Sports Visualization in the Wild: The Impact of Technical Factors on User Experience in Augmented Reality Sports Spectating

Wei Hong Lo, *University of Otago, Dunedin, New Zealand*
Holger Regenbrecht, *University of Otago, Dunedin, New Zealand*
Stefanie Zollmann, *University of Otago, Dunedin, New Zealand*

Abstract—The maturity of Augmented Reality (AR) technology allows for expansion into real-world applications, including visualizations for on-site sports spectating. However, it's crucial to understand the factors influencing user experience in AR applications. To optimize user experience, we conducted a user study where participants adjusted parameters to determine noticeable and disruptive values of latency, registration accuracy, and jitter using a mobile indirect AR prototype that simulates a rugby stadium experience. Our findings indicate that latency has the highest disruptive impact on users' experience, with registration accuracy following closely. Furthermore, when noticeable latency, registration accuracy, and jitter were combined, the user experience was negatively affected in a non-linear, combinatorial manner. This suggests that addressing factors individually is necessary but not enough for successful user experiences. By understanding these factors, developers can optimize AR experiences when creating immersive AR sports experiences and other large-scale AR applications to ensure maximum enjoyment for users.

In recent years, Augmented Reality (AR) research has progressed significantly with improved hardware and greater access to AR software development kits (SDKs). As a result, researchers are exploring diverse AR use cases in fields like tourism, education, medical surgery, etc. However, there is limited research on AR for large-scale environments, and it remains unclear how well-known technical factors influence user experience.

This article builds on our prototype in [13] and addresses the gap in understanding technical factors affecting user experience in large-scale AR environments, particularly within sports spectating contexts in stadiums. We have developed an AR scenario focusing on the use of AR in sports spectator contexts that allows sports fans to enhance their on-site experience with situated visualizations [19], aiming to provide real-time game statistics and information computed from a vision-based player tracking system using stadium-installed birds-eye cameras.

While this architecture is common for providing similar information to live TV audiences, which often use pre-calibrated cameras on smooth-moving tripods with a slight delay, replicating it on handheld devices in real time introduces technical factors such as latency, registration accuracy, and jitter that can affect the AR experience. These factors stem from both the client device used for the AR experience and the dynamic content provider. Client device factors mainly result from camera registration and tracking inaccuracies, while dynamic content provider factors influence the visualizations of dynamic content.

Previous work has explored the use of AR in on-site sports spectating scenarios, where spectators utilize mobile devices or AR head-mounted displays (HMD) to access additional information through situated visualization [13]. Although the study identifies relevant technical factors for AR experiences, it does not fully investigate the influence of these factors on user experience. Consequently, it remains unclear which optimizations are necessary for an effective large-scale

Digital Object Identifier 10.1109/XXX.0000.0000000

AR experience and to what extent technical factors impact user experience.

Addressing this gap, we pose an important yet unexplored question for situated visualization in the wild: How do technical factors, such as system limitations or errors, affect user experience in a large-scale AR system, particularly when dynamic content is presented? To answer this question, we designed a user study with the following research questions to determine which technical aspects should be improved in an AR system to enhance user experience.

RQ1) Among the selected technical factors, are there any factors that are more disruptive than others to the user experience in a large-scale environment AR application with dynamic content?

RQ2) For each of the technical factors, what are the noticeable levels and disruptive levels of those factors where spectators might start to be negatively affected in terms of user experience?

RQ3) If we were to combine the noticeable levels of the different factors that affect the user experience, would the effects accumulate?

We developed these research questions to determine the relative importance of various technical factors, identify their noticeable and disruptive levels, and examine the combined effects when these factors coexist. By addressing these questions, we seek to uncover insights that will help guide the optimization of AR systems for large-scale environments, ultimately enhancing user experiences in these contexts. Previous research has focused on measuring mental load and task performance [6]. However, sports spectating differs as spectators typically want to observe and access additional information in a relaxed, enjoyable setting—similar to watching television—without the pressure to complete tasks accurately or within a specific time frame. Consequently, we shifted our study's focus to user experience rather than performance metrics, emphasizing the unique characteristics of sports spectating contexts.

Our research intends to serve as a foundation for enhancing user experiences impacted by technical factors and limitations in large-scale AR environments. By identifying acceptable levels of latency, registration accuracy, and jitter in sports spectating scenarios, we provide guidance for optimizing AR experiences. We also offer insights from our user study, including usability issues and suggestions for future work.

Our findings highlight latency as the most influential factor affecting user experience in sports spectating contexts, followed closely by registration accuracy. We also discovered that combining multiple noticeable values of factors significantly increases disruption to the user experience. To the best of our knowledge, no previous user studies have explored tailored manipulation of these technical factors in AR. Our research addresses the gap of the lack of user evaluation in AR publications [10], particularly in relation to technical factors influencing user experiences.

Background

Our research centers on AR in sports, a growing area of study, where a quick search will reveal that most efforts have focused on training, performance analysis, and broadcasting. We aim to enhance the on-site sports spectating experience by providing spectators with better insights through situated visualization. Our use-case also does not require millimeter accuracy as it is used in a large environment, compared to most AR applications done in a small AR workspace.

Traditionally, AR research has predominantly concentrated on human factors or technical aspects such as optimizing tracking algorithms or display properties. Human factors involve designing products tailored to the target audience, ensuring safe designs, and effective user interfaces. In AR studies involving human participants, performance measures often include goal achievement, task completion rates, ease of use, and knowledge gain [6]. On the other hand, technical papers propose novel algorithms and techniques for tracking, registration, and display types, focusing on measurements like tracking rates, accuracy, and pose errors [5] without investigating their impact on user experience. Until now, investigations into human and technical factors have largely remained separated.

Technical Evaluations in AR

Kim et al. [10] identified tracking and localization methods as the most popular topics in AR research for over two decades. A vast body of computer vision-based research has explored various algorithms for Simultaneous Localization and Mapping (SLAM) [12], including SLAM-based methods for large-scale environments. While tracking and localization are crucial technical factors for user experience, these studies typically focus on technical accuracy results rather than user experience measures. From this technical research, we derived registration accuracy and jitter as key factors for our user study.

Latency is another factor influenced by the tracking and localization methods of an AR application. Computationally expensive methods increase latency due to extended processing times. Efforts to reduce latency have been investigated in recent years, and researchers have examined how networking technologies could be improved to support the massive offloading of mobile AR applications, which introduce latency to digital content visualizations [2]. However, similar to tracking and localization research, most work on optimizing latency is technically focused and lacks user involvement. Consequently, the effect of latency on user experience remains unclear.

User Experience Evaluations in AR

A survey by Dey et al. [4] shows that user evaluations in AR primarily target interaction and perception, often focusing on performance-based evaluations, such as the NASA TLX questionnaire [6], to determine mental load. However, user experience typically extends beyond this, encompassing usability, pragmatic, and hedonic aspects of a system [17]. On the other hand, usability metrics and acceptance questionnaires do not evaluate experiential and emotional aspects. In terms of user experience evaluation methods, not many of them have been applied in the context of AR use cases.

Some studies include questions about enjoyment, while others evaluate technical factors like latency, frame-rate, resolution, and jitter to measure their effects on presence [14]. Some claim to measure user satisfaction, but their questionnaires primarily assess usability. Iwata et al. [8] utilized intrinsic motivation to measure learning motivation, which includes some aspects of user experience in the interest/enjoyment sub-scale, and conducted a free-writing style survey to gain better insights from participants. However, these studies may not provide a comprehensive evaluation of user experience, as their primary focus lies elsewhere, merely collecting user experience-related feedback from participants.

Mixed Reality Solutions for AR Evaluations

The challenges of conducting reliable AR evaluations arise from various technical aspects, such as constant lighting changes or item placement in the real world [16]. Simulating AR using Virtual Reality (VR) has been explored as a potential solution, allowing for complete control over a virtual environment where factors can be isolated and manipulated individually. This enables evaluation formats that are not possible in actual AR scenarios [16]. Mixed reality setups can also provide simulated AR experiences that are not yet technologically feasible, such as a wide field-of-view AR. In some cases, mixed reality approaches like CAVE systems might even simplify evaluations compared to actual AR setups.

However, there are still challenges with using mixed reality solutions to evaluate AR, such as incorrect depth perception and the lack of tactile feedback from the real environment. Despite these limitations, mixed reality solutions remain the most suitable option for evaluating and manipulating isolated technical factors in large environments, as in our use case. Consequently, we implemented an indirect AR approach [20] where AR is simulated with a 360° video for our user study.

Technical Factors Affecting AR User Experience in Sports Spectating

Olsson et al. [15] conducted a large-scale online survey that involved more than two thousand participants in AR mobile applications. In their survey, users identified technical and functional problems as the main weaknesses of AR applications. As an example, they often mentioned imprecision in localizing as a technical problem. This is related to our target registration accuracy and jitter factors. Additionally, latency is one of the main factors in AR applications [7] and is also associated with software deficiencies (for tracking) and hardware limitations (networking infrastructure and device limitations).

Latency, registration accuracy, and jitter are common technical factors in an AR sports spectating scenario. In our research, we focus mainly on the two sources of these factors — the client device camera and the player tracking system (Figure 1). We did not include the sports data provider as they would also have their technical factors based on their system, which is well outside our control. More potential factors could affect the AR user experience, but we want to focus on the factors caused by technical limitations. Therefore, this thesis does not consider the design or human factors such as visualization style, information cluttering, depth perceptions, ergonomics, etc.

Latency

Latency in video see-through (VST) AR systems is the difference in time from augmented visualization to reality. There are a few sources of system delay in a VR system [9], which is mostly similar to VST AR systems. The five sources of system delay in temporal order are as follows:

1) *Tracking delay* - Refers to the delay caused by the tracking system. This delay is present in various areas involving tracking, such as the sports data provider's in-house tracking system,

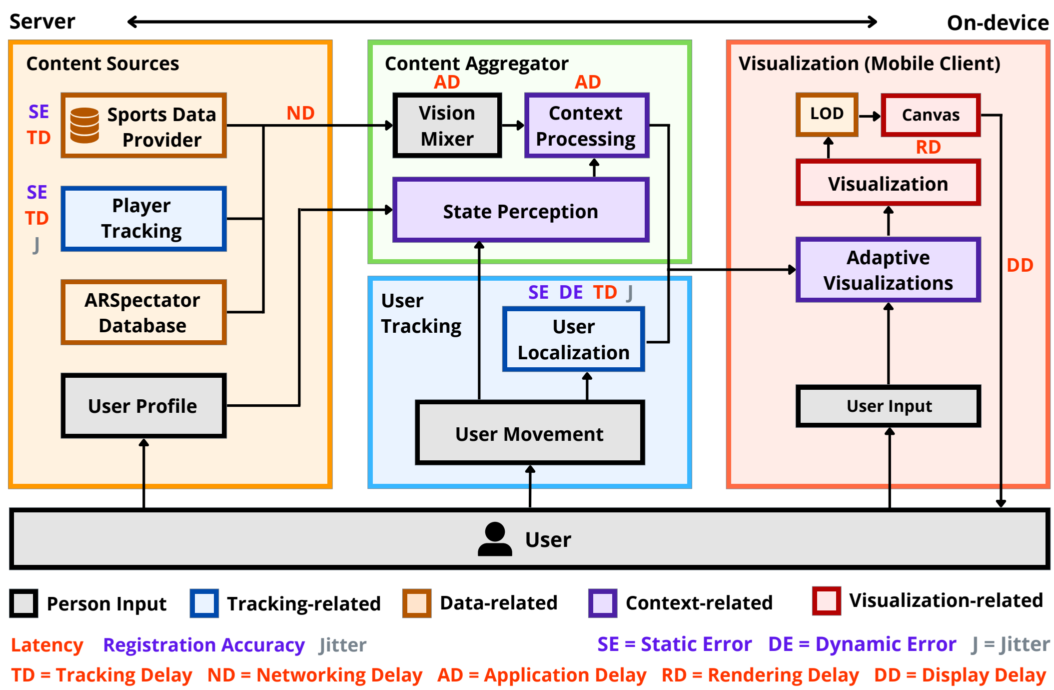


FIGURE 1. The various parts of the AR sports spectating overview where the three technical factors that we are investigating exist. The three technical factors are latency, registration accuracy and jitter.

our vision-based player tracking system, and our user localization component that tracks users' pose and position in the stadium.

2) *Networking delay* - Represents the delay in data communication over the network. In our case, *networking delay* mainly occurs in the content source, particularly when the content source is transferred to the content aggregator.
3) *Application delay* - This delay is task-specific and not related to other delays. In our scenario, *application delay* mostly occurs in the content aggregator, where the vision mixer may experience delays in annotation or context processing to determine what users want to see based on their state and incoming information.
4) *Rendering delay* - Refers to the delay in rendering graphics through the hardware's graphics component. *Rendering delay* depends on the complexity of the scene and varies over time. In our use case, *rendering delay* represents the delay in producing the scene with all visualizations. VST AR applications will have higher rendering delays compared to OST AR setups, as the entire environment needs to be rendered.
5) *Display delay* - This delay occurs in the final step when the hardware presents the visualizations to the user by activating the correct pixels. In our use case, *display delay* should be around 16.7 milliseconds or less, as most display hardware now operates at a minimum of 60Hz, with some VR HMDs reaching over 100Hz.

Our primary focus is on investigating the relative latency of player tracking, which is the latency between two data streams resulting in misregistration, commonly referred to as lag. The relative latency encompasses the *tracking delay* required to obtain the players' positions on the field and the *networking delay* needed to deliver all information to the mobile client. We aim to simulate the delay between the tracking data stream and the actual scene captured on camera. This latency has a significant impact on dynamic content on the field, particularly visualizations that follow a player, such as arrows or highlights (Figure 2, left).

We focused on the relative latency of player tracking as simulating OST or VST AR experiences with indirect AR isn't feasible due to display latency (100-150ms). Including camera latency in indirect AR would introduce confounding variables. Even in VST AR, latency above 50ms is perceptible and above 150ms affects presence [3]. We narrowed our study's scope by disregarding *application delay*, *rendering delay*, and *display delay*, and focused on *tracking delay* and *net-*

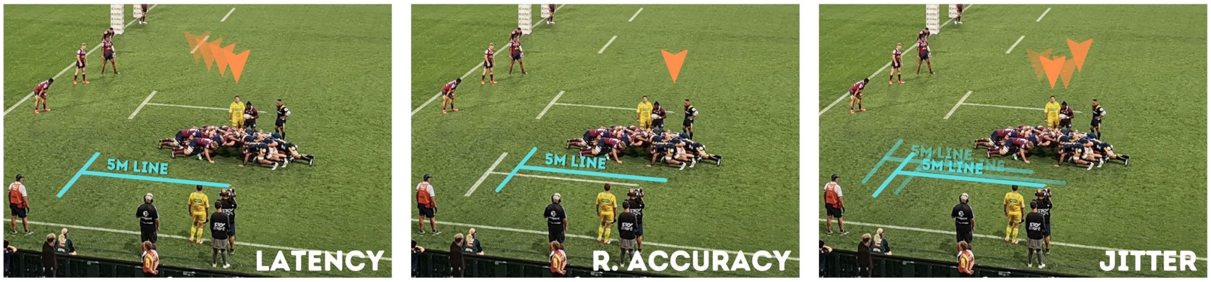


FIGURE 2. An illustration of technical factors evaluated: latency, registration accuracy, and jitter. Visualizations are shown in blue (static) and orange (dynamic) where the arrow is targeting the person in yellow.

working delay, which are more consistent and reliably simulated in our prototype.

Registration Accuracy

Registration accuracy in our scenario aims to mimic the inconsistencies in tracking and initialization. In real applications, this could be due to a misalignment in the initialization/localization phase. However, this is often present, especially in image-based registration. It is found that the human brain can handle inconsistencies in such registrations, and an approximate registration might be sufficient [11]. Therefore, we would like to determine how much inconsistency would be acceptable.

There are two types of registration errors, namely *static errors* and *dynamic errors* [1]. *Static errors* are registration errors that occur even when the user is stationary and not moving. For our use case, this translates into registration errors from the content source — the sports data provider and player tracking, and the user localization (device camera). On the contrary, *dynamic errors* occur if the user moves their head, usually caused by the temporal mismatch between the user's head movement and the display. This error would only occur in the user localization part, where the display lags behind the users' movement. We are only interested in the *static errors* for our case. Similar to latency, we are not evaluating the technical factors caused by head movements but rather the technical factors related to position tracking.

The registration accuracy is simulated by a consistent offset of the visualization from the original location, reflecting an error in placement or alignment. The accuracy of the client device camera registration affects every visualization in the environment. In contrast, the accuracy of the player tracking registration only affects the dynamic spatial visualization showing, for instance, the positioning of the players (Figure 2, middle).

Jitter

Jitter refers to the small, irregular movements of visualizations caused by precision errors in tracking and anchoring. There are two types of jitter in extended reality (XR) systems: *spatial jitter* and *temporal jitter*. *Spatial jitter* typically results from noise in the device signal and the user's hand movements. In our sports spectating scenario, this includes noise from player tracking software, user localization, and the mobile client device, such as a phone or headset.

Temporal jitter, on the other hand, is caused by inconsistent latency, making visualizations sometimes delayed and other times not. Although *temporal jitter* might be present in our sports spectating scenario, we focus on *spatial jitter* in our research as *temporal jitter* are more towards the research area of networking.

Reducing *spatial jitter* by smoothing out the visualizations could increase latency. In our system, *spatial jitter* primarily comes from noise and tremors in the device camera and player tracking system. Although we recognize that users' arm or head movements can introduce *spatial jitter*, we only simulate *spatial jitter* caused by noise in the signal of player tracking or user localization. Noise-induced *spatial jitter* from the device camera affects all visualizations (static and dynamic), while *spatial jitter* from player tracking only impacts dynamic visualizations (Figure 2, right).

Evaluation of AR UX Factors in Sports Spectating

We designed a user study for AR sports spectating using mobile indirect AR, focusing on mobile VST due to accessibility and familiarity. To eliminate confounding factors, we recreated the AR stadium experience in a lab using a mobile prototype, allowing accurate content alignment and control of latency, registration accuracy, and jitter. The prototype uses a scaled stadium model as an anchor for visualizations. This study received

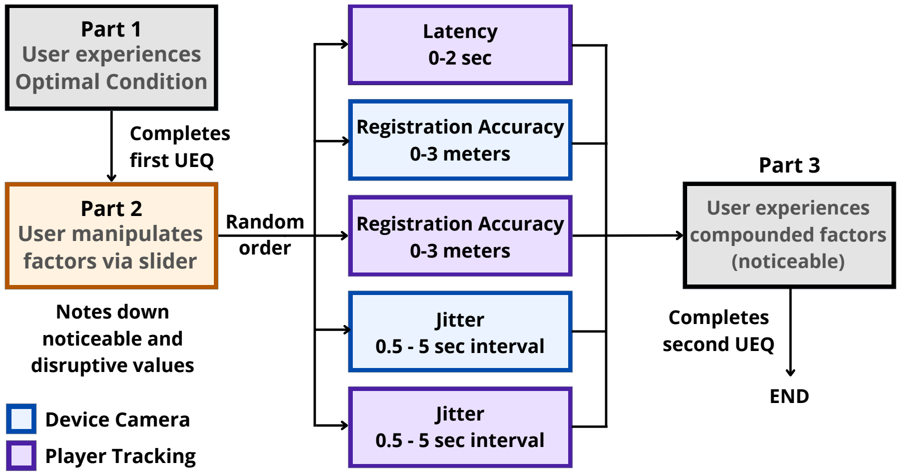


FIGURE 3. The flow of the user study investigating the noticeable value and disruptive value of the three technical factors — latency, registration accuracy and jitter. In all cases, except latency, we look into two sources; the device camera and the player tracking system.

ethical approval from the Ethics Committee of the University of Otago and adhered to health and safety precautions, including pandemic-related measures.

Study Design

To evaluate the impact of specific latency, registration accuracy, and jitter on user experience, we designed a within-subject study that allowed participants to manipulate these factors. We aimed to determine which factor had the most significant negative impact on user experience and identify noticeable and disruptive levels for each AR technical factor. The noticeable level is when participants notice an impact on their user experience, while the disruptive level is when they feel the application is almost unusable. The dependent variables are 1) user experience and 2) the value at which a factor is noticeable and disruptive. The independent variables are the three technical factors.

The user study has three parts (Figure 3): 1) experiencing the prototype in optimal conditions, 2) manipulating AR factors to obtain noticeable and disruptive values, and 3) experiencing the prototype with compounded factors. In each part, participants view a one-minute looping indirect AR session using a 360° video of a rugby game with overlaid visualizations. The first and last parts use the same video for comparison, while a different clip is used during manipulation. Participants only observe the game, and after parts one and three, they complete the User Experience Questionnaire (UEQ) [1] for comparison.

FIGURE 4. User study session where a participant controls a slider to manipulate the technical factor. The goal is to set the slider until they obtain the noticeable and disruptive levels. Shown in the image is latency.

During the second part, which involves manipulating factors, the technical factor and a brief instruction statement appear at the top of the screen. A slider and a button are located at the bottom of the screen (Figure 4). The slider controls the linear manipulation of the specific factor, while the button is for participants to press when they notice an effect on user experience. Upon selecting the noticeable value, the slider resets, and the button's wording changes to "Disruptive," al-

[1] https://www.ueq-online.org/Material/Handbook.pdf

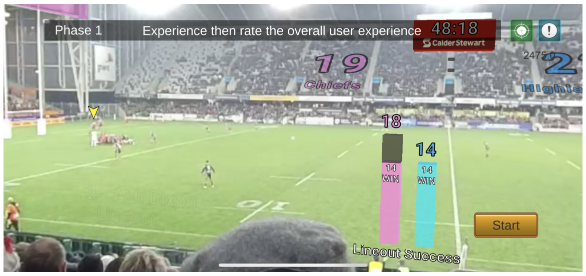


FIGURE 5. Screenshot taken out of phase 1 of the user study (optimal condition user experience rating) showing a situated 3D bar chart of the line-out statistic of both teams.

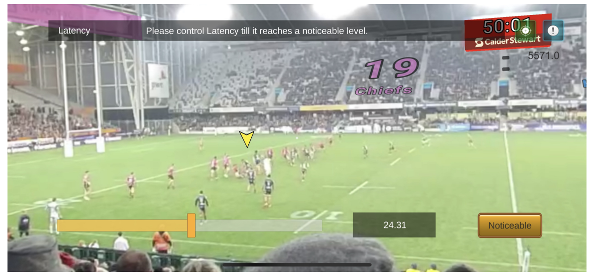


FIGURE 6. Our user study prototype showing the latency factor delayed by 24 frames (1 second) controlled by the slider at the bottom. Optimally, the arrow should point at the referee (in white outfit).

lowing participants to manipulate the factor until it represents a disruptive user experience. After that, a rating panel appears for participants to rate the factor's negative impact on user experience from 1 to 7. This repeats throughout the second part of the study with a different randomized technical factor.

Part three is an identical clone of part one, except that the noticeable value of all the device camera factors and the latency of the player tracking system is incorporated into the scene. These are based on what the participants selected earlier in part two. Again, there is no interaction required from the participants.

The visualizations used in the user study were kept simple to avoid complicating the study and introducing more variables. They included scores and timers on the opposite stands, on-field graphics, event visualizations, and a yellow arrow pointing at the referee. The arrow pointing at the referee simulates player tracking, using the referee's differently colored jersey to avoid confusion with other players. Occasional statistical visualizations, such as bar charts, were also used as shown in Figure 5.

We conducted three pilot tests to refine the experimental procedure and study parameters. The initial parameters were adjusted for latency, registration accuracy, and jitter. We addressed several issues identified during the pilot tests, such as redefining the noticeable value, informing participants about what they were manipulating, and heuristically setting the jitter magnitude based on discussion with our expert panel.

Finally, we adjusted the final parameters: For latency, the slider ranged from optimal to a delay of 48 frames (2 seconds). Registration accuracy parameters shifted from optimal to a maximum of 300 cm (3m), based on pilot feedback. Jitter frequency ranged from 5 seconds (less frequent) to 0.5 seconds (frequent) per sequence, with a magnitude of 1m. Each sequence had random jitters in either the x or y-axis, occurring one to six times, mimicking typical AR applications.

Apparatus

For this study, we captured 360° videos of a real rugby game in a stadium using an Insta One X 360° camera. We utilized the video as a dynamic background in the indirect AR prototype. We manually aligned a 3D model of the stadium with the 360-degree video to ensure a precise and consistent fit. We then overlaid visualizations onto the environment to simulate an actual in-stadium experience during a live game.

The study was conducted using an iPhone X, with some questionnaire elements (factor rating) completed on the device. Other questionnaires (UEQ and additional questions) were completed on paper. We also observed participant behavior and recorded noteworthy observations on paper. The prototype was developed using Unity[2] and Vuforia[3].

Participants

We recruited 20 participants from the university for this study, mostly through word of mouth. There are no prerequisites; no sports knowledge was needed to rate the user experience. Participants we recruited are between 20 and 30 years old ($\bar{x} = 24$, $\sigma = 2.66$), consisting of 11 women and 9 men. Among all participants, two participants have been to a game in the stadium at least ten times, while two participants have never been to a game.

[2] https://unity.com
[3] https://developer.vuforia.com

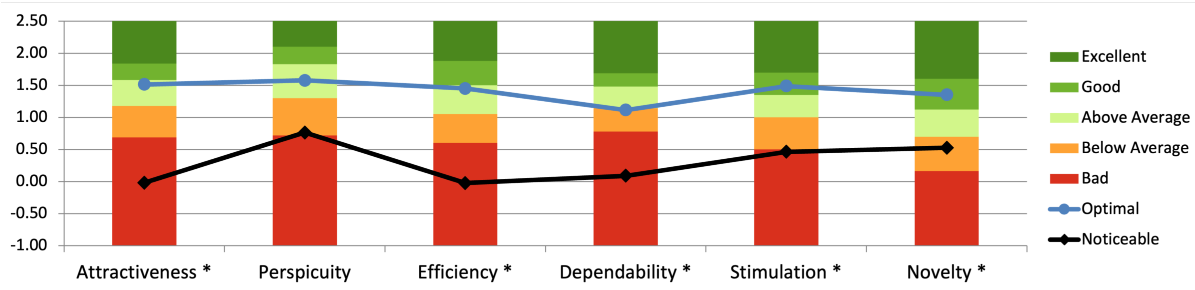


FIGURE 7. UEQ comparison between optimal configuration compared to noticeable configuration. Components with * have significant differences between the conditions. Both curves are mapped to the original UEQ benchmark. As expected, the optimal configurations scored higher than the noticeable configurations. The scores for the optimal configurations are mostly at above average and good. The scores for the noticeable configuration were mostly in the bad range. Our statistical analysis shows that there are significant differences between all components in the UEQ except for perspicuity.

Procedures

In the study, participants first read the information sheet and completed a demographic questionnaire and consent form. After a brief introduction, they began with the first part of the study, where they explored an optimal indirect AR experience of a real rugby game with overlaid visualizations. They then rated their user experience using the UEQ, which served as a baseline for comparison with other conditions.

Next, participants were briefed on the second part, which involved manipulating five technical factors using a slider (Figure 6). They adjusted the factors until they noticed an impact on their user experience and then pressed a button to reset the condition. They repeated this process for all five factors in a randomized order. In the third part, participants experienced a combination of the noticeable values for all factors. They completed another UEQ and indicated whether they would still use the prototype if it resembled the noticeable value condition. Finally, they were invited to mention other factors they believed could impact user experience.

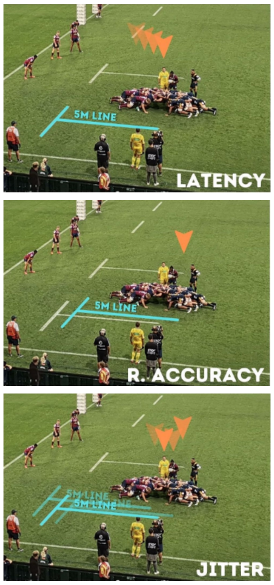
Results

The study analyzed participants' ratings of the technical factors using a 7-point Likert scale. The results showed that latency had the highest negative impact on user experience ($\bar{x}$ = 5.6, σ = 0.94), followed by registration accuracy (player tracking) ($\bar{x}$ = 5.5, σ = 0.95), registration accuracy (device camera) ($\bar{x}$ = 4.9, σ = 1.29), and jitter (device camera) ($\bar{x}$ = 4.85, σ = 1.81). Jitter (player tracking) had the smallest impact ($\bar{x}$ = 4.25, σ = 1.59). A Friedman test found a significant effect of the factors on the rating (Friedman's chi-squared = 10.507, df = 4, p = 0.033). Wilcoxon signed-rank tests showed significant differences between between latency (player tracking) and jitter (player tracking) (p = 0.007), registration accuracy (player tracking) and jitter (player tracking) (p = 0.007), as well as between registration accuracy for device camera and player tracking (p = 0.035). There were also various effect size across the factors.

Interestingly, for registration accuracy, the offset of dynamic content was rated to have a greater impact than the offset on the entire scene ($\bar{x}$ = 5.5 compared to 4.9), despite the factor of the device camera that also affects the dynamic content. This is probably due to participants only focusing on the dynamic content during the player tracking registration accuracy factor; therefore, it makes a bigger impact than when everything is offset, and the participants focus on the overall picture. Participants found that the jittering of all visualizations was more disturbing ($\bar{x}$ = 4.85) compared to when only the player tracking visualization was affected ($\bar{x}$ = 4.25). In general, all of the tested technical factors do affect the user experience, as each one averaged above the midpoint of 3.5.

Comparing the UEQ results for optimal and noticeable configurations showed lower measurements for the noticeable configuration, as expected. Two-sample t-tests revealed significant differences ($p<0.05$) between all UEQ components except for perspicuity (p=0.08). In terms of individual scores against the benchmark obtained from the UEQ Data Analytics Tool - version 4 (Figure 7), the optimal condition scored mostly above average and good except for dependability, which was on the border of below average. Meanwhile, the noticeable conditions scored mostly bad except for perspicuity and novelty, which still hovers in the below-average range.

In terms of consistency for both UEQ, both con-



Technical Factor	Noticeable Value	Disruptive Value
Latency (Player Tracking)	373.27ms	654.90ms
Registration Accuracy (Device Camera)	61.52cm	105.78cm
Registration Accuracy (Player Tracking)	62.35cm	101.86cm
Jitter - Interval (Device Camera)	4.62s	3.00s
Jitter - Interval (Player Tracking)	4.50s	2.81s

FIGURE 8. Average results for each factor comparing the noticeable value with the disruptive value. The factors from Figure 2 are placed on the left side for easier comprehension. Device camera factors affect all visualizations while player tracking only affects dynamic content such as the arrow as shown.

ditions have a Cronbachs Alpha of >0.7, with the noticeable condition having a higher average of the Cronbachs Alpha (0.87) compared to the optimal condition (0.81), showing high confidence in the scale results. For inconsistencies, in both optimal and noticeable conditions, there is a participant (different participant in both conditions) with 3/6 categories where the answers are inconsistent with the scale (rating discrepancy for specific attributes in the same category). This is checked by seeing how much an item's best and worst evaluation on a scale differs. Since it was two different participants under different conditions, we assumed that it was not intentional and that the participant probably did not understand everything on the scale properly. The UEQ does contain some aspects that might not be relevant to every scenario (such as secure vs not secure) which might cause some confusion and lead to discrepancies on the scales.

Discussion and Limitations

Answering **RQ1**, jitter had the least impact on user experience compared to latency and registration accuracy. Registration accuracy for dynamic content, such as pointing at the wrong player, greatly impacted user experience, and latency had a considerable effect with a disruptive value below one second.

The study determined the average noticeable and disruptive values for each factor (Figure 8), which helped answer **RQ2**. Latency noticeably impacted user experience at 373.27ms and was disruptive at 654.90ms. Registration accuracy had noticeable values of 61.52cm (camera) and 62.35cm (player), with disruptive values of 105.78cm and 101.86cm. Jitter values were comparable for camera and dynamic content, with noticeable values at 4.62s and 4.50s, and disruptive values at 3.00s and 2.81s.

The study found that using only the noticeable values from the client device in the comparison condition still had a strong negative impact on user experience. User experience disruption appeared to be non-linear, as UEQ ratings dropped sharply, indicating a poor user experience which answers **RQ3**. This suggests that while it is important to separate factors in research, they alone are not sufficient for ensuring a successful user experience. The compounding effect of these factors increases participants' frustration, leading to lower ratings. One participant (P19) noted that the individual factors did not impact their experience much, but collectively, they had a significant effect.

A limitation of the study is that we used different clip snippets for manipulation and comparison, introducing confounding variables such as differing amounts of horizontal movement in the clip. The comparison clip featured more horizontal movement, potentially amplifying the effects of latency and registration accuracy. Unfortunately, due to COVID-19 restrictions, access to a full stadium for capturing additional video data was limited. Consequently, the dynamic arrow visualization was placed on the referee rather than a player, which is not typical in real games. The study was conducted on a mobile phone to closely simulate the intended use case without considering stereoscopic videos, which we think might not have a significant impact due to the distance of the content from users.

Distance of action from the field could be another confounding variable, as closer action had more pixels than action further away. The action in the video clip starts relatively close to the camera and then moves toward the opposite side of the field. For the same length of distance occurring 68m away (width of the field), the one that is nearer to the spectator is 3.7 times more pixels than the further counterpart. The calculation is done by physically measuring the length of the screen with a ruler and then converting it into pixels with the pixel-per-inch of the iPhoneXR. Some participants may have based their evaluations on the closer action, evident in their disruptive values (P5, P15, P16, P17). This was seen in cases where participants set disruptive values smaller than noticeable values. We observed that the participant did it very quickly for the disruptive value, strengthening the earlier mentioned point.

At least six participants complained or showed signs of arm fatigue, including one participant (P2) that mentioned the arm fatigued happened a little over a minute. This prompts a reconsideration of mobile AR ergonomics for such studies. Additionally, three participants (P4, P5 and P9) had difficulty manipulating the slider, indicating a usability issue that could potentially be resolved by increasing the slider handle size. While most UEQ scales were pragmatic, the hedonic scales of *Stimulation* and *Novelty* were still indicative of user experience, as participants aimed to gather more information from the game despite not being given specific tasks.

Conclusion

This paper serves as an initial step in researching how to enhance user experiences impacted by technical factors and limitations. Our study primarily examines specific aspects of three technical factors — mainly *tracking delay* and *networking delay* in latency, *static errors* in registration accuracy, and *spatial jitter* for jitter. Since technical factors affecting the system are unavoidable, our objective is to minimize them to an acceptable level. These factors should be considered during the design phase of any AR application.

Our findings indicate that even when compounded together, noticeable technical factors can significantly impact user experience. To ensure a coherent user experience, it is essential to keep both content latency and registration accuracy low. While jitter is still disruptive, it is less detrimental to the user experience than latency and registration accuracy, provided it occurs only occasionally. However, the results and values presented in this study are specific to large-scale AR environments. There are also many other considerations besides technical factors regarding XR user experience, as mentioned by [18].

Though not explored further in this study, future research could potentially develop a formula to calculate acceptable values for these factors based on distance. Additionally, conducting a similar study using stereo 360° videos may provide valuable insights into the relationship between these factors and depth perception. We believe this research will address the gap identified by other researchers concerning the lack of user evaluation in AR publications [15], [10], particularly when considering technical factors in relation to user experiences.

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 participants and our HCI and Visual Computing labs members.

REFERENCES

1. Ronald Azuma and Gary Bishop. Improving static and dynamic registration in an optical see-through HMD. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pages 197–204, 1994.
2. Tristan Braud, Farshid Hassani Bijarbooneh, Dimitris Chatzopoulos, and Pan Hui. Future networking challenges: The case of mobile augmented reality. In *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, pages 1796–1807. IEEE, 2017.
3. Frederick P Brooks. What's real about virtual reality? *IEEE Computer graphics and applications*, 19(6):16–27, 1999.
4. Arindam Dey, Mark Billinghurst, Robert W Lindeman, and J Swan. A systematic review of 10 years of augmented reality usability studies: 2005 to 2014. *Frontiers in Robotics and AI*, 5:37, 2018.
5. Qing Hong Gao, Tao Ruan Wan, Wen Tang, and Long Chen. A stable and accurate marker-less augmented

reality registration method. In *2017 International Conference on Cyberworlds (CW)*, pages 41–47. IEEE, 2017.

6. Niels Henze and Susanne Boll. Evaluation of an off-screen visualization for magic lens and dynamic peephole interfaces. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, pages 191–194, 2010.
7. Richard L Holloway. Registration error analysis for augmented reality. *Presence: Teleoperators & Virtual Environments*, 6(4):413–432, 1997.
8. Takahiro Iwata, Tetsuo Yamabe, and Tatsuo Nakajima. Augmented reality go: extending traditional game play with interactive self-learning support. In *2011 IEEE 17th International Conference on Embedded and Real-Time Computing Systems and Applications*, volume 1, pages 105–114. IEEE, 2011.
9. Jason Jerald. Latency compensation for head-mounted virtual reality. *Department of Computer Science, University of North Carolina at Chapel Hill*, 2004.
10. Kangsoo Kim, Mark Billinghurst, Gerd Bruder, Henry Been-Lirn Duh, and Gregory F Welch. Revisiting trends in augmented reality research: A review of the 2nd decade of ISMAR (2008–2017). *IEEE transactions on visualization and computer graphics*, 24(11):2947–2962, 2018.
11. Ernst Kruijff, J Edward Swan, and Steven Feiner. Perceptual issues in augmented reality revisited. In *2010 IEEE International Symposium on Mixed and Augmented Reality*, pages 3–12. IEEE, 2010.
12. Ruyu Liu, Jianhua Zhang, Kejie Yin, Jiaxin Wu, Ruihao Lin, and Shengyong Chen. Instant SLAM initialization for outdoor omnidirectional augmented reality. In *Proceedings of the 31st International Conference on Computer Animation and Social Agents*, pages 66–70, 2018.
13. Wei Hong Lo, Stefanie Zollmann, and Holger Regenbrecht. Stats on-site—Sports spectator experience through situated visualizations. *Computers & Graphics*, 2021.
14. Thibault Louis, Jocelyne Troccaz, Amélie Rochet-Capellan, and François Bérard. Is it real? measuring the effect of resolution, latency, frame rate and jitter on the presence of virtual entities. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*, pages 5–16, 2019.
15. Thomas Olsson and Markus Salo. Online user survey on current mobile augmented reality applications. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality*, pages 75–84. IEEE, 2011.
16. Patrick Renner and Thies Pfeiffer. [POSTER] Augmented Reality Assistance in the Central Field-of-View Outperforms Peripheral Displays for Order Picking: Results from a Virtual Reality Simulation Study. In *2017 IEEE international symposium on mixed and augmented reality (ISMAR-Adjunct)*, pages 176–181. IEEE, 2017.
17. Panagiotis D Ritsos, Dennis P Ritsos, and Antonis S Gougoulis. Standards for augmented reality: A user experience perspective. In *International AR standards meeting*, pages 1–9, 2011.
18. Camille Sagnier, Emilie Loup-Escande, Domitile Lourdeaux, Indira Thouvenin, and Gérard Valléry. User acceptance of virtual reality: an extended technology acceptance model. *International Journal of Human–Computer Interaction*, 36(11):993–1007, 2020.
19. Sean White and Steven K. Feiner. SiteLens : Situated Visualization Techniques for Urban Site Visits. In *Proceedings of the 27th international conference on Human Factors in computing systems (CHI2009)*, pages 1117–1120. ACM, 2009.
20. Jason Wither, Yun-Ta Tsai, and Ronald Azuma. Indirect augmented reality. *Computers & Graphics*, 35(4):810–822, 2011.

Wei Hong Lo is a Research Assistant at the University of Otago in New Zealand. His main research interest is in human-computer interaction, XR user experience, visualizations, and applications. Wei Hong received his Ph.D. degree in XR User Experience and Visualization from the University of Otago, New Zealand. Contact him at weihong.lo@otago.ac.nz

Holger Regenbrecht is a professor at the University of Otago in New Zealand. His research interests include virtual and augmented reality, human–computer interaction, applied computer science and information technology, 3D teleconferencing, psychological aspects of mixed reality, three-dimensional user interfaces, and computer-aided therapy and rehabilitation. He received the Dr. Ing. (Ph.D.) degree in applied computer science from Bauhaus University Weimar, Germany, in 2000. He is a Member of the IEEE and ACM.

Stefanie Zollmann is an Associate Professor at the University of Otago in New Zealand. Her main research interests are XR for sports and media, visualization techniques for augmented reality, but also include capturing for XR and immersive experiences. She received a PhD degree (Dr. techn) from Graz University of Technology, Austria in 2013.