

# Border: webARとジェスチャー制御による 仮想楽器を用いたライブパフォーマンス

西田 騎夕<sup>1,a)</sup> 城 一裕<sup>1,b)</sup>

**概要:** 本発表では 2018 年にカールスルーエ・メディア・アンド・デザイン・センター, ビヨンドフェスティバルにおいて初演した観客参加型の即興演奏作品 “Border” について述べる. 本作品では携帯型端末上でのウェブアプリケーションによるオーギュメントド・リアリティ (AR), リアルタイムでの 3D 映像と音の配信および, ジェスチャー制御による仮想楽器を用いた演者と観客とのインタラクションについて探求した.

**キーワード:** WebAR, ジェスチャー制御, 共創メディア

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

KIYU NISHIDA<sup>1,a)</sup> KAZUHIRO JO<sup>1,b)</sup>

**Abstract:** 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 Augmented Reality (AR), real-time 3D audio/video streaming, advanced web audio– on smart mobile devices, and gesture-controlled virtual instruments.

**Keywords:** Web AR, Gesture Control, Collaborative Media

### 1. 序論

近年, メディア技術を用いた実験的なライブパフォーマンスの試みが急速に増加しており, 音楽の表現と美学に関する様々な議論が生じている [15]. 例えば, 共創的でインタラクティブなメディアでは観客が受動的聴取者から能動的参加者へと変化するため, 演者と観客との関係性についての議論が生じている [5], [10].

インタラクティブで共創的なライブパフォーマンスを実現する様々な方法が幅広く提案されており [8], [18], その中には様々な技術的, 舞台芸術的, 音楽的, そして全般的

な美に焦点を当てた集中的な調査と開発に基づいているものもある. この分野における次のパラダイムシフトは, よりウェブに接続された環境を構築し, ユーザーにマルチメディアへの容易なアクセスを提供する携帯型端末 – 簡易的なセンサーからスマートフォンやウェアラブルまで – を用いたものである [16].

CPU や GPU の処理能力の向上に加えて端末やセンサーの小型化などの技術発達はインタラクティブな観客参加型の作品における携帯型端末の使用方法を変化させている. 携帯型端末はサーバーからの情報を受け取るだけでなく, モーションセンサー, カメラ, マイクロフォンを通して周辺環境の情報を統合することを可能にしている. これらの発達は携帯型端末上でのオーギュメントド・リアリティ (AR) を可能にした. AR はヴァーチャルリアリティ

<sup>1</sup> 九州大学芸術工学部音響設計学科  
Kyushu University School of Design Department of Acoustic,  
Fukuoka, Japan

<sup>a)</sup> nishida.kiyu.621@s.kyushu-u.ac.jp

<sup>b)</sup> jo@design.kyushu-u.ac.jp

(VR)とは対照的にコンピューターによって生成されたコンテンツを現実世界の上に描画できるため、コンサートにおいて観客のパフォーマンスの能動的参加を向上させることができるかと期待されている。

AR環境において聴取者に安定した没入感のある音体験をいかに提供するかということは大きな課題である。例えば、Higher-order Ambisonics (HOA) はVRやAR環境向けに広く用いられているフォーマットである。HOAはスピーカー [19] とヘッドホン [4], [14] のそれぞれで聴取することが可能なフォーマットであり、定位に優れ、高水準の没入感を実現できる。最新の立体音響制作のためのソフトウェアは、VRやAR環境においてインタラクティブな音源制御や室内音響効果のシミュレーションのための先進的な3Dリアルタイム音響処理を可能としている [2], [3]。

HTML5<sup>\*1</sup>やWebGL<sup>\*2</sup>などウェブ技術の発達とともに、ウェブブラウザを備えた全ての携帯型端末はアプリケーションをダウンロード、インストールすることなくサーバーに接続するだけで即座にウェブベースのアプリケーションを利用することが可能となった。ウェブベースのアプリケーションは保守が容易であり、異なるオペレーティング・システム(OS)やウェブブラウザ上でも円滑に機能するという点においてコンサートなどの場面ではネイティブのアプリケーションと比較して優位である [7]。

## 2. 動機

観客と演者双方に新たな体験を提供するために、3Dのヴァーチャルなコンテンツをライブ演奏に取り入れる試みが複数の芸術作品によって探求されてきた。例えば、*Reflets*[1]ではヴァーチャルな物体を観客と演者の間に立てられた半透明のパネルに映すことで共創的なミックス・リアリティ環境を構築した。また、*Con i piedi per terra* [11]ではインタラクティブなAR環境を提供するスマートフォン向けのアプリケーションを用いることによって観客が能動的にパフォーマンスに関与できる方法を提案した。これらの作品とは対照的に本作品“Border”ではARとインタラクティブ技術を用いることによりライブ演奏に複数の没入感のある視聴覚コンテンツを付加することを可能とした。このために、ライブループを基本としてパフォーマンスが行われた。すなわち演者はある一定の時間自身の演奏を録音・録画し、その映像がAR環境でそれぞれ繰り返し再生された。それぞれ録画された演者の“クローン”はAR上のそれぞれの場所に表示された。観客はスマート端末からウェブアプリケーションに接続することによってARを体験した。図1はスマート端末上に表示されたARの例を示している。加えて、演者がAR環境において音を制御できるようにするために、Kinectのモーショントラッ

\*1 HTML5, <https://dev.w3.org/html5/spec-LC/>

\*2 WebGL, <https://www.khronos.org/registry/webgl>



図1 観客のスマートフォンに表示されたAR(後編集)

Fig. 1 The AR view through the audience members' smartphones (post-edit)

キングを用いたジェスチャー制御による仮想楽器を開発した。

## 3. システム設計

本章ではARと観客の参加を実現したウェブベースのアプリケーション、ジェスチャー制御による仮想楽器のシステム設計について述べる。

### 3.1 ウェブベースのアプリケーション

観客と演者に双方向性のあるコミュニケーションを提供するために、ARと音によるインタラクションを実現するウェブアプリケーションを開発した。サーバーの構築にはNode.js<sup>\*3</sup>を使用し、広帯域のネットワーク容量と安定性を確保するため専用のWi-Fiルーターを用いて、ローカルエリアネットワーク(LAN)を形成した。観客はパフォーマンスが始まる前に携帯型端末を無線LANに接続し、前方スクリーンに示されたIPアドレスをウェブブラウザに入力することによってウェブアプリケーションに接続した。

#### 3.1.1 オーギュメントド・リアリティ環境

本作品では、ウェブベースのAR環境を構築するためにクロスプラットフォームのJavaScriptライブラリーであるAR.js<sup>\*4</sup>を用いた。AR.jsはWebGLとWebRTC<sup>\*5</sup>をサポートする全ての携帯型端末のウェブブラウザ上で機能する。AR.jsはQRコードを簡略化したような特有のARマーカ―を認識することによって仮想環境を現実のカメラ映像の上に描画する。

図2は演者をAR上に表示させるためのシステム図である。Kinectからカメラ映像と深度情報が取得され、Processing<sup>\*6</sup>上で、得られた深度情報を用いてカメラ映像から背景を白いピクセルでマスクすることによって、カメラ映像から演者のみを取り出された。この映像はMax<sup>\*7</sup>を通してCamTwist<sup>\*8</sup>に送られた。CamTwistからの映像はサー

\*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/>

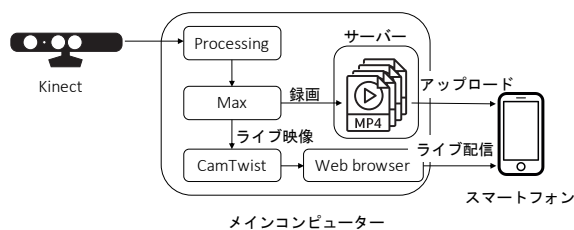


図 2 AR システム図

Fig. 2 AR system processing chart

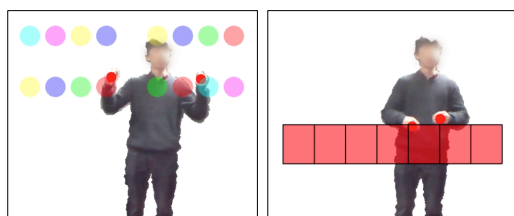


図 3 ジェスチャー制御による楽器：  
 選択モード（左）と演奏モード（右）

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

バー上のウェブブラウザでウェブカメラ入力として認識され、WebRTC を用いてリアルタイムに観客の携帯型端末へと配信された。同時に、Processing からの映像は Max である一定の時間録画、サーバーに保存、観客の携帯型端末にロードされ、図 1 に示すように次々とループ再生された。ロードされた映像とリアルタイムに配信された映像の各フレームは HTML5 の画像処理を行う要素である canvas にコピー、全ての白いピクセルが削除され、演者のみが AR 上に表示された。

### 3.1.2 観客とのインタラクション

観客とインタラクションを図るために、演者が演奏した音がリアルタイムに観客の携帯型端末に配信された。この機能は、WebRTC、CamTwist、Max、Web Audio API<sup>\*9</sup>などを用いて実装された。これにより観客は携帯型端末を傾けることでリアルタイムにこの音の再生音量を制御することができた。これによって、演者は観客の携帯型端末を音楽的表現に使用することが可能となり、さらに観客を能動的参加者にする事ができる。

### 3.2 ジェスチャー制御による仮想楽器

演者の動きに多様性を与え、様々な種類の楽器や音色を変化させられるように、ジェスチャー制御による仮想楽器を Kinect を用いたモーショントラッキングによって設計した。図 3 に示すようにこの楽器では演者が簡単な身振りを使って、楽器を選択する“選択モード”（それぞれの丸は各楽器を表している）と選択された楽器を演奏する“演奏モード”の 2 つのモードを選択できる。それぞれの選択さ

\*9 Web Audio API, [https://developer.mozilla.org/en-US/docs/Web/API/Web\\_Audio\\_API](https://developer.mozilla.org/en-US/docs/Web/API/Web_Audio_API)

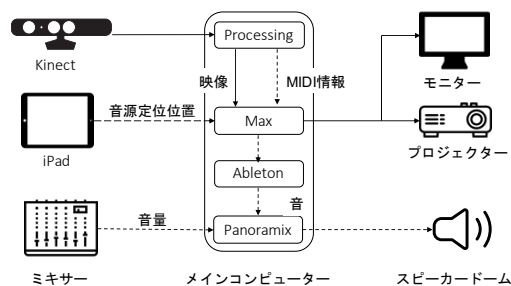


図 4 ジェスチャー制御による楽器 システム図

Fig. 4 Gesture-controlled virtual instrument signal processing chart

れた仮想楽器はそれぞれに音色、サンプル音源、オーディオエフェクトと対応する。図 3 の右図の演奏モードでは演者の周囲に MIDI ノートと対応する複数の演者が“触れる”と音を発する区域が表示される。図 4 はジェスチャー制御による仮想楽器のシステム図を示している。

仮想楽器のインターフェースは Processing と Max を用いて実装され、Kinect のモーショントラッキングによって制御された。さらに音の再生や録音に Ableton Live<sup>\*10</sup>、リアルタイムでの 3D 音響処理に Panoramix<sup>\*11</sup> を用いた。Kinect によってトラッキングされたジェスチャーに従って Processing は MIDI 鍵盤が押されたのを検知、また同時に映像を出力した。この映像はリアルタイムに Syphon<sup>\*12</sup>を通して Max に配信され、観客が携帯型端末上の AR 環境を理解しやすいようにステージ上のスクリーンに投影された。同時に、演者にフィードバックを提供するためにステージ上のビデオモニターにも出力された。アプリケーション間の情報のやり取りは Open Sound Control (OSC) protocol [17] と User Datagram Protocol (UDP) を用いて行われた。Max でマッピングされた MIDI 情報は Ableton live に送信され、音を発生させた。Ableton live において発生した音は Soundflower<sup>\*13</sup>を通して Panoramix へ送られ、並行して音源の定位位置は Tosca<sup>\*14</sup>を用いて OSC として送られた。加えて、iPad が Panoramix、Ableton Live、and Max の補助的な制御に用いられた。

没入感のある音体験を提供するために 3D 音響のレンダリング手法として HOA を用いた。音は“energy preserved”デコーダーを用いて、43 個のスピーカーからなるスピーカーカードームで再生された。

### 4. ライブパフォーマンス

私たちは‘Border’をドイツ、カールスルーエ市の ZKM で行われた Beyond Festival 2018, Future Sound で発表し

\*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/>

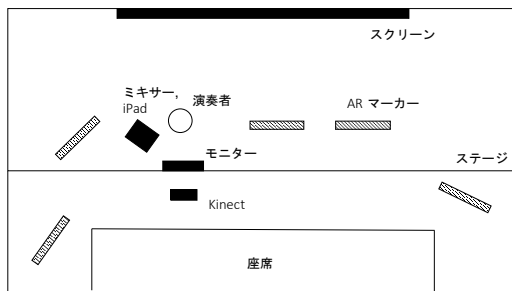


図 5 ステージプラン (周囲の 43 チャンネルスピーカードームは除く)

Fig. 5 Stage setting (the surrounding 43-channel speaker dome is not depicted)

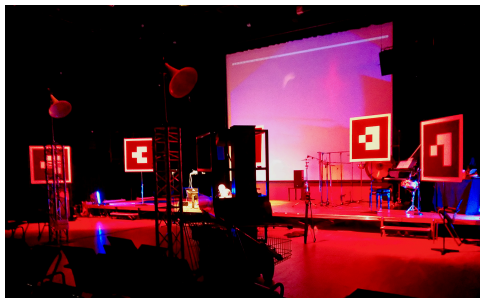


図 6 ステージの様子

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

た．図 5 はスピーカードームを除いたステージプランを，図 6 は，ステージの様子を示している．

## 5. 議論

本章ではインタラクティブな AR 環境，ジェスチャー制御による仮想楽器，観客の体験について議論する．

### 5.1 オーギュメントド・リアリティ環境

私たちは仮想世界と現実世界とを統合したライブパフォーマンスの第一歩としてインタラクティブな AR 環境を設計した．このコンセプトを完遂するまでにはいまだに数多くの問題点が存在する．ここに鍵となる 2 つの検討事項について示す．

#### 5.1.1 携帯型端末の処理能力の限界

携帯型端末上で映像を加工し，AR 上に表示することは高い CPU/GPU の処理能力を必要とする．これとネットワークの帯域限界の観点から，本作品では Kinect から出力される映像の四分の一の解像度の映像でしか AR 上に表示できなかった．このことは没入感を下げる要因となる．今後より高性能な携帯型端末の普及，またサーバー処理のような新しい技術の普及によってこの問題は解消できるだろう．

#### 5.1.2 ウェブベースアプリケーションの限界

AR.js は AR マーカーの位置を基準として現実世界の上

にヴァーチャルな物体を表示する．つまり，私たちは演者のクローン映像を AR マーカーからの相対座標でしか表示することができず，現実世界と仮想世界の絶対座標を一致させることは困難である．このことはウェブブラウザにおける AR の表現領域を制限してしまっている．現実世界と仮想世界をより自然に統合することのできるウェブブラウザで使用できる AR のライブラリーの拡張はより動的で没入感のある AR 環境を作り出すことを可能とするだろう．実際のところ，ネイティブアプリケーションを用いればこのようなことは達成できるが，前述の通り，そのようなアプリケーションは多数の観客のいるコンサートのような場面では有用ではない．

### 5.2 ジェスチャー制御による仮想楽器

本作品では，多少の入力はタブレットを通じて行われたが，ほとんどの音は演者の身振りなどの動きによって制御された．演者は，このような入力装置の分割は実用的ではないと報告している．代替案としては視線の動き [9] や手の動き [6], [12] などの補助的なトラッキングを用いることが考えられる．本作品で用いられたトラッキング方法をこれらの方法を合わせて使用し，インターフェースを向上させることによって演者の表現力をより引き出すことができるだろう．

演者はまたモニターによるフィードバックは視線を下げさせるため，観客とのコミュニケーションを失わせてしまうと報告している．この問題への解決策は [1] に示されているようなミックスド・リアリティのパネルを用いることであろう．このようなディスプレイは演者に観客とのコミュニケーションを失わせることなくフィードバックを提供することができる．またこれを本作品で用いたような方法による AR と組み合わせて使用することで観客に一層の没入感を与えることができると考えられる．

本作品では演者がドラムのようなテンポを持続させる支配的な音を使用しなかったため，演者が仮想楽器の鍵盤を身振りにより叩いた瞬間と実際に音が出るまでの間の遅延は問題にならなかった．しかし，そのような音を使用したい場合には特別な方法を用いる必要があるだろう．

### 5.3 観客の反応

複数の観客がコンサート後に，演奏中にスマートフォンを手で持ち上げたままにしているのは疲れると報告している．さらにスマートフォンを用いて AR を通してパフォーマンスを見ている時間は演奏全体の時間に比べてとても少なかったという報告もある．この問題に対する解決策はダンボール製のヘッドマウント器具を使うことだろう．もちろん，この方法は Vive, HoloLens, Oculus Rift のようなヘッドマウントディスプレイに取って代わることはできないが，多くの観客を少ない労力で没入感のある作品に参加

させることを可能とする。

## 6. 結論

本論文ではジェスチャー制御の仮想楽器と3D音響、またオーギュメントド・リアリティ、リアルタイムでの映像と音声の配信、最新のウェブオーディオ技術を用いたウェブベースのアプリケーションを使用することにより演者と観客との関係性を探求したインタラクティブで没入感のあるパフォーマンス作品“Border”について述べた。多くのアーティストが観客と演者とのインタラクションについて探求しており、VRやARが観客の役割を受動的聴取者から能動的参加者へと変化させることは明らかである。“Border”では共創的なVR/ARのパフォーマンス作品へのアクセスを容易にするウェブテクノロジーと携帯型端末の使用方法について焦点を当てた。本論文で述べたパフォーマンスの様子は以下のリンクより視聴可能である。  
[https://youtu.be/\\_4yrxCeap1M](https://youtu.be/_4yrxCeap1M)

## 7. 謝辞

本研究の一部は、日本学術振興会科研費・若手研究(A)ポストデジタル以降の音を生み出す構造の構築 [17H04772]、トビタテ！留学JAPAN日本代表プログラムの助成を受け実施されている。本作品の制作機会をいただいたカールスルーエ造形大学メディアアート学部、発表機会をいただいた、カールスルーエ・メディア・アンド・デザイン・センター(ZKM)、Beyond Festival 2018に感謝する。また本論文をProc of the 2019 Int Conf on New Interfaces for Musical Expression (NIME'19)にて発表する際に共著して下さった湯口 彰重氏、Paul Modler氏、Markus Noisternig氏に感謝する [13]。

## 参考文献

- [1] Berthaut, F., Plasencia, D. M., Hachet, M. and Subramanian, S.: Reflets: Combining and Revealing Spaces for Musical Performances, *Proc of the 2014 Int Conf on New Interfaces for Musical Expression (NIME'15)*, Baton Rouge, US, pp. 116–120 (2015).
- [2] Carpentier, T.: Panoramix: 3D mixing and post-production workstation, *Proc of the 42nd Int Computer Music Conf (ICMC'16)*, Utrecht, Netherlands, pp. 7–12 (2016).
- [3] Carpentier, T., Noisternig, M. and Warusfel, O.: Twenty years of Ircam Spat: looking back, looking forward, *Proc of the 41st Int Computer Music Conf (ICMC'15)*, Denton, TX, USA, pp. 7–12 (2015).
- [4] Daniel, J., Rault, J.-B. and Polack, J.-D.: 3D binaural sound reproduction using a virtual ambisonic approach, *Proc of the 105th AES Conv*, San Francisco, USA, p. 4795 (1998).
- [5] Freeman, J.: LARGE AUDIENCE PARTICIPATION, TECHNOLOGY, AND ORCHESTRAL PERFORMANCE, *Proc of the 41st Int Computer Music Conf (ICMC'13)*, Denton, TX, USA, pp. 757–760 (2015).
- [6] Han, J. and Gold, N.: Lessons Learned in Exploring the Leap Motion Sensor for Gesture-based Instrument Design, *Proc of the 2014 Int Conf on New Interfaces for Musical Expression (NIME'14)*, Goldsmiths, UK, pp. 371–374 (2014).
- [7] Heino, A.: Garth Paine's 'Future Perfect', in development at IRCAM/ZKM (2018). <https://www.ircam.fr/person/garth-paine/>.
- [8] Hindle, A.: SWARMED: Captive Portals, Mobile Devices, and Audience Participation in Multi-User Music Performance, *Proc of the 39th Int Computer Music Conf (ICMC'13)*, Daejeon, Korea (2013).
- [9] Hornof, A. J.: The Prospects For Eye-Controlled Musical Performance, *Proc of the 2014 Int Conf on New Interfaces for Musical Expression (NIME'14)*, Goldsmiths, UK (2014).
- [10] Jo, K. and Tanaka, A.: A. The Music Participates In. In F. Schroeder (ed.) *Performing Technology: User Content and the New Digital Media*, Cambridge Scholars Publishing, pp. 34–50 (2009).
- [11] Mazzanti, D., Zappi, V., Caldwell, D. and Brogni, A.: Augmented Stage for Participatory Performances, *Proc of the 2014 Int Conf on New Interfaces for Musical Expression (NIME'14)*, Goldsmiths, UK, pp. 29–34 (2014).
- [12] Modler, P. and Myatt, A.: Video Based Recognition of Hand Gestures by Neural Networks for the Control of Sound and Music, *Proc of the 2008 Int Conf on New Interfaces for Musical Expression (NIME'08)*, Genova, Italy, pp. 358–359 (2008).
- [13] Nishida, K., Yuguchi, A., Jo, K., Modler, P. and Noisternig, M.: Border: A Live Performance Based on Web AR and a Gesture-Controlled Virtual Instrument, *Proc of the 2019 Int Conf on New Interfaces for Musical Expression (NIME'19)*, Porto Alegre, Brazil (2019).
- [14] Noisternig, M., Musil, T., Sontacchi, A. and Holdrich, R.: 3D binaural sound reproduction using a virtual ambisonic approach, *Proc of the IEEE Symp VECIMS*, Lugano, Switzerland, pp. 174–178 (2003).
- [15] Oh, J. and Wang, G.: AUDIENCE-PARTICIPATION TECHNIQUES BASED ON SOCIAL MOBILE COMPUTING, *Proc of the 37th Int Computer Music Conf (ICMC'11)*, Huddersfield, UK, pp. 665–672 (2011).
- [16] Turchet, L., Fischione, C., Essl, G., Keller, D. and Barthelet, M.: Internet of Musical Things: Vision and Challenges, *IEEE Access*, Vol. 6, pp. 61994–62017 (2018).
- [17] Wright, M. and Freed, A.: Open sound control: a new protocol for communicating with sound synthesizers, *Proc of the 23rd Int Computer Music Conf (ICMC'97)*, Thessaloniki, Hellas, Greece, pp. 101–104 (1997).
- [18] Zeitbloom et al.: BioMorph (2003). <http://www.onarchitektur.de/index.php/Biomorph>.
- [19] Zotter, F., Pomberger, H. and Noisternig, M.: Energy-Preserving Ambisonic Decoding, *Acta Acust United Acust*, Vol. 98, No. 1, pp. 37–47 (2012).