Border: A Live Performance Based on Web AR and a Gesture-Controlled Virtual Instrument

Kiyu Nishida
School of Design, Kyushu University, Fukuoka, Japan/
University of Arts and Design
Karlsruhe, Germany
nishida.kiyu.621@s.kyushu-u.ac.jp

Akishige Yuguchi
Graduate School of Information Science
Nara Institute of Science and Technology
Nara, Japan

Kazuhiro Jo
Faculty of Design, Kyushu University, Fukuoka, Japan/
YCAM
Yamaguchi, Japan
jo@design.kyushu-u.ac.jp

Paul Modler
Dep. of Media Art - Sound
University of Arts and Design
Karlsruhe, Germany
pmodler@hfg-karlsruhe.de

Markus Noisternig
IRCAM-CNRS-Sorbonne
Université (UMR STMS)
Paris, France
markus.noisternig@ircam.fr

ABSTRACT

Recent technological advances, such as increased CPU/GPU processing speed, along with the miniaturization of devices and sensors, have created new possibilities for integrating immersive technologies in music and performance art. Virtual and Augmented Reality (VR/AR) have become increasingly interesting as mobile device platforms, such as up-to-date smartphones, with necessary CPU resources entered the consumer market. In combination with recent web technologies, any mobile device can simply connect with a browser to a local server to access the latest technology. The web platform also eases the integration of collaborative situated media in participatory artwork. In this paper, we present the interactive music improvisation piece 'Border,' premiered in 2018 at the Beyond Festival at the Center for Art and Media Karlsruhe (ZKM). This piece explores the interaction between a performer and the audience using web-based applications – including AR, real-time 3D audio/video streaming, advanced web audio, and gesture-controlled virtual instruments – on smart mobile devices.

Author Keywords

Web AR, Gesture Control, Collaborative Media

CCS Concepts

•**Human-centered computing → Mixed / augmented reality;** Web-based interaction; •**Applied computing →** *Performing arts;*

1. INTRODUCTION

In recent years, the number of live performances experimenting with media-based technologies has rapidly increased, raising many questions regarding musical expression and aesthetics [14]. Collaborative situated and interactive media, for example, questions the relationship of a performer with the audience, as the audience members change their roles from passive listeners to active participants [6, 10].

A wide range of different ways to achieve interactively and collaborated live performances have been proposed [8, 17]. Some of them are based on intensive research and development focusing on various technical, scenographic, musical and overall aesthetic aspects. The next paradigm shift in this area involves mobile devices connected to the web – from simple sensors to smartphones and wearables – that create a much more connected environment and provide users with easy access to multi-media content [15].

Recent technological advances, such as increased CPU and GPU processing speed, along with the miniaturization of devices and sensors, have changed the way mobile devices are used in interactive and participatory artwork. Mobile devices not only receive media streams from servers, but also help to integrate local environmental information through motion sensors, cameras, and microphones. This enables Augmented Reality (AR) on mobile devices, which, in contrast to Virtual Reality (VR), layers computer-generated content on top of the existing reality. Using motion and gesture tracking provides natural human-computer interaction methods and enables embodied generative music and sound control. With such a system, the listener becomes a performer playing the virtual world such as a 'virtual instrument' [5], and AR is expected to enhance audience participation in a concert situation [11].

One of the main remaining challenges in AR is to provide the listener with a consistent immersive audio experience. Higher-order Ambisonics (HOA), for example, is a spatial audio format that is widely used for VR/AR applications. It can be decoded over both loudspeakers [18] and headphones [4, 13] and combines good sound localization with a high level of immersion. Up-to-date software tools for spatial audio production provide advanced 3D real-time audio processing methods for interactive sound source control and room effect simulation in VR/AR environments [2, 3].

In combination with web technologies, such as HTML5[1] and WebGL[2], any mobile device with a browser can simply connect to a server without needing to download and install an app prior to the performance. Web-based applications are easy to maintain and offer seamless functionality on different platforms and operating systems (although not all

NIME'19, June 3-6, 2019, Federal University of Rio Grande do Sul, Porto Alegre, Brazil.

[1] HTML5, `https://dev.w3.org/html5/spec-LC/`

[2] WebGL, `https://www.khronos.org/registry/webgl`

web browsers support VR/AR content).

2. MOTIVATION

Several artworks have explored ways to integrate 3D virtual content into live performances providing new experiences for both audience members and performers. The collaborative mixed reality environment for *Reflets* [1], for example, displays virtual objects onto semi-transparent panels placed between the audience and the performer. *Con i piedi per terra* [11] is another example of a participatory performance that uses AR. A specific application was developed for this performance that provides interactive AR content on smartphones and enables audience members to contribute to the performance. Our piece, 'Border,' on the contrary, gives a single performer the ability to add multiple layers of immersive audiovisual content to a live music performance using AR and interactive technologies. This is achieved by extending the timeline of the performance through live looping. The performer recorded his/her movements over a certain time interval, and the recordings were then continuously repeated in the AR environment. Each 'instance' of the performer (i.e. each of the recorded video clips) was then displayed at a certain position in the AR environment. The audience members could access the AR video stream by connecting their smart devices to a web-based application. Figure 1 shows the AR view on a smartphone. To enable the performer to control sound and music in this AR environment, we developed a gesture-controlled virtual instrument, which uses Kinect body motion tracking.

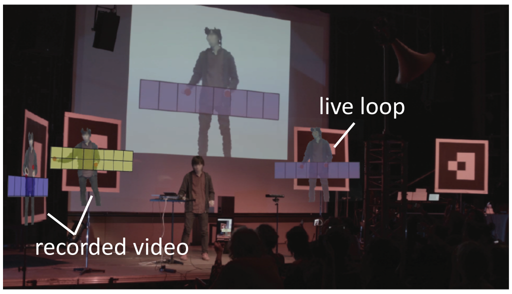

Figure 1: The AR view through the audience members' smartphones (post-edit)

3. ARCHITECTURE AND DESIGN

In this section, we describe the web-based application, including the AR and audience interaction, and the gesture-controlled virtual instrument.

3.1 Web-Based Application

We decided to design a web-based application so that a performer could communicate and interact with the audience members. The web application was built on the cross-platform JavaScript run-time environment Node.js[3]. To maintain high bandwidth access and good network stability, an isolated local area network (LAN) was installed using a central Wi-Fi router. Audience members could access the web-based application by connecting their mobile devices to the wireless LAN and then opening the web address indicated on the screen before the start of the performance.

3.1.1 *Augmented Reality Environment*

In our piece, the web-based AR environment was implemented using the cross-platform JavaScript library AR.js[4] which runs on every mobile device browser supporting WebGL and WebRTC[5]. AR.js works with specific visual markers (i.e. a sort of simplified QR-code) to insert specific 3D scenes in the video. Figure 2 shows the processing chart of our AR system for inserting the live performer into the AR stream on a mobile device. The Kinect sensor provides video and depth information. Using the obtained depth information, the performer's body shape area is extracted from the video stream, and the background of this layer mask is filled with solid white in Processing[6]. This live video stream is then sent through Max[7] to CamTwist[8]. The video from CamTwist is recognized as a web camera source by a web browser on the server and is transmitted in real time using WebRTC. At the same time, the live video stream was recorded for a certain time interval in Max, then stored on the server and uploaded – in the background – to the mobile devices through the web interface, where it is played back in a loop as an additional layer to already playing videos, as shown in Figure 1. Each frame of both the loaded loops and the transmitted video is copied onto an HTML5 canvas, a container element used to process graphics. Deleting every white pixel of the layer mask extracts the performer instance, which is then displayed in the AR stream.

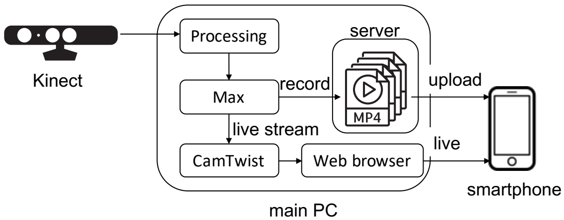

Figure 2: AR system processing chart

3.1.2 *Interaction with Audience Members*

To interact with audience members, the performer sends an audio stream to a mobile device in real time. The implementation of this functionality combines tools such as WebRTC, CamTwist, Max, and the Web Audio API[9]. Audience members controlled the playback volume of the looped sounds by tilting their smartphones. This means the performer can use the mobile devices in the audience to add an additional layer of musical expression to his/her performance and the audience members become active performers, using their mobile devices as musical instruments.

3.2 Gesture-Controlled Virtual Instrument

To support various kinds of instruments and timbres with the performer's motions, a gesture-controlled virtual instrument was designed based on body tracking with a Kinect sensor. As shown in Figure 3, with this instrument, a performer can switch between two modes using simple movements, the 'selection mode' for choosing an instrument, and the 'play mode' for playing the selected instrument. Each selected virtual instrument represents a unique timbre, a sampled sound, or an audio effect. In the play mode, shown in the right sub-figure, the performer's personal space (i.e. the region of space immediately surrounding the body) is partitioned into a limited number of areas representing MIDI notes which trigger sounds each time the performer 'touches' one of these areas. Figure 4 shows the system architecture of the gesture-controlled virtual instrument.

[3] Node.js, https://nodejs.org/en/

[4] AR.js, https://github.com/jeromeetienne/AR.js

[5] WebRTC, https://webrtc.org/

[6] Processing, https://processing.org/

[7] Max, https://cycling74.com/

[8] CamTwist, http://camtwiststudio.com/

[9] Web Audio API, https://developer.mozilla.org/en-US/docs/Web/API/Web_Audio_API

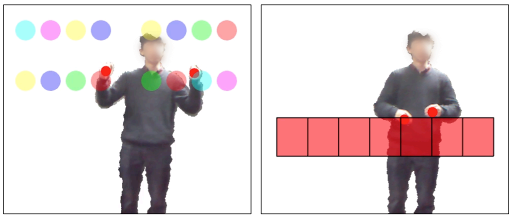

Figure 3: Gesture-controlled virtual instrument: selection mode (left) and play mode (right)

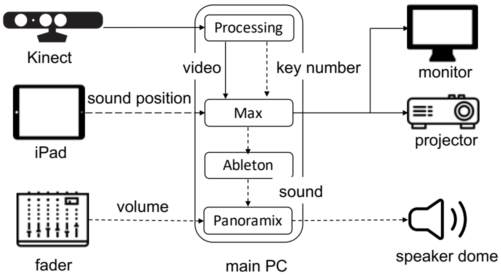

Figure 4: Gesture-controlled virtual instrument signal processing chart

The interface of the virtual instrument was implemented in Processing and Max, and controlled through the Kinect motion tracking; Ableton Live[10] was used for audio recording and playback, and Panoramix[11] for real-time spatial audio processing. According to the gestures tracked by the Kinect, Processing detected when a MIDI key was activated and provided the associated video stream. The video was then streamed in real-time to Max using Syphon[12] and projected on a video screen on stage to help the audience understand the computer-generated AR scenes on their mobile devices. In parallel, the video stream was sent to a computer screen on stage to provide the performer with some visual feedback. All building blocks communicated via the Open Sound Control (OSC) protocol [16] and User Datagram Protocol (UDP). The MIDI notes were interpreted in Max, triggering the sounds in the digital audio workstation Ableton Live. The Ableton Live audio outputs were connected to Panoramix using Soundflower[13], although the sound source positions are sent as OSC message bundles using ToscA[14]. In addition, an iPad was set up for extended remote control of Panoramix, Ableton Live, and Max.

Moreover, HOA was used for 3-D audio rendering in real-time in order to provide an immersive audio experience. The audio was played back over a 43-speaker dome using the 'energy preserving' decoder [18].

4. LIVE PERFORMANCE

We performed 'Border' at the Beyond Festival 2018 'Future Sound' at the ZKM in Karlsruhe, Germany. Figures 5 and 6 show the stage plan, without the surrounding speaker array, and a photo of the stage. The photo shows the five AR markers, each with a dimension of 1x1 square meters to guarantee that smartphones at the far end of the auditorium can recognize them. Each AR marker defines the position of a 'visual instance' of the performer. An audience of 120 attended the concert, and on average, 80 participated actively with their smartphones. At the beginning of the concert, instructions were given to the audience on how to set up their smartphones and on how to interact with the AR environment. Then, the audience held their smartphones in their hands. During the performance, the performer introduced the way to interact with a sound using the smartphone, and ways to actively participate in the show were demonstrated.

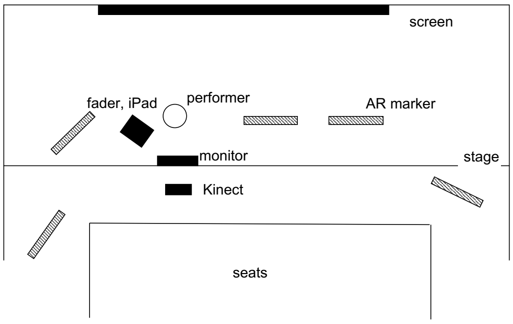

Figure 5: Stage setting (the surrounding 43-channel speaker dome is not depicted)

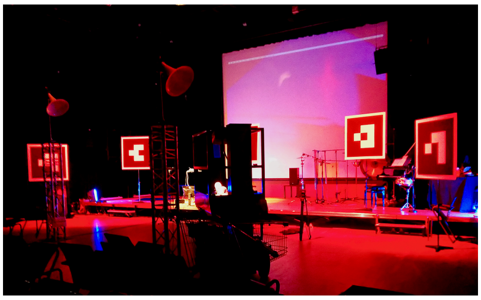

Figure 6: Performance at the 'Beyond Festival 2018' at the Center for Art and Media Karlsruhe (ZKM)

5. DISCUSSION

This section discusses the interactive AR environment, the gesture-controlled virtual instrument, and the experience of the audience members.

5.1 Augmented Reality

We designed our interactive AR environment as a first step in integrating a virtual world with the real world. There are still many steps to achieve this concept completely. We describe two key considerations here.

5.1.1 Management of Computers

Both processing a video and displaying a video as a virtual object take a large amount of CPU and GPU power. In terms of that and the bandwidth for the communication lines, we had to use a quarter of the video resolution compared to the original resolution from the Kinect sensor, which caused a significant drop in the sense of immersion.

5.1.2 Limits of a web-based application

AR.js displays virtual objects on real-world video streams based on the position of an AR maker. One problem with this method is that maintaining the spatial relationship between markers is difficult, specifically, synchronizing the virtual positions with the absolute real ones. For this reason, we displayed the performer instances using the relative positions of the AR markers, which limited the expression of AR in the web browser. Extending the web browser's AR library with the ability to synthesize a virtual world in sync with the real world more fluently would allow us to create

[10] Ableton live, `https://www.ableton.com/en/`

[11] Panoramix [2], `https://forumnet.ircam.fr/product/panoramix-en/`

[12] Syphon, `http://syphon.v002.info/`

[13] Soundflower, `https://soundflower.en.softonic.com/`

[14] ToscA, `http://forumnet.ircam.fr/product/tosca-en/`

more dynamic and rich effects. In fact, this also could be achieved through a dedicated native application; however, as described above, such an application would be less useful in a concert with a heterogeneous audience.

5.2 Gesture-Controlled Virtual Instrument

In this performance, the sound was mostly controlled by gestural movements of the performer, although for certain instructions, a tablet interface was used. The performer experienced this separation of the interface into two parts as less practical for control. An alternative method could be to use an additional tracking system such as eye movements [9] or certain hand gestures [7, 12]. In combination with the motion tracking used in this performance, the integration of these options would improve the control, and so expand the expressiveness of the performer significantly.

The performer also said the visual feedback provided by the monitor was not easy to see because he had to look down, and it causes to lose visual communication with the audience. A solution to this problem is to use a mixed reality panel as mentioned in [1]. Such a display enables the performer to get feedback and visual communication. It also enables the audience members to get more immersion, combining with AR provided by their mobile devices.

On the other hand, the latency between the moment the performer hits a key with his gesture on the virtual instrument and the actual occurrence of the sound was not perceived as disruptive, because the performance did not have a dominant beat such as a drum track with a strict tempo. However, special care must be taken in such an environment if it follows a dominant beat.

5.3 Perspective of Audience Members

Some people reported after the concert that holding up a smartphone during the performance was exhausting and fatiguing. Others reported that they watched the AR through the smartphone for only a very short period, compared to the total length of the performance. A possible solution is using smartphones on cardboard carriers. Although this may not replace a head-mounted display (e.g., Vive, HoloLens or Oculus Rift), it provides inexpensive accessible technologies for integrating a large number of audience members in immersive and participatory music and performance art, with a low physical effort and minimum fatigue.

6. CONCLUSIONS

This paper presents the immersive and interactive performance 'Border' which explores the performer/audience interaction using web-based applications including augmented reality, real-time 3D audio/video streaming, advanced web audio technologies, and a gesture-controlled virtual instrument. Many artists have explored the interaction between a performer and the audience, and it is obvious that interactive VR/AR changes the role of the audience members from passive listeners to active participants. 'Border' highlights how the use of web technologies and mobile devices can provide easy access to collaborative VR/AR in art performances. A video of the performance presented in this paper is available at: `https://youtu.be/_4yrxCEap1M`

7. ACKNOWLEDGMENTS

This project was supported in part by JSPS KAKENHI (Grant No. JP17H04772) and the Japanese Public-Private Partnership Student Study Abroad Program 'TOBITATE! Young Ambassador Program'. The authors further thank Christian Sünkeberg for IT support, the colleagues of the Media Art Department of the Karlsruhe University of Arts and Design for fruitful discussions, as well as the Center for Art and Media Karlsruhe (ZKM) and the Beyond Festival 2018 for the invitation to premiere this piece.

8. REFERENCES

[1] F. Berthaut, D. M. Plasencia, M. Hachet, and S. Subramanian. Reflets: Combining and revealing spaces for musical performances. In *Proc of the 2014 Int Conf on New Interfaces for Musical Expression (NIME'15)*, pages 116–120, Baton Rouge, US, May 2015.

[2] T. Carpentier. Panoramix: 3d mixing and post-production workstation. In *Proc of the 42nd Int Computer Music Conf (ICMC'16)*, pages 7–12, Utrecht, Netherlands, June 2016.

[3] T. Carpentier, M. Noisternig, and O. Warusfel. Twenty years of ircam spat: looking back, looking forward. In *Proc of the 41st Int Computer Music Conf (ICMC'15)*, pages 7–12, Denton, TX, USA, June 2015.

[4] J. Daniel, J.-B. Rault, and J.-D. Polack. 3d binaural sound reproduction using a virtual ambisonic approach. In *Proc of the 105th AES Conv*, page 4795, San Francisco, USA, June 1998.

[5] E. Egozy and E. Y. Lee. *12*: Mobile phone-based audience participation in a chamber music performance. In *Proc of the 2018 Int Conf on New Interfaces for Musical Expression (NIME'18)*, pages 7–12, Blacksburg, VA, USA, June 2018.

[6] J. Freeman. Large audience participation, technology, and orchestral performance. In *Proc of the 41st Int Computer Music Conf (ICMC'13)*, pages 757–760, Denton, TX, USA, June 2015.

[7] J. Han and N. Gold. Lessons learned in exploring the leap motion sensor for gesture-based instrument design. In *Proc of the 2014 Int Conf on New Interfaces for Musical Expression (NIME'14)*, pages 371–374, Goldsmiths, UK, June 2014.

[8] A. Hindle. Swarmed: Captive portals, mobile devices, and audience participation in multi-user music performance. In *Proc of the 39th Int Computer Music Conf (ICMC'13)*, Daejeon, Korea, June 2013.

[9] A. J. Hornof. The prospects for eye-controlled musical performance. In *Proc of the 2014 Int Conf on New Interfaces for Musical Expression (NIME'14)*, Goldsmiths, UK, June 2014.

[10] K. Jo and A. Tanaka. A. the music participates in. in f. schroeder (ed.) performing technology: User content and the new digital media. *Cambridge Scholars Publishing*, pages 34–50, 2009.

[11] D. Mazzanti, V. Zappi, D. Caldwell, and A. Brogni. Augmented stage for participatory performances. In *Proc of the 2014 Int Conf on New Interfaces for Musical Expression (NIME'14)*, pages 29–34, Goldsmiths, UK, June 2014.

[12] P. Modler and A. Myatt. Video based recognition of hand gestures by neural networks for the control of sound and music. In *Proc of the 2008 Int Conf on New Interfaces for Musical Expression (NIME'08)*, pages 358–359, Genova, Italy, June 2008.

[13] M. Noisternig, T. Musil, A. Sontacchi, and R. Holdrich. 3d binaural sound reproduction using a virtual ambisonic approach. In *Proc of the IEEE Symp VECIMS*, pages 174–178, Lugano, Switzerland, June 2003.

[14] J. Oh and G. Wang. Audience-participation techniques based on social mobile computing. In *Proc of the 37th Int Computer Music Conf (ICMC'11)*, pages 665–672, Huddersfield, UK, June 2011.

[15] L. Turchet, C. Fischione, G. Essl, D. Keller, and M. Barthet. Internet of musical things: Vision and challenges. *IEEE Access*, 6:61994–62017, September 2018.

[16] M. Wright and A. Freed. Open sound control: a new protocol for communicating with sound synthesizers. In *Proc of the 23rd Int Computer Music Conf (ICMC'97)*, pages 101–104, Thessaloniki, Hellas, Greece, June 1997.

[17] Zeitbloom et al. Biomorph, 2003. `http://www.onarchitektur.de/index.php/Biomorph`.

[18] F. Zotter, H. Pomberger, and M. Noisternig. Energy-preserving ambisonic decoding. *Acta Acust United Acust*, 98(1):37–47, November 2012.