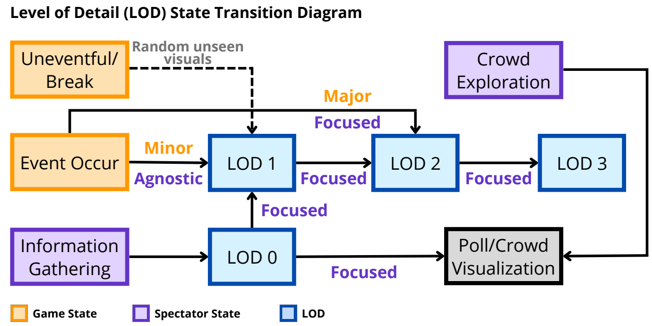

Figure 4: State transition diagram of the different LOD. Orange text are game-based and purple text are spectator-based states.

from the actual game rather than provide value. Usually, visualizations would take the form of LOD-1 in such cases; however, when coupled with other considerations such as event significance and user profile, it is possible to start directly with a LOD-2 visualization.

Regardless of which method triggered the LOD-1/LOD-2 visualizations (via information gathering spectator state or events in the game state), in order to increase the LOD, spectators would need to focus on the visualization. If the state perception determines that the spectator is focused, the visualization should evolve into the next LOD after a couple of seconds. Hence, this gives spectators a choice; if they do not focus on the visualization, it will disappear after a while. After getting to LOD-3, the visualizations will stay there if the spectators are still focused on it (it will take more time to consume such detailed content) and disappear after a while if the spectators are no longer focusing on the visualization.

4 Implementing the Context-aware Adaptive Interface for Sports Spectating

We implemented our context-aware adaptive interface in a VR prototype that serves as an indirect AR. Indirect AR prototypes are often a good method of simulating an AR experience in an off-site scenario [39]. Our indirect AR prototype is created with the Unity game engine, using a pre-captured 360° video of an existing game to simulate the experience of a Rugby Union game. Situated visualizations are then visualized in the stadium environment as what an AR HMD would show, using the 3'C's situated visualization framework for sports spectating [22] as guidance.

4.1 Requirements for Sports Spectating

In a sports spectating experience, there are three main points to take note of, which form our requirements for an XR sports spectating interface:

1. Spectators are there to have a good time. Whether experienced sports spectators or casual spectators, everyone's end goal is to have an **enjoyable experience**, hence, the interface should support this goal.
2. The stadium environment is **noisy and messy**. Unlike watching a movie in the cinema, the stadium environment is filled with life, from shouting sports fans to physical movements such as cheering and consuming food and beverages. The interface developed must adapt to such an environment.
3. **Interaction between fans** such as cheering, discussions, and celebrations are vital to a good stadium atmosphere. Therefore, the interface created should still allow fans to communicate with each other or even foster better interaction among the crowd.

Addressing these requirements, the XR interface for sports spectating needs to be unobtrusive so that spectators can still enjoy the stadium's atmosphere. Rather than being the most efficient information retrieval method, the interface's focus would be to provide an enjoyable experience that supplements the viewing experience. It would also need to withstand and adapt to the highly unpredictable nature of a sports game where for instance, fans could suddenly erupt in celebration.

To not disrupt the viewing experience, the interface would need to show information only when needed and at proper canvases in the stadium. This is to prevent sports spectators from getting distracted from the action happening on the field. The noisy and messy stadium environment also prevents us from implementing voice-based controls and other body measurements since spectators might be screaming and moving around. We excluded gesture controls, and spectators watching the sports would most likely not want to be distracted by the gestures of neighbouring spectators, in addition to the potentially high cognitive load needed to perform hand gestures. Image targets are also unsuitable for our use case as most actions are happening far from the field, and it would not make sense for sports spectators to be, e.g. carrying around image target cards.

More promising are considerations around the head and eye tracking, controllers, and the use of a virtual PIP. We decided that since head tracking is more commonly available in existing devices, we would go with head tracking for now. We opted for a semi-transparent dot in the centre of the screen, acting as a reticle (on-screen cursor). As for controllers, it would not be convenient for on-site spectators to carry around a bulky controller such as those used with VR headsets. If used at all, we aim to use a small controller with minimal control, such as the Hololen's clicker. As part of the ARSpectator experience, we implemented a context-aware adaptive interface to view visualizations in the stadium.

4.2 The Context-aware Adaptive Interface

The context-aware adaptive interface is an entirely hands-free interface. Using head-tracking as the interaction method, users interact with the virtual elements by looking at a visualization for two seconds to interact. Visualizations, by default, appear depending on the game context and spectator state as described in the state inference model earlier. This factors in the ongoing event happening on the field and the users' head movement.

For the game context, if there is, e.g. a penalty in the game, the penalty-related visualization might appear. At the same time, if there is not much action, some random visualizations that users have not explored will appear. This scenario is similar to how sports visualizations are fed to viewers in a sports broadcast. For the users' head movement, if the interface detects that users are looking for information, it will automatically spawn a menu that allows users to select what they want to see by simply looking at the various visualization options (Fig. 1, right).

The visualization options would appear on the canvas that the visualization is anchored. For example, suppose the score summary is to appear above the VIP seating box. In that case, the option to toggle it will also be co-located, reducing the need for spectators to look down to select a visualization. A ring will circle the cursor dot to indicate that an option is being selected. The visualizations also allow for cycling through different LOD simply by looking at them. If the user is not interested in the visualization and does not look at it, it will simply disappear after five seconds without proceeding to the next LOD.

5 Summary and Future Work

This paper explores the domain of using context-aware AR in on-site sports spectating. We looked into the design space for a context-aware interface and proposed a state inference model which could be used for various sports. We then implemented a context-aware adaptive interface and used a LOD system to show the appropriate amount of details and the proper time. However, our prototypes only partially implemented the overall state inference model, leaving room for improvement and future iterations.

The next steps are the evaluation of the interfaces and the inclusion of more context into the model. Our implementation at the moment does not consider visualization occlusion with players on the field and does not consider user profile in the LOD implementation. Our own experience of the interface suggest that for our context-aware adaptive interface to be successful, it has to be reasonably accurate in predicting what the user wants to see. This observation mirrors the finding by Rogers et al. [27] stating that fans will enjoy a sporting event more when they feel competent and autonomous. If the adaptive interface has not reached that level of competency, a manual approach that gives more control to the user is preferable. However, more research is needed to back up this point and potentially find the sweet spot to balance out autonomy and competency.

It is also necessary to test out such interfaces for an extended period of game time and with other participants simultaneously to test out the social acceptance of our interfaces. It was noted that when people are supposed to interact, the social acceptability of using VR HMD reduces [29]. This phenomenon might be the same for AR HMDs. Due to the limitations of the COVID-19 pandemic, we struggled to get enough live games where we could do longer high-quality recordings, not to mention conduct on-site studies. The duration of the video clips we use in our prototype is insufficient to see if the interface aided in game understanding or led to realizations of any particular insight. However, we hope this research helps propel AR use in on-site sports spectating when the relevant hardware becomes more advanced and accessible.

Acknowledgments

This project is supported by an MBIE Endeavour Smart Ideas grant. We thank Animation Research Ltd, Forsyth Barr Stadium, the Highlanders, Otago Rugby (ORFU), and OptaPerform for their support. We also would like to thank our HCI and Visual Computing labs members.

References

[1] G. D. Abowd, A. K. Dey, P. J. Brown, N. Davies, M. Smith, and P. Steggles. Towards a better understanding of context and context-awareness. In *International symposium on handheld and ubiquitous computing*, pp. 304–307. Springer, 1999.

[2] D. Ahlström, K. Hasan, and P. Irani. Are you comfortable doing that? Acceptance studies of around-device gestures in and for public settings. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*, pp. 193–202, 2014.

[3] A. Baca, P. Dabnichki, M. Heller, and P. Kornfeind. Ubiquitous computing in sports: A review and analysis. *Journal of Sports Sciences*, 27(12):1335–1346, 2009.

[4] M. Billinghurst, R. Grasset, and J. Looser. Designing augmented reality interfaces. *ACM Siggraph Computer Graphics*, 39(1):17–22, 2005.

[5] M. Buric, M. Ivasic-Kos, and M. Pobar. Player tracking in sports videos. In *2019 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, pp. 334–340. IEEE, 2019.

[6] M. Carter, F. Allison, J. Downs, and M. Gibbs. Player identity dissonance and voice interaction in games. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, pp. 265–269, 2015.

[7] R. Cavallaro, M. Hybinette, M. White, and T. Balch. Augmenting Live Broadcast Sports with 3D Tracking Information. *IEEE MultiMedia Magazine*, 2011. doi: 10.1109/MMUL.2011.61

[8] Y. Cheng, Y. Yan, X. Yi, Y. Shi, and D. Lindlbauer. Semanticadapt: Optimization-based adaptation of mixed reality layouts leveraging virtual-physical semantic connections. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, pp. 282–297, 2021.

[9] V. Clay, P. König, and S. Koenig. Eye tracking in virtual reality. *Journal of eye movement research*, 12(1), 2019.

[10] A. M. da Silva, S. G. Gustavo, and F. P. A. de Medeiros. A Review on Augmented Reality applied to Sports. In *2021 16th Iberian Conference on Information Systems and Technologies (CISTI)*, pp. 1–6. IEEE, 2021.

[11] R. Dasgupta, R. Dasgupta, and Srivastava. *Voice User Interface Design*. Springer, 2018.

[12] E. L. Deci and R. M. Ryan. Self-determination theory. 2012.

[13] B. Ens and P. Irani. Spatial Analytic Interfaces: Spatial User Interfaces for In Situ Visual Analytics. *IEEE computer graphics and applications*, 37(2):66–79, 2016.

[14] R. Grasset, T. Langlotz, D. Kalkofen, M. Tatzgern, and D. Schmalstieg. Image-Driven View Management for Augmented Reality Browsers. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2012)*, 2012.

[15] H. Ishii. The tangible user interface and its evolution. *Communications of the ACM*, 51(6):32–36, 2008.

[16] J. D. James and G. T. Trail. The relationship between team identification and sport consumption intentions. *International Journal of Sport Management*, 9(4):427–440, 2008.

[17] P. Jost, S. Cobb, and I. Hämmerle. Reality-based interaction affecting mental workload in virtual reality mental arithmetic training. *Behaviour & Information Technology*, 39(10):1062–1078, 2020.

[18] G. Klein and D. Murray. Parallel tracking and mapping for small AR workspaces. In *2007 6th IEEE and ACM international symposium on mixed and augmented reality*, pp. 225–234. IEEE, 2007.

[19] S. M. LaValle, A. Yershova, M. Katsev, and M. Antonov. Head tracking for the Oculus Rift. In *2014 IEEE international conference on robotics and automation (ICRA)*, pp. 187–194. IEEE, 2014.

[20] D. Lindlbauer, A. M. Feit, and O. Hilliges. Context-aware online adaptation of mixed reality interfaces. In *Proceedings of the 32nd annual ACM symposium on user interface software and technology*, pp. 147–160, 2019.

[21] W. Liu, X. Li, and D. Huang. A survey on context awareness. In *2011 International Conference on Computer Science and Service System (CSSS)*, pp. 144–147. IEEE, 2011.

[22] W. H. Lo, S. Zollmann, and H. Regenbrecht. Stats on-site—Sports spectator experience through situated visualizations. *Computers & Graphics*, 2021.

[23] E. O. Lopez-Caudana. Active Noise Cancellation: The Unwanted Signal and the Hybrid Solution. *Adaptive Filtering Applications, Dr. Lino Garcia (Ed.)*, pp. 49–84, 2011.

[24] S. Melax, L. Keselman, and S. Orsten. Dynamics based 3D skeletal hand tracking. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, p. 184, 2013.

[25] J. Mercier-Ganady, F. Lotte, E. Loup-Escande, M. Marchal, and A. Lécuyer. The Mind-Mirror: See your brain in action in your head using EEG and augmented reality. In *2014 IEEE Virtual Reality (VR)*, pp. 33–38. IEEE, 2014.

[26] X. Qian, F. He, X. Hu, T. Wang, A. Ipsita, and K. Ramani. ScalAR: Authoring Semantically Adaptive Augmented Reality Experiences in Virtual Reality. In *CHI Conference on Human Factors in Computing Systems*, pp. 1–18, 2022.

[27] R. Rogers, K. Strudler, A. Decker, and A. Grazulis. Can Augmented-Reality Technology Augment the Fan Experience?: A Model of Enjoyment for Sports Spectators. *Journal of Sports Media*, 12(2):25–44, 2017.

[28] D. Schmalstieg, A. Fuhrmann, G. Hesina, Z. Szalavári, L. M. Encarnaçao, M. Gervautz, and W. Purgathofer. The studierstube augmented reality project. *Presence: Teleoperators & Virtual Environments*, 11(1):33–54, 2002.

[29] V. Schwind, J. Reinhardt, R. Rzayev, N. Henze, and K. Wolf. Virtual reality on the go? a study on social acceptance of vr glasses. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pp. 111–118, 2018.

[30] B. Shneiderman, A. Publ, and N. J. Norwood. 1.1 Direct manipulation: a step beyond programming. *Sparks of innovation in human-computer interaction*, p. 17, 1993.

[31] M. Stein, H. Janetzko, A. Lamprecht, T. Breitkreutz, P. Zimmermann, B. Goldlücke, T. Schreck, G. Andrienko, M. Grossniklaus, and D. A. Keim. Bring it to the pitch: Combining video and movement data to enhance team sport analysis. *IEEE transactions on visualization and computer graphics*, 24(1):13–22, 2018.

[32] X. Sun and A. May. The role of spatial contextual factors in mobile personalization at large sports events. *Personal and Ubiquitous Computing*, 13(4):293–302, 2009.

[33] Z. Szalavári and M. Gervautz. The personal interaction Panel–a Two-Handed interface for augmented reality. In *Computer graphics forum*, vol. 16, pp. C335—-C346. Wiley Online Library, 1997.

[34] P. Temoche, E. Ramirez, and O. Rodriguez. A low-cost data glove for virtual reality. In *Proceedings of XI International Congress of Numerical Methods in Enginnering and Applied Sciences (CIMENICS).[internet]*, vol. 2012, 2012.

[35] C.-H. Tsai and J.-Y. Huang. Augmented reality display based on user behavior. *Computer Standards & Interfaces*, 55:171–181, 2018.

[36] A. S. Tulshan and S. N. Dhage. Survey on virtual assistant: Google assistant, siri, cortana, alexa. In *International symposium on signal processing and intelligent recognition systems*, pp. 190–201. Springer, 2018.

[37] J. Tumler, F. Doil, R. Mecke, G. Paul, M. Schenk, E. A. Pfister, A. Huckauf, I. Bockelmann, and A. Roggentin. Mobile augmented reality in industrial applications: Approaches for solution of user-related issues. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 87–90. IEEE, 2008.

[38] R. B. Ventura and M. Porfiri. Galvanic Skin Response As a Measure of Engagement During Play in Virtual Reality. In *Dynamic Systems and Control Conference*, vol. 84270, p. V001T17A003. American Society of Mechanical Engineers, 2020.

[39] J. Wither, Y.-T. Tsai, and R. Azuma. Indirect augmented reality. *Computers Graphics*, 35(4):810–822, 2011.

[40] N. Xi, J. Chen, F. Gama, M. Riar, and J. Hamari. The challenges of entering the metaverse: An experiment on the effect of extended reality on workload. *Information Systems Frontiers*, pp. 1–22, 2022.

[41] L. I. Yang, J. Huang, T. Feng, W. Hong-An, and D. A. I. Guo-Zhong. Gesture interaction in virtual reality. *Virtual Reality & Intelligent Hardware*, 1(1):84–112, 2019.

[42] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C.-L. Chang, and M. Grundmann. Mediapipe hands: On-device real-time hand tracking. *arXiv preprint arXiv:2006.10214*, 2020.

[43] R. Zhang, L. Wu, Y. Yang, W. Wu, Y. Chen, and M. Xu. Multi-camera multi-player tracking with deep player identification in sports video. *Pattern Recognition*, 102:107260, 2020.

[44] S. Zollmann, T. Langlotz, M. Loos, W. H. Lo, and L. Baker. Arspectator: Exploring augmented reality for sport events. In *SIGGRAPH Asia 2019 Technical Briefs*, pp. 75–78. 2019.