Towards Accessible Sports Broadcasts for Blind and Low-Vision Viewers

Gaurav Jain
Columbia University
New York, NY, USA

Basel Hindi
Columbia University
New York, NY, USA

Connor Courtien*
Hunter College
New York, NY, USA

Xin Yi Therese Xu*
Pomona College
Claremont, CA, USA

Conrad Wyrick*
University of Florida
Gainesville, FL, USA

Michael Malcolm*
SUNY at Albany
Albany, NY, USA

Brian A. Smith
Columbia University
New York, NY, USA

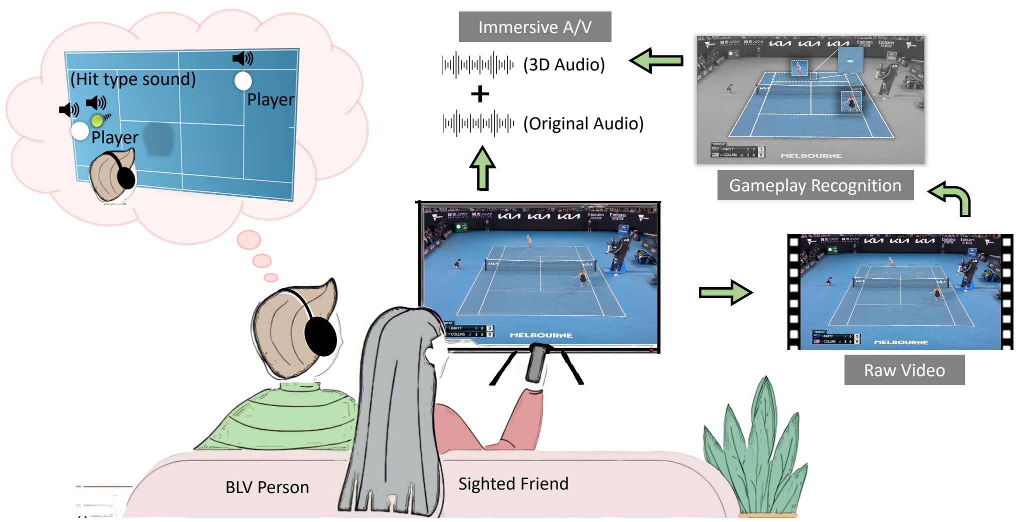

Figure 1: Illustration of Immersive A/V. Immersive A/V is a technique for making sports broadcasts accessible and immersive to blind and low-vision (BLV) viewers by automatically extracting gameplay information and conveying it through an added layer of spatialized audio cues. Immersive A/V conveys players' positions and actions (i.e., hit types) as detected by computer vision-based gameplay recognition, enabling BLV viewers to visualize the action. Using Immersive A/V, BLV people can co-watch sports with their sighted friends via this single, universal format.

ABSTRACT

Blind and low-vision (BLV) people watch sports through radio broadcasts that offer a play-by-play description of the game. However, recent trends show a decline in the availability and quality of radio broadcasts due to the rise of video streaming platforms on the internet and the cost of hiring professional announcers. As a result, sports broadcasts have now become even more inaccessible to BLV people. In this work, we present Immersive A/V, a technique for making sports broadcasts —in our case, tennis broadcasts— accessible and immersive to BLV viewers by automatically extracting gameplay information and conveying it through an added layer of spatialized audio cues. Immersive A/V conveys players' positions and actions as detected by computer vision-based video analysis, allowing BLV viewers to visualize the action. We designed Immersive A/V based on results from a formative study with BLV participants. We conclude by outlining our plans for evaluating Immersive A/V and the future implications of this research.

*This work was done while Connor Courtien, Xin Yi Therese Xu, Conrad Wyrick, and Michael Malcolm were interns at Columbia University.

CHI EA '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9422-2/23/04.
https://doi.org/10.1145/3544549.3585610

CCS CONCEPTS

• **Human-centered computing → Accessibility systems and tools**.

KEYWORDS

Visual impairments, sports, accessibility, computer vision

ACM Reference Format:

Gaurav Jain, Basel Hindi, Connor Courtien, Xin Yi Therese Xu, Conrad Wyrick, Michael Malcolm, and Brian A. Smith. 2023. Towards Accessible Sports Broadcasts for Blind and Low-Vision Viewers. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3544549.3585610

1 INTRODUCTION

Sports influence people's social and cultural lives by fostering a sense of community and shared identity [7, 23, 38]. For blind and low-vision (BLV) people, however, watching sports is an exclusionary and isolating experience due to the lack of accessible sports broadcasts. Despite being one of the most-watched content categories on television for BLV viewers, sports broadcasts remain among the most challenging to access [35]. TV commentators assume that their audience is able to visually follow the action on the screen, which leads them to provide commentary that focuses more on providing additional context and analysis of the game rather than describing the action itself.

As a result, BLV people watch sports through radio broadcasts because radio announcers convey information non-visually and, thus, tend to be more descriptive than TV commentators [29]. While radio offers a play-by-play description of the game to BLV viewers, it comes at the cost of not being able to enjoy sports with their friends and family — who prefer to watch sports through video-based broadcasts such as TV [3].

Past research has observed a decline in the availability and quality of radio broadcasts over the past few decades [33, 42]. This is primarily due to the cost of hiring professional announcers [33] and the rise of video streaming platforms such as YouTube, Facebook Live, Hulu, and ESPN, that have replaced the traditional streaming platforms [42]. The ease of access to video streaming platforms has made even more sports content available —including college sports and local matches— that would not have been aired on TV before. Ironically, this increased availability of sports broadcasts has led to an even lower accessibility of sports content for BLV viewers.

Prior work has made efforts toward making football [1, 13] and tennis [8, 21, 33, 41] games accessible to BLV viewers. However, most approaches require the field to be instrumented with specialized hardware to infer gameplay, which is not always feasible. As a result, only a fraction of all sports broadcasts can be made accessible through these approaches. The lack of low-cost means to infer gameplay data is a challenge to sports broadcast accessibility.

One promising direction is to infer gameplay by directly parsing the broadcast video feed. Although research on video analysis has explored techniques to track game elements, such as actions [46], balls [18], and players [31], much of the focus has been on contexts beyond accessibility, failing to consider how computer vision systems can assist users in understanding gameplay. Our goal, by contrast, is to understand how computer vision can help *BLV people* better perceive and visualize sports gameplay.

In this work, we present *Immersive A/V*, a technique for making sports broadcasts —in our case, tennis broadcasts— accessible and immersive to blind and low-vision (BLV) viewers by automatically extracting gameplay information from source broadcast and conveying it through an added layer of spatialized audio cues. Immersive A/V conveys players' positions and actions as detected by computer vision-based video analysis, enabling BLV viewers to visualize the action. Figure 1 illustrates Immersive A/V conveying action in a tennis match on a plane orthogonal to the viewer in the 3D soundscape, making it possible for them to enjoy sports with their sighted peers. We designed Immersive A/V based on a formative study with BLV participants. We conclude by outlining our plans for evaluating Immersive A/V and the future implications of this research.

2 RELATED WORK

Our work builds from three main threads of research: *(i)* approaches to visual media accessibility, *(ii)* sports media accessibility for BLVs, and *(iii)* sports video analysis using computer vision.

Approaches to visual media accessibility. One common approach for making visual media accessible to BLV people is through *text-based descriptions* (e.g., alt-text for images [28, 50], audio descriptions for videos [2, 49]). Prior work, however, shows that descriptions do not provide BLV people a rich spatial understanding of the visual content [30, 36, 39], which is important for understanding sports gameplay.

Another approach to visual media accessibility is using tactile graphics, which conveys rich spatial information via touch [4, 45]. Tactile graphics have been successfully used to understand spatial layout of paintings [40], floor plans [14, 27], and more [9, 15, 22, 25, 37, 44]. However, BLV users explore the tactile surface by touch, which makes it less suitable for perceiving dynamic visual content (e.g., videos). This becomes even more difficult for sports videos, given the fast-paced nature of sports.

Audio, in the form of sonification or audio-cues, has also been explored for general image accessibility [16, 34], as well as for particular forms of images such as time series charts [17, 43]. However, limited work has been done to make videos, specifically sports videos, accessible via audio. In this work, we use 3D spatialized audio to make tennis videos accessible to BLV people, with an aim to help them visualize the gameplay.

Sports broadcast accessibility for BLV people. Past research has explored different ways of making sports broadcasts accessible to BLV people, leveraging tactile graphic displays for football games [8, 33, 41] and 3D-audio for tennis games [1, 13]. However, most existing approaches rely on specialized hardware installations to collect game data, which may not always be feasible. For example, Action Audio [1] relies on data acquired by ball tracking systems [20], where 3D-ball position is tracked using multiple high-performance cameras mounted throughout the tennis court. Our work explores inferring gameplay directly from broadcast videos using computer vision so that all tennis broadcasts can be made equally accessible to BLV people.

Sports video analysis using computer vision. Research within the computer vision community has explored techniques to analyze sports videos. However, much of the focus is on downstream applications where user experience is not considered. Example of applications include recognizing actions [46], tracking ball [18] and

Table 1: Self-reported demographics of our formative study participants. Gender information was collected as a free response where our participants identified themselves as female (F), non-binary (NB), and male (M). The country codes refer to India (IN) and the United States (US). Participants indicated their sports fandom as per Hunt et al.'s [19] scale which classifies sports fans into five categories: (1) temporary, (2) local, (3) devoted, (4) fanatical, and (5) dysfunctional.

PID	Gender	Age	Race	Country	Occupation	Vision ability	Onset	Sports Fandom (1–5)	Tennis Familiarity (1–5)
P1	M	37	White	US	Game developer	Totally blind	Birth	4: Fanatical fan	5: Extremely familiar
P2	F	60	White	US	Retired	Totally blind	Age 25	2: Local fan	2: Slightly familiar
P3	F	28	White	US	FMLA claims expert	Totally blind	Birth	1: Temporary fan	2: Slightly familiar
P4	M	32	Asian	IN	Self-employed	Low vision	Age 20	1: Temporary fan	1: Not at all familiar
P5	M	23	Black	US	Editor	Low vision	Age 15	4: Fanatical fan	2: Slightly familiar

players [31], automatic refereeing and analysis [11, 12, 48], and augmenting broadcasts [6]. Our work builds on sports video analysis research to develop Immersive A/V, which is designed to enhance BLV people's experience of watching sports. Next, we present our formative study that informed our design and development of Immersive A/V.

3 FORMATIVE STUDY

To inform the design of Immersive A/V, we conducted semi-structured interviews and observation sessions with five BLV participants. Specifically, we focus on answering two questions:

Q1. What challenges do BLV people face when watching sports?
Q2. What are BLV viewers' information preferences for achieving a better understanding of the gameplay?

3.1 Methods

Participants. We recruited five BLV participants (three males, two females; aged 23–60) by posting to social media platforms. Table 1 summarises the participants' information. All interviews were conducted remotely via Zoom and lasted for about 60–75 minutes. Participants were compensated $25 for this IRB-approved study. Additionally, we draw on the experiences of our two BLV co-authors — both of whom have low vision.

Procedure. To answer the first question about BLV people's challenges when watching sports, we used a recent Critical Incident Technique (CIT) [10], in which we asked participants to recall and describe a recent time when they watched sports. We asked participants to describe their likes and dislikes about this experience, the challenges they faced while viewing the game, and the ways in which they navigated those challenges.

To answer the second question about BLV people's information preferences, we observed participants as they viewed tennis games via television (TV) and radio broadcasts. We shared our screen over Zoom and played several short clips from professional tennis matches for both TV and radio. After each clip, we asked participants to describe the gameplay and elaborate on aspects of gameplay they wanted to learn more about.

Interview Analysis. We transcribed the interviews and analyzed them via a grounded theory approach [5]. We used NVivo [32] to first create an initial set of codes, which we then iteratively refined to identify emerging themes for both questions.

3.2 Challenges Faced by BLV People when Watching Sports

We found two major challenges that BLV people face when watching sports.

3.2.1 Feeling excluded when co-watching sports with friends and family. Our participants noted that it is challenging for them to co-watch sports with friends and family because of mismatched preferences for the mediums through which they watch sports. BLV people prefer radio commentary, whereas their sighted friends and family prefer a visual medium such as TV. P2 mentioned that not being able to watch sports through a medium they could equally understand made them feel excluded: "*Well, I feel like I was kind of left out with the family conversation.*" P1 explained that feelings of exclusion are even more pronounced for sports because: "*you don't like having what is supposed to be fun, make you feel excluded*".

3.2.2 Inappropriate amount of information. We observed that participants felt underwhelmed when watching tennis on TV and overwhelmed when watching tennis on radio. Participants noted that, unlike other sports, TV commentators in tennis are silent during the play. As a result, participants lose interest: "*I feel like there's a lot of stuff that I'm just not getting in, so I don't feel very immersed in it. And so my mind wanders*" (P1). On the contrary, radio announcers spoke too fast for them to be able to follow the game events, which made them feel frustrated sometimes.

3.3 Information Preferences for Better Gameplay Understanding

We discovered BLV viewers' two major information preferences for better understanding gameplay.

3.3.1 Preference for spatial information about the gameplay. After listening to tennis clips for both TV and radio, participants expressed desire to more closely follow *where* the actions were happening on court: "*I never got a sense of where they were hitting it on the court. Because I know when you're really playing, if the player is up close to the net, then you try to hit it back in the far corner, you know, to make them have to run to make the play. I didn't get a sense whether that was happening or not.*" (P2).

3.3.2 Preference for neutral, objective information about the gameplay. Participants expressed their preference for fact-based reporting of information versus information interpreted from someone else's perspective. For instance, "*the announcers [often] color things from their home team's perspective*" (P4), and if a BLV viewer supports the other team, they "*probably wouldn't want [announcers'] opinions as much because I could form my own opinions*" (P4).

3.4 Design Goals

Based on our formative study findings, we present our design goals for Immersive A/V as follows:

G1: Facilitating spatial understanding of the gameplay. As noted, perceiving spatial aspects of gameplay is difficult in a non-visual format (Section 3.3.1), one of our goals is to intuitively facilitate a spatial understanding of the gameplay for BLV people.

G2: Providing an appropriate amount of information to facilitate immersion. Since immersion within the game is important to BLV viewers (Section 3.2.2), one of our aims is to ensure that an enhanced gameplay understanding is not achieved at the cost of immersion.

G3: Providing a single format that both BLV and sighted viewers can enjoy. To instill a sense of affiliation in their sports watching experience (Section 3.2.1), one of our goals is to provide a single, universal format that BLV people can co-watch with friends and family.

G4: Supporting agency in gameplay understanding. As mentioned in Section 3.3.2, it is important for BLV people to form their own opinion about the gameplay. Thus, one of our aims is to provide factual information that enables BLV people to view the game from their own perspective.

4 IMMERSIVE A/V

In this section, we present the Immersive A/V pipeline, which is a technique for making sports broadcasts —in our case, tennis broadcasts— accessible and immersive to blind and low-vision (BLV) viewers by automatically extracting gameplay information from source broadcast and conveying it through an added layer of spatialized audio cues. Immersive A/V conveys players' positions and actions as detected by computer vision-based video analysis, enabling BLV viewers to visualize the action. Immersive A/V is designed to address BLV people's challenges and information preferences for watching sports, which we learned through our formative study.

Figure 2 shows the Immersive A/V pipeline, which generates spatialized audio cues from a raw tennis video (i.e., video from a television broadcast) by processing it through three major components: *(i)* rally segmentation, *(ii)* gameplay recognition, and *(iii)* audio feedback generation.

4.1 Rally Segmentation.

A tennis broadcast video consists of a sequence of rallies, with non-play periods between them (e.g., commercial breaks and players changing sides). In order to effectively overlay spatialized audio cues onto the original video —which is important to create a coherent A/V format (**G3**)— the raw video is segmented to get a sequence of rallies. A *rally* in tennis is analogous to what one might call a play or a point in other sports: it is an exchange of shots between players, ending when one player fails to make a successful return.

To segment rallies, we used the observation that during a rally, the camera in tennis broadcasts is steadily positioned behind one of the players, overlooking the full court. We used a support vector classifier (SVC) to label the HOG features extracted from each frame as *"play"* or *"non-play"*. Each frame of the raw video is passed through this classifier to obtain a list of labels, onto which we applied Kalman filter to get continuous video segments that will be passed through the gameplay recognition module.

4.2 Gameplay Recognition.

The gameplay recognition component extracts gameplay information from the segmented rallies in a format that matches BLV people's information preferences for watching sports (**G1, G4**). Specifically, we extract spatial information (**G1**) about players' positions, as well as factual information such as hits and their types (forehand vs. backhand) to support BLV people in forming their own understanding of the gameplay (**G4**). The gameplay recognition component constitutes the following three modules:

Frame-level Game Data Extraction. Figure 2a shows this module, which processes each video segment, to extract locations of court lines, ball, and the two players.

For locating court lines, we use the fact that court lines are always white. Thus, we filtered out all white pixels and applied Hough transform to identify lines. We then compare this set of lines with a reference court structure to find the court lines. However, this approach generated mis-detections due to non-court white pixels in the frame (for example, audience members wearing white clothes and advertisement boards). To address this, we annotated our own data and trained a custom court segmentation model to mask out everything but the court. We used FCN [26] with a Resnet50 backbone for this purpose. This considerably improved the robustness in detecting court lines.

To extract the ball and player position, we experimented with a variety of deep learning network architectures and chose one that performed best. For ball tracking, we used TrackNet [18], which is state-of-the-art in detecting small, fast-moving objects such as tennis balls. To extract player position, we first experimented with the pre-trained YOLOv5 [24] model's 'person' class. However, due to the presence of ball boys/girls and line judges on the court, we had to train our own custom model to accurately detect only the two players. To train this model, we annotated over 1500 frames by labeling the two players as "top" and "bottom". As shown in Figure 2a (players), the player farther away from the camera is labeled as "top". The extracted game data is then fed into the top-view projection module (Figure 2b).

Top-view Projection. Figure 2b shows the top-view projection. The extracted game data from the previous module is in pixel coordinates, i.e., (x, y) coordinate with respect to the image's 2D plane. Since the actual game is in 3D, directly inferring gameplay in terms of pixel coordinates would lead to inconsistencies. Thus, we project the pixel coordinates onto a 2D plane using the correspondence between an actual court —whose size is fixed and known— and the court lines that we detected. We apply the same projection to transform the player and ball coordinates as well.

We repeat the game data extraction and top view projection for all frames in the segmented rally to obtain a sequence of coordinates for the players and the ball, which we feed into the hit detection module (Figure 2c).

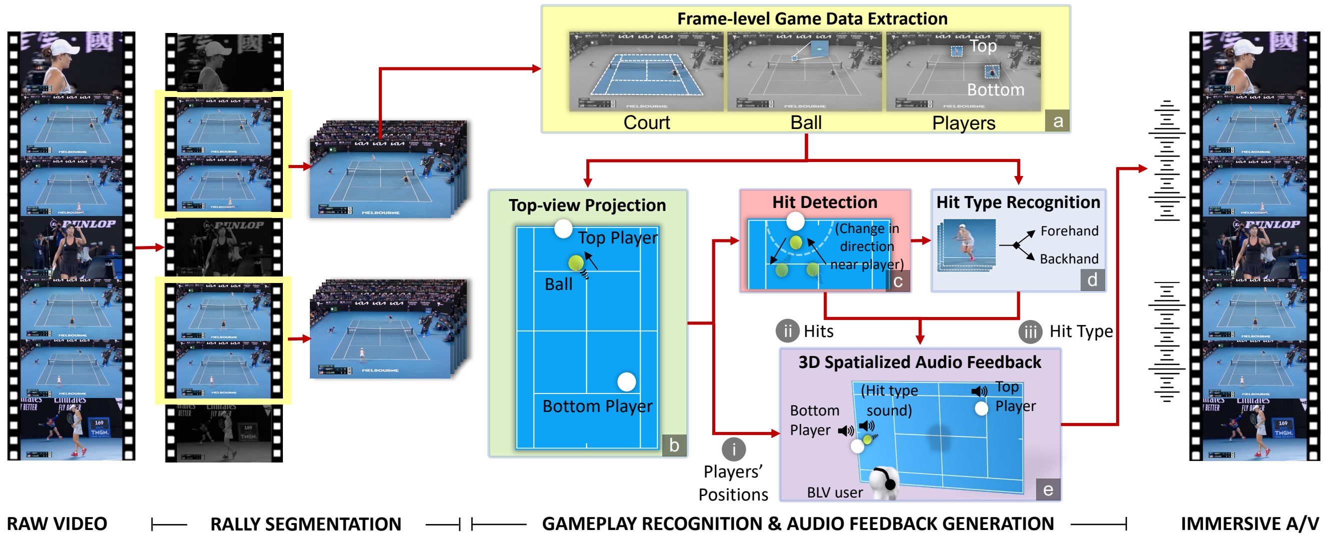

Figure 2: Overview of the Immersive A/V pipeline. The pipeline generates spatialized audio directly from the raw video of sports broadcasts on TV. The raw video is segmented to get sequences where points are in play (i.e., rallies), removing periods between points such as commercial breaks and players changing sides. We then perform the following steps. (a–b) For each rally, we recognize gameplay by extracting information about the court, players, and the ball. (c–d) Using this information, we detect *when* a player hits the ball, and *what* type of hit they used to do so. (e) Finally, 3D spatialized audio feedback is generated to provide BLV people spatial information about (i) players' positions, (ii) the hits, and (iii) hit types. This audio is overlaid onto the original audio from the raw video.

Hit Detection. Figure 2c shows the hit detection module. In this module, we detect specific frames at which a player hits the ball. We do so by using the fact that at the time of a hit, the ball changes direction perpendicular to the net. Thus, we label all such frames as "hit frames", where the ball coordinates depict a change in this direction. However, we noticed that inconsistencies in ball tracking result in abrupt movement of the ball, leading to spurious "hit frames" being detected. For example, irregularities in ball coordinates due to occlusion by either player, often lead to mis-detected hits. Another such instance is when the ball bounces, which results in minor changes in direction sometimes.

To address this, we only consider a change in direction as a hit when it is within a specific euclidean distance from one of the players as illustrated in Figure 2c. We fine-tuned this distance parameter based on the camera angle, by trial and error. We also provide the location of hits as the position of the player who hit the ball, in order to better support the spatial understanding of gameplay (**G1**).

Hit Type Recognition. Figure 2d shows the hit type recognition module. In this module, we recognize the type of hit as a forehand or a backhand, given a "hit-frame." We trained a custom hit type recognition model which predicts the hit type based on a sequence of 9 player crops, where the central frame corresponds to the detected "hit frame". This is done to capture the action of hitting a ball. We obtain these player crops from the game data we extracted in the previous modules. We aggregate the predictions by averaging the prediction probabilities across each of the 9 player crops to obtain the final hit type prediction.

4.3 Audio Feedback Generation.

This component transforms the gameplay recognition data into 3D spatialized audio cues, which is then overlayed onto the original video. To design the audio feedback, we draw on prior work [1, 13] that uses 3D spatialized audio cues for making tennis accessible and several design iterations and pilot studies that we conducted with our two BLV co-authors. Throughout the design process, we focused our discussions around our design goals (Section 3.4). Specifically, **G1**: facilitating spatial understanding of the gameplay, **G2**: providing appropriate amount of information to facilitate immersion, and **G4**: supporting agency in gameplay understanding.

Figure 2e illustrates the audio feedback generation, which relies on three pieces of information: *(i)* players' positions, *(ii)* hits, and *(iii)* hit types. We render 3D audio on a plane orthogonal to the viewer in the 3D soundscape using Unity's Steam Audio plug-in [47].

Conveying Players' Positions. Players' positions are conveyed as continuous, point audio sources that move around on the plane. Our design iterations revealed that even though the players were rendered in separate ears, it was difficult to uniquely identify their movements. To address this, we altered the pitch for the two players, iteratively tweaking the difference between the two sounds until it was easy to distinguish the two players' movements without losing immersion (**G2**). We also experimented with different audio configurations for rendering the players' positions. One notable instance was to render the players as discontinuous, point audio sources that only activates at three points: the two ends and the middle of the court. We tested this configuration with our two BLV co-authors and found that although discontinuous sounds were less cognitively demanding, they were less immersive than continuous sounds (**G2**).

Conveying Hit and Hit types. In addition to players' positions, audio cues are intermittently played whenever the ball is hit. The audio cues originate from the location of the hit and distinguish forehands from backhands. We differentiate a forehand from a backhand using the pitch. We chose the bell sound for hits, which resembles the sound of the ball used for playing blind tennis. Prior research inspired this design choice [1, 13]. Spatially rendering the hit sounds enabled our two BLV co-authors to not only gain a spatial understanding of the hits (**G1**) but also make their inferences about the game (**G4**). Prior work [13] uses a grid overlaid onto the court to specify where the ball bounces on the court. However, our design iterations revealed value in spatially rendering hit sounds, so that BLV people can get a feel for themselves where the ball is being hit on the court.

Overlaying Original Audio Onto 3D Spatialized Audio. Finally, we overlaid these audio cues onto the original audio from the raw video. Two main reasons drove this design choice. First, adding natural sounds of the game makes it feel more realistic and immersive (**G2**), as pointed out by our two BLV co-authors. Second, integrating the original audio with an accessible format would enable BLV people to co-watch tennis games with their friends and family — one of the challenges they faced, which we learned from our formative study (**G3**).

5 FUTURE WORK

We are conducting user studies with BLV participants with three evaluation goals. First, we evaluate how Immersive A/V affects BLV viewers' ability to understand tennis gameplay compared to their existing means of viewing tennis: Television and Radio broadcasts. Second, we quantitatively analyze BLV people's overall experience viewing tennis games using Immersive A/V and these existing means. Third, we see how BLV participants rank the three audio formats in order of their preference for watching tennis games.

Our initial results indicate a strong preference for Immersive A/V as it provides BLV viewers a more accurate understanding of the gameplay. We also found that Immersive A/V facilitates more immersion, with many participants valuing its ability to offer them agency in interpreting the gameplay for themselves and forming their own opinions about the players' strategies during the game.

Besides a user study with BLV participants, we are also evaluating the proposed pipeline's overall technical performance. Through this evaluation, we aim to present the operating characteristics of Immersive A/V and to understand how system errors affect viewers' gameplay understanding. Overall, the results show that the pipeline detects players with over 95% accuracy and detects hits and their types with 80% accuracy. We have also found that BLV participants could compensate for the incorrect or missing information by themselves through context.

In the future, we aim to demonstrate potential applications of the proposed framework and discuss how our design of Immersive A/V can be extended to make *all* online videos accessible. We envision integrating Immersive A/V as a plug-in for video streaming platforms, similar to how captions can be automatically generated on platforms like YouTube.

ACKNOWLEDGMENTS

We thank Chloe Ho, Hazel Zhu, Jacqueline Gibson, Jizhong Wang, Siwanta Thapa, and Venkat Ramamoorthi for data annotation, and all our study participants.

REFERENCES

[1] Action Audio. 2021. Making Sports Broadcasts Accessible to People Living With Blindness or Low Vision. https://action-audio.com/

[2] American Council of the Blind. 2022. The Audio Description Project. https://adp.acb.org/guidelines.html

[3] Saki Asakawa and Amy Hurst. 2021. "What just happened?": Understanding Nonvisual Watching Sports Experiences. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, Virtual Event USA, 1–3. https://doi.org/10.1145/3441852.3476525

[4] Matthew Butler, Leona M Holloway, Samuel Reinders, Cagatay Goncu, and Kim Marriott. 2021. Technology Developments in Touch-Based Accessible Graphics: A Systematic Review of Research 2010-2020. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445207

[5] Kathy Charmaz. 2006. *Constructing grounded theory*. Sage Publications, London ; Thousand Oaks, Calif.

[6] Zhutian Chen, Shuainan Ye, Xiangtong Chu, Haijun Xia, Hui Zhang, Huamin Qu, and Yingcai Wu. 2022. Augmenting Sports Videos with VisCommentator. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (Jan. 2022), 824–834. https://doi.org/10.1109/TVCG.2021.3114806

[7] Morgan Cottril. February 12, 2020. The Importance of Sports in Culture. https://fghsnews.com/2603/diversity/the-importance-of-sports-in-culture/

[8] Expedio Design. 2019. Footbraile. https://www.expediodesign.com/portfolio-footbraille

[9] Olutayo Falase, Alexa F. Siu, and Sean Follmer. 2019. Tactile Code Skimmer: A Tool to Help Blind Programmers Feel the Structure of Code. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, Pittsburgh PA USA, 536–538. https://doi.org/10.1145/3308561.3354616

[10] John C. Flanagan. 1954. The critical incident technique. *Psychological Bulletin* 51, 4 (1954), 327–358.

[11] Anurag Ghosh and C. V. Jawahar. 2018. SmartTennisTV: Automatic indexing of tennis videos. *arXiv:1801.01430 [cs]* (Jan. 2018). http://arxiv.org/abs/1801.01430 arXiv: 1801.01430.

[12] Anurag Ghosh, Suriya Singh, and C. V. Jawahar. 2017. Towards Structured Analysis of Broadcast Badminton Videos. *arXiv:1712.08714 [cs]* (Dec. 2017). http://arxiv.org/abs/1712.08714 arXiv: 1712.08714.

[13] 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*. Vol. 12932. Springer International Publishing, Cham, 396–417. https://doi.org/10.1007/978-3-030-85623-6_24

[14] Cagatay Goncu, Anuradha Madugalla, Simone Marinai, and Kim Marriott. 2015. Accessible On-Line Floor Plans. In *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, Florence Italy, 388–398. https://doi.org/10.1145/2736277.2741660

[15] Cagatay Goncu and Kim Marriott. 2011. GraVVITAS: Generic Multi-touch Presentation of Accessible Graphics. In *Human-Computer Interaction – INTERACT 2011*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Pedro Campos, Nicholas Graham, Joaquim Jorge, Nuno Nunes, Philippe Palanque, and Marco Winckler (Eds.). Vol. 6946. Springer Berlin Heidelberg, Berlin, Heidelberg, 30–48. https://doi.org/10.1007/978-3-642-23774-4_5

[16] Giles Hamilton-Fletcher, Marianna Obrist, Phil Watten, Michele Mengucci, and Jamie Ward. 2016. "I Always Wanted to See the Night Sky": Blind User Preferences for Sensory Substitution Devices. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 2162–2174. https://doi.org/10.1145/2858036.2858241

[17] Leona M Holloway, Cagatay Goncu, Alon Ilsar, Matthew Butler, and Kim Marriott. 2022. Infosonics: Accessible Infographics for People who are Blind using Sonification and Voice. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–13. https://doi.org/10.1145/3491102.3517465

[18] Yu-Chuan Huang, I-No Liao, Ching-Hsuan Chen, Tsì-Uí İk, and Wen-Chih Peng. 2019. TrackNet: A Deep Learning Network for Tracking High-speed and Tiny Objects in Sports Applications*. In *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. 1–8. https://doi.org/10.1109/AVSS.2019.8909871 ISSN: 2643-6213.

[19] Kenneth A. Hunt, Terry Bristol, and R. Edward Bashaw. 1999. A conceptual approach to classifying sports fans. *The Journal of Services Marketing* 13, 6 (1999), 439–452. https://doi.org/10.1108/08876049910298720

[20] Hawk-Eye Innovations. 2001. https://www.hawkeyeinnovations.com/

[21] IrisVision. Retrieved July 15, 2022. IrisVision. https://irisvision.com/product/

[22] Hiroo Iwata, Hiroaki Yano, Fumitaka Nakaizumi, and Ryo Kawamura. 2001. Project FEELEX: Adding Haptic Surface to Graphics. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '01*. ACM Press, Not Known, 469–476. https://doi.org/10.1145/383259.383314

[23] Grant Jarvie, James Thornton, and Hector Mackie. 2017. *Sport, Culture and Society: An Introduction* (3 ed.). Routledge, Third edition. | Abingdon, Oxon ; New York, NY : Routledge is an imprint of the Taylor & Francis Group, an Informa Business, [2017].

[24] Glenn Jocher et al. April 2021. YOLOv5. https://ultralytics.com/yolov5

[25] Shaun K. Kane, Meredith Ringel Morris, and Jacob O. Wobbrock. 2013. Touchplates: Low-Cost Tactile Overlays for Visually Impaired Touch Screen Users. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, Bellevue Washington, 1–8. https://doi.org/10.1145/2513383.2513442

[26] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully Convolutional Networks for Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), 3431–3440. https://openaccess.thecvf.com/content_cvpr_2015/html/Long_Fully_Convolutional_Networks_2015_CVPR_paper.html

[27] Anuradha Madugalla, Kim Marriott, Simone Marinai, Samuele Capobianco, and Cagatay Goncu. 2020. Creating Accessible Online Floor Plans for Visually Impaired Readers. *ACM Transactions on Accessible Computing* 13, 4 (Oct. 2020), 1–37. https://doi.org/10.1145/3410446

[28] Tom McEwan and Ben Weerts. 2007. ALT Text and Basic Accessibility. https://doi.org/10.14236/ewic/HCI2007.64

[29] Raymond Montgomery. February 2018. How Does a person who is blind watch the Big Game? https://www.bosma.org/Blog/February-2018/How-does-a-person-who-is-blind-watch-the-Big-Game

[30] Meredith Ringel Morris, Jazette Johnson, Cynthia L. Bennett, and Edward Cutrell. 2018. Rich Representations of Visual Content for Screen Reader Users. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–11. https://doi.org/10.1145/3173574.3173633

[31] Yuri Nishikawa, Hitoshi Sato, and Jun Ozawa. 2018. Multiple sports player tracking system based on graph optimization using low-cost cameras. In *2018 IEEE International Conference on Consumer Electronics (ICCE)*. 1–4. https://doi.org/10.1109/ICCE.2018.8326126 ISSN: 2158-4001.

[32] NVivo. 1997. NVivo. https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home

[33] Hiroyuki Ohshima, Makoto Kobayashi, and Shigenobu Shimada. 2021. Development of Blind Football Play-by-play System for Visually Impaired Spectators: Tangible Sports. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–6. https://doi.org/10.1145/3411763.3451737

[34] Peter Meijer. Retrieved August 2022. The vOICe. https://www.seeingwithsound.com/

[35] Bridget Pettitt, Katharine Sharpe, and Steven Cooper. 1996. AUDETEL: Enhancing television for visually impaired people. *British Journal of Visual Impairment* 14, 2 (May 1996), 48–52. https://doi.org/10.1177/026461969601400202 Publisher: SAGE Publications Ltd.

[36] Venkatesh Potluri, Tadashi E Grindeland, Jon E. Froehlich, and Jennifer Mankoff. 2021. Examining Visual Semantic Understanding in Blind and Low-Vision Technology Users. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–14. https://doi.org/10.1145/3411764.3445040

[37] Denise Prescher, Jens Bornschein, Wiebke Kohlmann, and Gerhard Weber. 2018. Touching Graphical Applications: Bimanual Tactile Interaction on the HyperBraille Pin-Matrix Display. *Universal Access in the Information Society* 17, 2 (June 2018), 391–409. https://doi.org/10.1007/s10209-017-0538-8

[38] Arthur A Raney and Jennings Bryant. 2006. *Handbook of Sports and Media*. Chapter 19: Why we watch and enjoy mediated sports.

[39] Kyle Rector, Keith Salmon, Dan Thornton, Neel Joshi, and Meredith Ringel Morris. 2017. Eyes-Free Art: Exploring Proxemic Audio Interfaces For Blind and Low Vision Art Engagement. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (Sept. 2017), 1–21. https://doi.org/10.1145/3130958

[40] Andreas Reichinger, Stefan Maierhofer, and Werner Purgathofer. 2011. High-Quality Tactile Paintings. *Journal on Computing and Cultural Heritage* 4, 2 (Nov. 2011), 1–13. https://doi.org/10.1145/2037820.2037822

[41] Santander. 2019. Fieeld. https://www.santander.com/en/press-room/press-releases/santander-presents-fieeld-a-deviceenabling-blind-people-to-watch-football-using-their-fingertips

[42] Brad Schultz and Wei Wei. 2013. Sports broadcasting: History, technology, and implications. In *Routledge Handbook of Sport Communication*. Routledge. Num Pages: 9.

[43] Alexa Siu, Gene S-H Kim, Sile O'Modhrain, and Sean Follmer. 2022. Supporting Accessible Data Visualization Through Audio Data Narratives. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–19. https://doi.org/10.1145/3491102.3517678

[44] Alexa F. Siu, Son Kim, Joshua A. Miele, and Sean Follmer. 2019. shapeCAD: An Accessible 3D Modelling Workflow for the Blind and Visually-Impaired Via 2.5D Shape Displays. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, Pittsburgh PA USA, 342–354. https://doi.org/10.1145/3308561.3353782

[45] Nancy Staggers and David Kobus. 2000. Comparing Response Time, Errors, and Satisfaction Between Text-based and Graphical User Interfaces During Nursing Order Tasks. *Journal of the American Medical Informatics Association : JAMIA* 7, 2 (2000), 164–176. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC61470/

[46] Takamasa Tsunoda, Yasuhiro Komori, Masakazu Matsugu, and Tatsuya Harada. 2017. Football Action Recognition Using Hierarchical LSTM. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, Honolulu, HI, USA, 155–163. https://doi.org/10.1109/CVPRW.2017.25

[47] Unity Technologies. 2018. Unity. https://unity.com/

[48] Roman Voeikov, Nikolay Falaleev, and Ruslan Baikulov. 2020. TTNet: Real-time temporal and spatial video analysis of table tennis. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, Seattle, WA, USA, 3866–3874. https://doi.org/10.1109/CVPRW50498.2020.00450

[49] World Wide Web Consortium (W3C). 2022. Making Audio and Video Media Accessible. https://www.w3.org/WAI/media/av/

[50] World Wide Web Consortium (W3C). 2022. W3C Image Concepts. https://www.w3.org/WAI/tutorials/images/