

推薦研究論文

SDM360²：インタラクティブ3Dコンテンツの 自由視聴点再生

塚田 学^{1,a)} 菰原 裕¹ 粕谷 貴司^{1,2} 新居 英明³ 高坂 茂樹⁴ 小川 景子⁵ 江崎 浩¹

受付日 2018年4月7日, 採録日 2018年5月30日

概要：インターネットを前提とした視聴サービスが登場し、なかでも空間に存在する視聴対象を解釈し、コンテンツとして活用するオブジェクトベースの視聴サービスの重要性が増している。2014年より、Software Defined Media (SDM) コンソーシアムでは、オブジェクトベースのメディアとインターネットを前提とした視聴空間の研究を行っている。現在、音楽イベントのDVDなどのパッケージメディアは、マイクやカメラなどの収録機材の位置によって大きく制約を受けるコンテンツである。こうした課題を解決するため、本研究では、クラシックコンサートとジャズセッションのイベントを収録し、インタラクティブに自由視聴点での三次元映像音声を再生するアプリケーション「SDM360²」を設計、実装した。SDM360²を95人の被験者に実際に体験してもらい、インタラクティブ3Dコンテンツの自由視聴点再生の有効性を検証した。さらにビルボードジャパンが開催した2017年Live Music HackasongでSDM360²のデモンストラーションを行い、審査員と一般の来場者の投票により、優秀賞を受賞した。

キーワード：映像音声メディア, 3Dコンテンツ, media networking, software defined media

SDM360²: An Interactive 3D Audio-visual Service with a Free-view-listen Point

MANABU TSUKADA^{1,a)} YU KOMOHARA¹ TAKASHI KASUYA^{1,2} HIDEAKI NII³ SHIGEKI TAKASAKA⁴
KEIKO OGAWA⁵ HIROSHI ESAKI¹

Received: April 7, 2018, Accepted: May 30, 2018

Abstract: Various audio-visual service based on Internet are deployed these days widely. Among these, object-based audio-visual services are getting more critical. We started Software Defined Media (SDM) consortium to investigate object-based audio-visual services and Internet-based audio-visual since 2014. The placement of microphone and camera limits the audience to watch at the free viewpoint of the contents of the package media such as DVD. In the study, we designed and implemented the system of interactive 3D audio-visual service with a free-view-listen point, named SDM360². 95 persons experienced SDM360² and answered the questionnaire in subjective evaluation. We also demonstrated the system in “Live Music Hackasong 2017” hosted by Billboard Japan. We received the second prize based on the vote of the judges and the audience.

Keywords: audio-visual media, 3D contents, media networking, software defined media

¹ 東京大学
The University of Tokyo, Bunkyo, Tokyo 113-8654, Japan
² 株式会社竹中工務店
Takenaka Cooperation, Koto, Tokyo 136-0075, Japan
³ VIVITA 株式会社
VIVITA, Inc., Minato, Tokyo 107-0061, Japan
⁴ エスイーディー株式会社
Sophisticated Engineers Department (SED), Minato, Tokyo
108-0075, Japan

1. はじめに

近年、スマートフォンが普及し、標準的に内蔵される映像音声の収録機能によって様々なイベントが収録される場

⁵ 慶應義塾大学
Keio University, Minato, Tokyo 108-8345, Japan
a) tsukada@hongo.wide.ad.jp

面が増えている。それにともない、収録されたコンテンツが、YouTube やニコニコ動画などの動画共有サービスで共有されるという、インターネットを利用して、伝達・共有・加工が行われる状況ができあがった。さらには、収録対象から映像素子に入力されたビットマップ情報とステレオマイクに入力された2チャンネルの音声情報をそのまま転送するだけでなく、空間に存在する収録対象を3次元モデルとして解釈し、複数の視聴オブジェクトに分解して伝送するオブジェクト指向の方式が注目を集めている。これにより、ヘッドマウントディスプレイ (HMD)、3D テレビ、立体音響装置などの受信側の設備に合わせた柔軟な3次元表現が可能となるだけでなく、他のコンテンツの視聴オブジェクトを別途受信して、組み合わせることで、今までにない表現への可能性が広がる。たとえば、音声と映像をオブジェクト化することで、スポーツ観戦者の要求に応じて応援スタンド側からの視聴や、選手1人称視点からの視聴を切り替えたり、アイドルグループの中のお気に入りの人物を近くで鑑賞したりというインタラクティブなコンテンツの製作が可能になる。

このように、インターネットを前提とした視聴サービスが登場し、なかでも空間に存在する視聴対象を三次元的に解釈するオブジェクトベースの視聴サービスの重要性が増加するなか、著者らは2014年からSoftware Defined Media (SDM) コンソーシアム^{*1}を設立し、オブジェクトベースのデジタルメディアと、インターネットを前提とした視聴空間の研究を続けてきた [1], [2]。

現在、オーケストラやライブなどのイベント収録を、後に視聴するためのパッケージメディアでは通常、あらかじめ決められたカメラとマイクの位置と角度で収録し、その情報が再現可能な方法で記録されていないため、視聴できる位置や方向などは非常に限られたものとなる。これはコンテンツが大幅な制約を受けていると考えられる。こうした課題を解決するため、本研究では、クラシックコンサートとジャズセッションのイベントを収録し、インタラクティブに自由視聴点での3次元映像音声を再生するアプリケーション SDM360² を開発した。

本論文では、まず3章で、本研究の目的を述べる。次に、4章で、慶應義塾大学コレギウムムジクム古楽アカデミーのコンサート収録と、Musilogue Band ジャズセッションの収録に関する構成を説明する。5章では、収録されたデータをもとに開発した、インタラクティブに自由視聴点での3次元映像音声を再生するアプリケーションであるSDM360²の設計と実装を述べる。6章では、SDM360²の性能評価を行い、95人の被験者による主観評価を行った結果を報告する。7章で、「ライブを拡張する」というテーマで行われたビルボードハッカソンで出展したデモンスト

レーションの報告を行う。2章で、関連研究の調査を行う。8章で、本論文の結論と今後の課題について述べる。

2. 関連研究

音響の録音・再生システムは、チャンネルベース、オブジェクトベース、シーンベースの3つに大別されることがある [3]。

チャンネルベースのシステムは、ステレオサウンド (2チャンネル) から始まりサラウンドサウンド (多チャンネル) へと発展し、2016年に試験放送の始まったスーパーハイビジョン [4] では、22.2 マルチチャンネルの立体音響システム [5] を採用している。チャンネルベースのシステムでは、収録においては一般的なマイクを利用できるという利点があるが、最終的に出力する音声の情報をそのまま記録するため、再生環境に合わせたチャンネル数の音声情報を記録しておく必要がある。また、最終出力の形で音声データを記録するため、視聴者の動きに追従した音声の提示することは難しい。

オブジェクトベースのシステムは、音源の音色のデータとその三次元の位置をメタデータとして記録し、再生環境においてスピーカの位置から音場をレンダリングする方式である。たとえば、映画館やホームシアタでの採用が進むDolby Atmos [6] や、AuroMax [7] などがある。また、オブジェクトベースの方式は、国際標準化機構 (ISO) と国際電気標準会議 (IEC) のMoving Picture Experts Group (MPEG) において、MPEG-H [8], [9] の標準化が進んでいる。オブジェクトベースの方式では、音源の位置を記録する必要があるが、収録に使うマイクは一般的なものを利用できるという利点がある。また、再生する音源と視聴者の相対的な位置関係から音場を計算できるため、視聴者の位置移動や頭部の回転に追従した音声の提示が可能である。

シーンベースのシステムは、ある受音点に到来する音を指向性を持った複数のマイクを組み合わせ、全周360度で空間の音全体を録音し、到来する音の方向を再現する技術である。アンビソニックス (Ambisonics) [10] の收音では、アンビソニックマイクという特殊なマイクを利用し、收音したデータはB-フォーマットと呼ばれる信号として記録される。このデータをもとに視聴者の聴取位置における、頭部の回転に追従した音声の提示が可能である。ただし、視聴者の位置移動に追従した音声の提示は難しい。Ricoh Theta V など近年発売された民生用収録機器にもアンビソニックマイクが搭載される例も多く、こうして収録された収録データはYouTube やFacebook の動画共有サイトでも再生に対応している。さらに、アンビソニックでは再現が難しい複数の音源から到来する音を正確に再現する高次アンビソニックス (Higher-order Ambisonics, HOA) [11] の研究も行われている。

音声を空間的に物理的な波面として再現することを目指す研究としては、Huygens-Fresnel の原理や、音波の物理

^{*1} <http://sdm.wide.ad.jp/>

的な伝搬を記述する Kirchhoff-Helmholtz 積分定理に基づき、マイクアレイで観測した音の波面をスピーカアレイで正確に再現することを目指している波面合成 (Wave Field Synthesis) [12] がある。さらに、収録環境に設置したマイクアレイと同じ構成で、再生環境にもマイクアレイを設置する境界音場制御 (Boundary Surface Control, BoSC) [13], [14] の方式の研究がなされている。この方式では、再生環境においてスピーカアレイから到来する音波を、再生側のマイクアレイで観測し、収録側のマイクアレイと同じになるような逆システムを設計することで、物理的な波面の再現を目指している。一般的に、空間的・物理的に音波を再現する方式は収録・再生システムが大型化し複雑になる。

映像収録技術においては、4K や 8K などの高精細映像化を進める方向と平行して、収録対象の三次元解釈を取り入れる方式が登場している。たとえば、視聴者が任意に選択する視点に合わせて三次元の空間を表示することが可能な自由視点映像の合成技術が注目されている [15], [16]。特に、サッカーのような屋外大空間で撮影される映像を対象とした合成方式として、単純化三次元モデルを用いる手法が提案されている [17], [18]。

映像音声の遠隔配信は IP ネットワークの高速化にともない、4K60p 映像を複数同時に配信する実証実験が開始されている。高精度な映像音声の双方向での遠隔配信を行うことで、遠隔コラボレーションシステムを構築すると、立体感や情感など、臨場感を高めることができる [19]。さらに、同様の技術を利用してデジタルシネマ劇場へのライブ配信することで、劇場をパブリックビューイングの会場として転用することが可能になる。これにより、同じく高精度化するホームシアタから劇場を差別化し、劇場施設の提供できる価値を大きく向上させると考えられ、ビジネス的な展開が期待されており、商用実証実験が行われた [20]。

遠隔のユーザ同士が快適に会話し、効率的に協調作業するためのコミュニケーションシステムの開発が進んでいる。NTT では、同じ部屋にいる感覚を同室感と名づけ、同室感コミュニケーションシステム「t-Room」の開発した [21], [22]。t-room では、ユーザ全員が周囲の音や映像に関して同じ認識や知覚を対称的に共有することを目指し、同じ大きさの部屋に複数の背面スクリーンを囲い込んで設置している。

バーチャルリアリティの視覚ディスプレイとしては CAVE [23] や多面型全天周ディスプレイ (CABIN) [24] をはじめとする没入型多面ディスプレイの研究が行われてきた。こうした 4 面の壁と地面を含めた没入多面ディスプレイは HMD とは違い、その場にいる多人数のユーザに同時に仮想世界の体験を提供できる。したがって、こうしたディスプレイに追従する音響の提供はヘッドホンではなく、複数のスピーカによるインタラクティブ高臨場感音場提示

手法が検討されてきた [25]。また、NHK 放送技術研究所は、2 眼立体の三次元コンピュータグラフィックス再生と、映像に追従してインタラクティブに三次元の音場をスピーカアレイを用いて再生するシステム [26] を組み合わせ、インタラクティブ 3D 映像音響再生システムを開発した [27]。

オープンデータのメタデータ記述手法として Resource Description Framework (RDF) が広く利用され、それらのデータが URI で相互参照され構築されるデータのネットワークは Linked Open Data (LOD) と呼ばれる。RDF や LOD は、現在、たとえば、人文社会系大規模データベース [28] や、ロックアウトマウスの表現型のデータベース [29] に使われ、データの横断的な利用を促進している。さらに、データ間の関係を示す語彙やデータの種類 (クラス) を表す語彙は RDF Vocabulary と呼び、音楽データを記述する RDF Vocabulary は Music ontology [30] として定義されて、英国放送協会 (BBC) などでも広く利用されている。

3. 本研究の目的

本研究の目的は、音楽イベントの収録データを利用して、自由視点映像音声のインタラクティブな再生を行うことである。2 章に述べたように、収録・再生を行う機器が大規模・複雑になる物理的な音波を再現する立体音響方式は採用せず、より簡易な機器を用いて、視聴者のインタラクティブな動作に応じて聞こえる音声に変化する体験によって、立体音響を提供することを目指す。その目的のため、既存のチャンネルベースのシステムではなく、ソフトウェアの制御によりインタラクティブな体験を提供するのに適していると考えられるオブジェクトベースの立体音響を採用する。本研究では、オブジェクトベースの収録データを利用した様々な演出を可能にするプラットフォームの要求事項を抽出するため、SDM360² を試作する。

こうしたプラットフォームにより、たとえば、ある映像が、ある奏者にズームアップするとその奏者が奏でる音も連動してズームアップするなどのインタラクティブなコンテンツを簡単に作成することが可能になる。図 1 は、プロ用の機器のほか、観客が持ち込む携帯デバイスによって収録されたデータから三次元映像音声コンテンツを協調制作するフレームワークを模式的に示したものである。

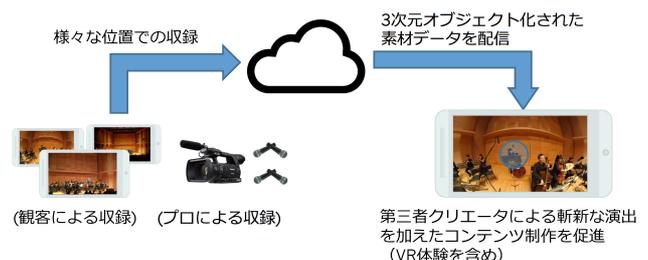


図 1 三次元映像音声コンテンツの協調制作フレームワーク
Fig. 1 Framework of collaborative 3D contents creation.

SDM360²と目指すプラットフォームは以下の要件を想定する。

三次元の映像・音声の演出をソフトウェアで制御：収録対象から映像素子に入力されたビットマップ情報と、ステレオマイクに入力された2チャンネルの音声情報としてそのまま伝送する従来方式では、映像音声の演出を制御することは難しいため、SDMでは、視聴オブジェクトを三次元表現を持った情報空間上で管理しながら、ソフトウェアの演出によって再生環境に適した映像と音響を作り出す。

ユーザ・インタラクション性：SDMでは、視聴者の関心事を再生側のソフトウェアシステムにフィードバックし、演出の制御や、視聴者の望む映像音声のソースを選択可能にする。現在のコンテンツ視聴においては、放送局において制作されたコンテンツを受け取り再生するため、視聴者が特定の関心事に対して、能動的にズームアップしたり、違うアングルから視聴を行ったりすることはできない。SDMでは、視聴者の関心事を再生側のソフトウェアシステムにフィードバックし、演出の制御や、視聴者の望む映像音声のソースを選択可能にする。ユースケースとしては、ファンの歌手を目の前にできる位置からの映像音楽の視聴などが考えられる。

ソフトウェアレンダリングによる拡張演出：SDMで交換される視聴オブジェクトは実空間由来のものである必要はなく、コンピュータを使って設計された映像音声による効果を組み合わせることも可能であり、それによって視聴者の臨場感や満足度を高めることが期待される。これはコンピュータを使って設計された映像音声による効果をソフトウェアレンダリングによって組み合わせた拡張演出と考えられる。たとえば実際のイベント会場では、見たり聞こえたりしないような映像音声の効果を追加することで、現実拡張を行うといったことが想定できる。

映像・音声の複数ソースのミキシング可能性：放送側が作成した映像音声のコンテンツを単一のソースからの受信だけでなく、複数のソースからコンテンツ素材の受信を可能にし、ソフトウェアによる受信側でのミキシングを実現する。さらに、複数ソースのコンテンツをミキシングする主体は、放送側と受信側の中間に位置するクラウドでの共同でのミキシングを行うことも想定した仕組みを構築する。これにより、創造性豊かなクリエイタの手によるコンテンツが、第三者クリエイタらの共同作業によって進化し、視聴者の好みによって選択される視聴環境を作る。

再利用性の高いデータ公開方式：収録されたメディアはアプリケーションに非依存に、複数の視聴オブジェクトなどに分解して解釈し、管理する機構が必要である。本研究ではSDM360²を開発することで、三次元の映像・音

声を利用したプラットフォームの要求に応えるデータベースの設計に関する知見を得る。

4. 素材データの収録

SDM360²の開発や、その他の3次元映像音声のアプリケーション開発に用いるため、収録データ素材として、慶應義塾大学コレギウムムジクム古楽アカデミーのコンサートと、Musilogue Band ジャズセッションを収録した。

4.1 慶應義塾大学コレギウムムジクム古楽アカデミーのコンサート収録

2016年1月10日、慶應義塾大学日吉キャンパス内に建つ509席の席数を有する藤原洋記念ホールで開催された慶應義塾大学コレギウムムジクム古楽アカデミーのコンサートを収録した。収録対象は、17世紀のドイツ宮廷音楽であるヨハン・フリードリヒ・ファッシュの「管弦楽組曲ト長調 FaWV K: G2」など、最大24人の演奏家によるすべてアコースティックの演奏である。楽器には、テオルベ、チェンバロ、バロック・ヴァイオリン、バロック・オーボエ、ヴィオラ・ダ・ガンバなどの現代では珍しい楽器が含まれており、楽器ごとの演奏と各楽器の合わさった音を分離して収録した [31]。

図2にカメラとマイクの配置を示す。メインマイクロホンとして単一指向性のSchoeps MK4+CMC6をプロセニウム（客席からみて舞台を額縁のように区切る構造物）付近のステージバトンに左右それぞれ舞台中央から2mの位置に吊るした。アンビエントマイクとしては、無指向性のDPA4006を左右のサイドバルコニーにそれぞれ1つずつ設置するとともに、主階席の最後席にORTFマイクを設置した。これらはすべて録音セッションを指示したプロの録音エンジニアの判断のもと、標準的な室内楽オーケストラ録音用のマイクとして、最適な位置を選んで配置された。舞台内にはセクションごとと同じく録音エンジニアの判断のもと標準的な個別録音として最適な位置を選び設置したが、それぞれのマイクロホンは単一指向性と双指向性の組合せとなっている。この組合せは従来から使われているM/S型のマイク配置であるため、マイクロホンからの出力に既存のマトリクス処理を施せばその係数を変えることでステレオ集音の広がり感を変化させることができる。また、さらに高度な処理を使うことで、側方成分を利用した主方向成分の効果的な強調といったことも期待できる。

加えて音響的な出力レベルは弱いユニークな楽器であるテオルベ、チェンバロには個別集音のための近接マイクが単一指向性マイクロホンを使って設置されている。ポイントレイマイクロホンとしては、それぞれ6個と8個の単一指向性マイクロホンを利用した、3D XYZ および2D 360度アレイを設置するとともに、Eigenmike™を設置した。また客席エリア内の代表点にはダミーヘッドマイク

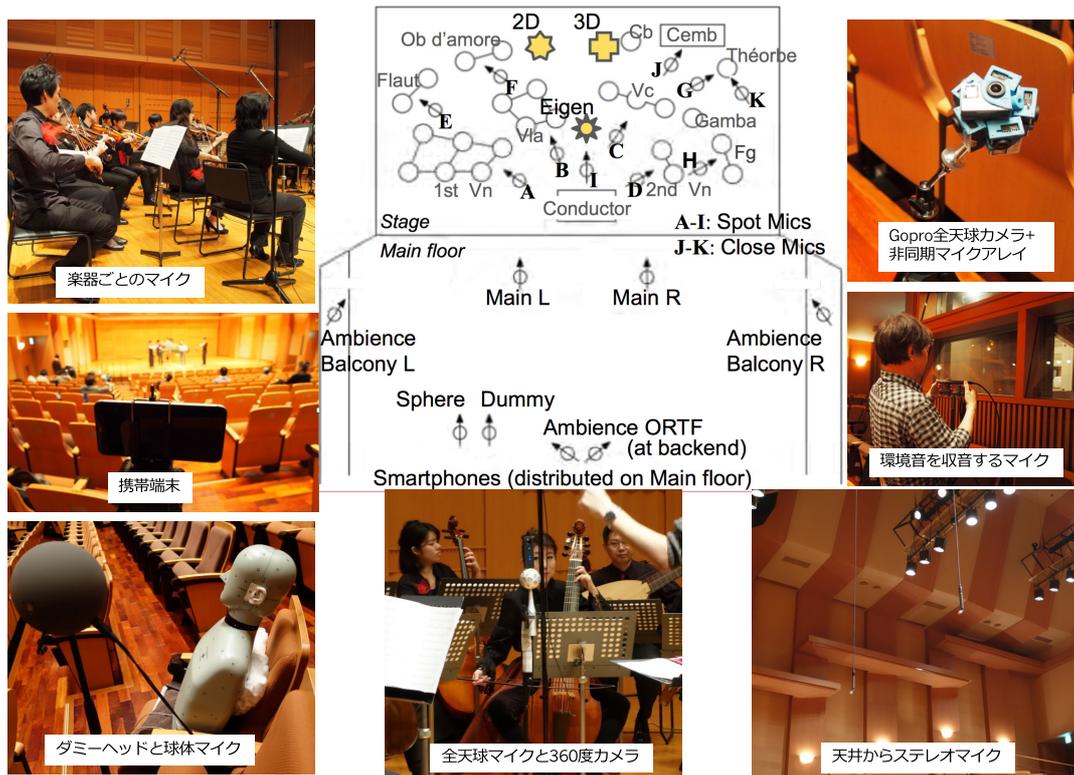


図 2 オーケストラ収録のカメラとマイクの配置

Fig. 2 Camera and microphone layout of orchestra recording.

と球体マイクも設置されている。2D と 3D アレイマイク、Eigenmike, その他のすべてのマイク出力は Dante 付きの Head Amp Unit を介し IP ネットワークで録音ブースに設置された PC 上の Digital Audio Workstation (DAW) に伝送された。

映像収録は、コンサート参加者がふだんは目にすることができない地点からの映像音声を活用することを狙い、舞台内の 2 点において、GoPro™ を 6 台組み合わせた 360 度動画と、Ricoh Theta™ S を使った 360 度動画収録を行った。さらに、コンサート参加者が持ち込む携帯端末による収録を活用できるか検証することを意図して、iPhone を 1 階席の近距離、中距離、遠距離に合計 3 点にスタンド上に取り付けて、映像音声の収録を行った。また同様に、2 階席にも 1 点 iPhone を設置した。さらに、また 1 階後方に設置したカメラで 4K Video も同時に撮影している。これらの映像音声の収録は、ローカルのストレージに記録することで行った。

4.2 Musilogue Band ジャズセッションの収録

2017 年 1 月 26 日に六本木ミッドタウン内にある Billboard Live Tokyo で Musilogue Band のコンサートを収録した。Billboard Live Tokyo のフロアは 3 層あり、3 階・4 階に位置するテーブル席、ソファ席、主に 5 階に位置するカジュアルな席を合わせて、300 人の観客を収容できる。リハーサル中に録音した映像、音声のコンテンツを格納し、

SDM360² のコンテンツを完成させ、収録から約 4 時間後には Live Music Hackasong での最終発表を行った。

図 3 に、バンドの編成、360 度カメラの設置場所、マイクの設置場所を示す。

バンド編成は、藤原佑介、藤谷一郎、金子巧と 3 人のアーティストによる、Drums, Electric Bass, Keyboard の 3 種の楽器演奏である。会場のリスナへの表現手法としてはすべての楽器にマイクあるいは電気信号を増幅する機材に接続し、Sound Reinforcement (SR) 用ミキシングコンソールで調整されたのちに大規模なスピーカで再生される。音圧レベルは場所にもよるがおおよそ 100dB SPL となる。

舞台内のマイク構成については楽器単体のほかに Audience マイクを 4 本設置する。音源よりの近いポジションとしてステージ袖の上手、下手に 1 本ずつ、もう 2 本は天井から無指向性のマイク DPA4090 を 2 本吊りさげでの構成とした。音源よりの近いマイクはステージの角の位置、高さ 1m くらいのところに設置する。音源より遠いマイクについてはステージから奥に 5m、高さ 5m のところにセンタを中心に約 1m の幅で 2 本設置した。

音声はすべて SR 用ミキシングコンソールにまとめられおり、コンソールの Head Amp 分岐をデジタルで DAW に転送するシステムとなっている。録音される音量レベルはすべてミキシングコンソールの設定に依存するため、後に整音作業が必須となる。

録音したデータを編集用の DAW で DrumsMix, Bass-

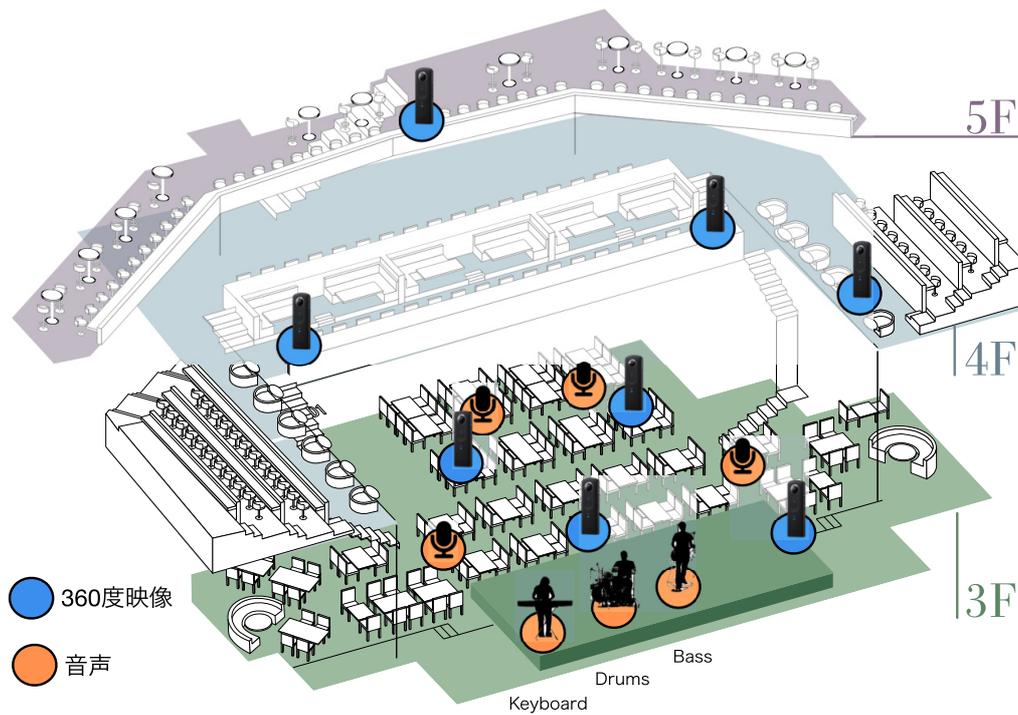


図 3 ビルボードライブ収録のカメラとマイクの配置
 Fig. 3 Camera and microphone layout for billboard recording.

Mix, KeyboardMix それぞれ楽器ごとの単体でまとめあげた音源に編集し、さらに観客席の左奥, 右奥, 左手前, 右手前のアンビエンス Mix を編集した。

さらに、映像データを収録するにあたりリコー社製品の全天球カメラ「Ricoh Theta S」を、図 3 に示す場所に、8 台配置した。ステージと同じ階である 3 階には、ステージ目の前、ステージから見て右手、観客席右手、観客席左手に 4 つ配置した。4 階の観客席はステージを取り巻くようにコの字型になっており、ステージから見て最端右手、右手、左手に 3 つ配置した。5 階に設置した Ricoh Theta S は全景を見渡せるよう、中央に配置した。

これらの 7 つの音源と、8 つの映像をデータ素材として、SDM360² アプリケーションで利用する。

5. SDM360²: インタラクティブ 3D コンテンツの自由視聴点再生

SDM コンソーシアムでは 3 章で述べた本研究の目的と要求事項を基に、インタラクティブに自由視聴点での三次元映像音声を再生するアプリケーション SDM360² を開発した [32]。SDM360² とは、映像と音声を全周 360 度で視聴できることから命名した。

現在、SDM360² では、収録した慶應義塾大学のクラシックコンサートと、ビルボードライブ・ジャパンで収録したジャズのセッションを視聴できる。

ユーザは、収録された演奏が行われている三次元空間をステージと観客席を含めインタラクティブに自由に移動で

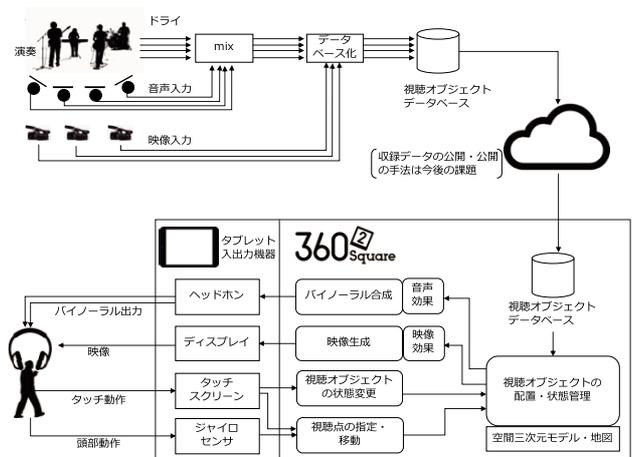


図 4 SDM360² の設計
 Fig. 4 System design of SDM360².

き、SDM360² はその地点で見える映像と聞こえる音声を再生する。また、演奏されている楽器や演奏パートを音声にズームインしたり、それらを自由に有効化・無効化したりすることができる。

5.1 設計

SDM360² の全体の設計概要を図 4 に示す。前章では、図 4 上部の、その楽器から発せられる生の音源であるドライ音源と、空气中を伝わる様々な音が混ざった音声であるマイク収録を、SR 用ミキシングコンソールで調整したものである音声データと、多種のカメラで撮影した動画デー

表 1 SDM360² で利用したデータ
Table 1 Data used in SDM360².

	コンサート (2016)	ジャズセッション (2017)
時間	約 1 分 30 秒	約 3 分
映像	360 度動画 1 つと iPhone 動画が 4 つ (合計約 400 MB)	360 度動画を合計 8 つ (合計約 1010 MB)
音声	wav ファイル 8 つ (約 100 MB)	wav ファイル 7 つ (約 22 MB)

データを保存する処理を解説した。

SDM360² で利用したデータは、4 章で解説した収録データの一部を切り取り利用した。データの概要は表 1 のとおりである。収録データをインターネット上で公開、共有する手法については今後の課題であり、最終章に記述した。よって本章では、SDM360² と収録データがインストール済みのタブレット端末における、SDM360² の映像音声の処理とユーザインタフェースを解説する (図 4 の下部)。

5.1.1 インタラクティブ映像音声の処理

視聴オブジェクトデータベースから取り出した映像と音声のメディアは、それに付随するメタデータをもとにあらかじめ作成しておいた空間三次元モデルの上に配置される。配置された視聴オブジェクトは状態を持ち、自由に有効化・無効化できるように管理される。

また、その 3 次元モデルの中に単一の視聴点を定義して、その場所で聞こえる音をバイノーラル合成する。その際、音の減衰率やリバーブ効果をつけるなどの処理を追加可能である。また、移動速度によって音声のドップラー効果を付与する。今回のコンテンツでは、最大で約 20m を 1 秒で移動 (時速 72km) という、実際には体験が難しい移動速度での視聴を体験できる。同様に、その場所で見える映像を生成しディスプレイで再生する。映像にも、音声可視化などの実際には収録されてないものを仮想的に追加するなどの処理を追加できる。

ユーザのタッチ動作のインタラクションより、視聴オブジェクトの状態を有効化・無効化でき、さらに視聴点を自由に移動したりできる。また、頭部動作により向いた方向に視聴方向を移動できる。

5.1.2 ユーザインタフェース

SDM360² は、HMD での鑑賞に応用可能であるが、現在は簡易なインタラクティブ性を重視してタブレットの入出力機器を想定して開発する。

ユーザは、ヘッドホンからバイノーラル出力を受け両耳で音を聴取し、ディスプレイにより映像を視認する。

一方、ユーザからのインタラクションは、表 2 に示す、スクリーンのタッチ動作で行う。視聴点自体を移動させる通常モードと、位置は固定し視線を移動させる位置固定モードによって、タッチ動作による結果が異なるが、それぞれ 1 指ドラッグ、2 指回転、2 指ピンチをサポートする。

表 2 タッチ操作と動作
Table 2 Touch gesture and action.

動作	通常モード	位置固定モード
1 指ドラッグ	ドラッグ方向へ視聴点の位置を移動	視聴点の位置を中心にドラッグ方向へ視線を移動
2 指回転	視聴点の位置を中心に視線を水平方向に回転	視線を中心に回転
2 指ピンチ	ズームイン・ズームアウト	



図 5 SDM360² のヘルプ画面
Fig. 5 Help display of SDM360².

また、視聴点の移動に関しては、画面につねに表示されるボタン UI を設置し、プリセットされた位置と視線に移動できるように設計する。さらに視線移動に関しては、タブレットのジャイロセンサを使い、タブレットの向いている方向を取得して、タブレットの背中方向へ視線移動できるモードを用意する。

5.2 実装

SDM360² は、Unity Version 2017.1.1f1 を用いて開発した。

本アプリケーションは起動すると SDM360² ロゴが現れ、コンテンツ選択画面へ遷移する。現在選択可能なコンテンツは、慶應義塾大学のクラシックコンサートと、ビルボードライブ・ジャパンでのジャズのセッションである。

クラシックコンサートを選択すると図 5 に示すクリーンショットのようなヘルプ画面に遷移して、使い方を提示する。ユーザは使い方を理解したのち、「Start」ボタンを押下することにより、コンテンツ視聴が開始する。

コンテンツ視聴は、図 6 に示すクリーンショットのような画面で行われる。図 6 は図中で音符を描画した立方体で示される音声オブジェクトの配置が見やすい視聴点に移動したときの映像である。実際に複数の場所で収録された音声は、3 次元モデル上に配置され、Unity の Audio Listener と呼ばれるコンポーネントを利用して、その聴取点で聞こえるようバイノーラル合成されている。このコンポー



図 6 通常モードのスクリーンショット
Fig. 6 Screenshot of normal mode.

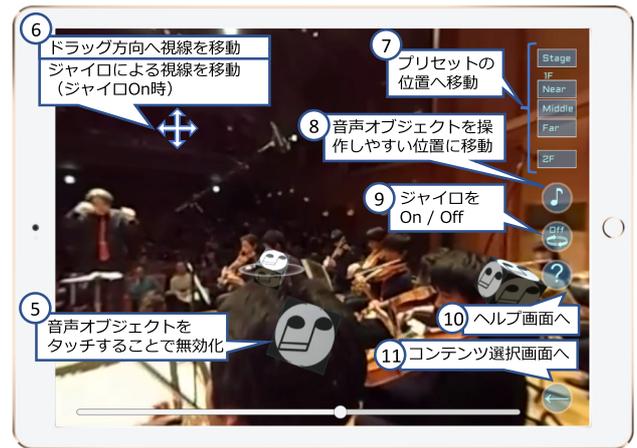


図 7 位置固定モードのスクリーンショット
Fig. 7 Screenshot of fixed mode.

ネットは、聴取者が主に左右の音像定位を行えるように、両耳間の音の到着時間の差 (Interaural Time Difference, ITD) や音量の差 (Interaural Level Difference, ILD) を利用している。さらに、前後上下の音像定位を行えるように、平均的な頭部伝達関数 (Head-Related Transfer function, HRTF) を用いたバイノーラル合成を行っている。図中 (1) で示されるように、実際には見えない音を音量に大きさの変化する白い輪によって可視化する映像効果を追加した。また、図中 (2) で示されるとおり音声オブジェクトにタッチすることで、その音声のパートの有効化・無効化することができ、無効化された音声オブジェクトは回転が止まり、音声可視化の白い輪が消える。図中 (3) に示すとおり、各音声オブジェクトにはメタデータが付属しており、タッチによってそのメタデータが詳細表示に切り替わる。また、図中 (4) の映像オブジェクトをタッチすることで、視聴点はその映像オブジェクトの位置へ移動し、位置固定モードに切り替わる。

位置固定モードのスクリーンショットを図 7 に示す。位置固定モードでは、Ricoh Theta S で撮影された 360 度動画を、球体内部に投影して 360 度ビューを実現している。これは、Easy Movie Texture (v3.56) が提供する MediaPlayerCtrl スクリプトによって実現している。位置固定モードにおいても、同様にタッチ動作により音声オブジェクトの有効化・無効化が可能である (図中 (5))。また、図中 (6) に示すとおり、ドラッグによって視線を全周に移動させることができる。また、ジャイロによる視線移動を有効にしている場合 (図中 (9) のボタン UI)、タブレットの背中方向へ視線を移動する。図中 (7)~(11) は、通常モードと位置固定モードの両方につねに表示されるボタン UI の説明である。図中 (7) のボタン群は、タッチ動作によりステージや客席などのプリセット位置へと移動するためのボタンである。図中 (8) は、音声オブジェクトを操作しやすいステージ上の視聴点へ移動するボタンであり、スク

リーンショットを図 6 に示した。図中 (10), (11) のボタンにより、ユーザはいつでも自由にヘルプ画面を参照し、コンテンツ選択画面へ遷移することができる。

6. SDM360² の評価

本章では前章で解説した SDM360² の実装を用いて実験による性能評価を行う。6.1 節では Unity のプロファイラを利用し、インタラクティブ自由視聴点の再生がシステムに与える負荷を検証する。6.2 節では、デモンストレーション展示を行った際、来場者からのアンケート回答をもとに主観評価を行う。

6.1 システムの性能評価

本節では前節で解説した実装を用いて実験による性能評価を行う。実験で用いた機器は iOS 11.2.6 がインストールされた Apple iPad Pro 10.5 インチ (2017) である。

実験は、iPad Pro で、SDM360² を実行し、タッチ操作を行い、各イベントのタイムスタンプを記録することでイベント発生時の Unity のプロファイラの分析結果を確認することで行った。コンテンツは、システムへの負荷が高い、8 個の 360 度動画と、7 個のオーディオ音源を含むビルボードでのジャズセッションを選択した。Unity のプロファイラは CPU 使用率、レンダリング、メモリ、オーディオの 4 種類の結果を報告する。大きく影響と思われる GPU については、該当機器では情報を取得できなかったため、GPU 使用率の分析は今後の課題である。

図 8 にビルボードでのジャズセッションの視聴を行った際の CPU 使用率、レンダリング、メモリ、オーディオの変動を示す。X 軸は、コンテンツ開始からのフレームの番号である。変動は、すべての操作イベントの発生ではなく、視聴点移動するイベントの影響が多かったため、図 8 では、影響の大きかったイベントのみを表示した。

全体として測定した値は、それぞれ図 5~7 で示した、

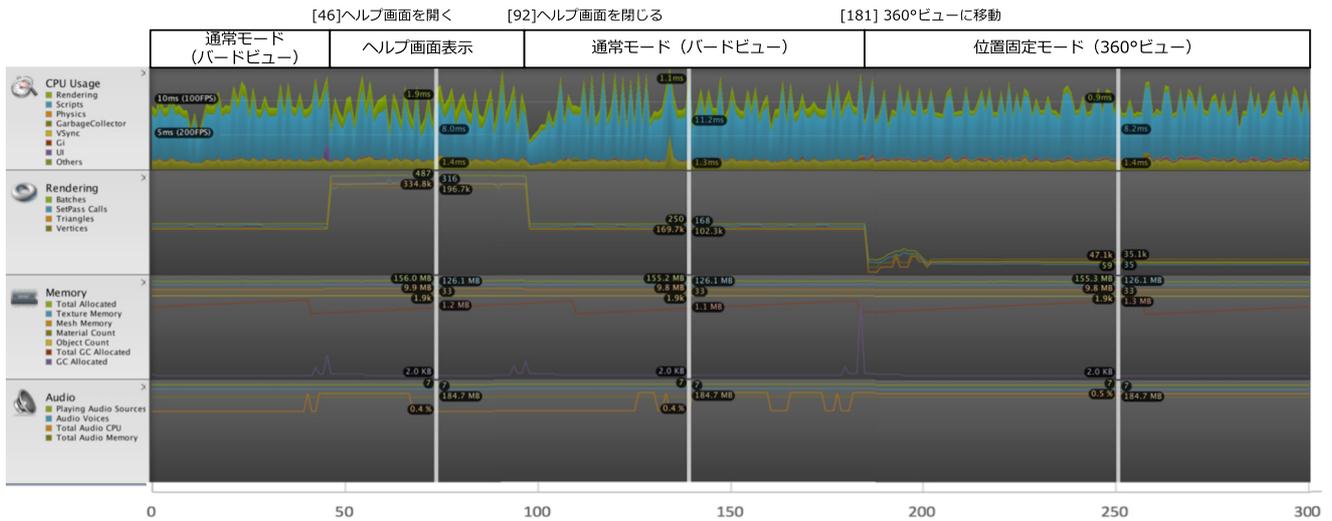


図 8 SDM360² の性能評価

Fig. 8 System measurement of SDM360².

表 3 各モードにおけるレンダリングの数

Table 3 Rendering in each mode.

	ヘルプ画面	通常モード	位置固定モード
Batches 数	487	250	59
SetPass calls 数	316	168	35
三角メッシュ数	334.8 k	169.7 k	47.1 k
頂点数	196.7 k	102.3 k	35.1 k

ヘルプ画面、通常モード、位置固定モードの3つの傾向のカテゴリに分かれた。図8のグラフでは、それぞれの代表的な値を示す位置にY軸の値を示している。

CPU利用率では、全体としておおむね100フレーム/秒を維持していることが分かる。全体として最もCPU時間を消費しているプロセスは、360度動画を球体内面に貼り付けるMediaPlayerCtrlスクリプトであり、8つの360度動画の貼り付けを行うためのCPU時間が全体の66.5%を占めている。図8では、MediaPlayerCtrlスクリプトの処理が大半を占めるスクリプト処理が、ヘルプ画面、通常モード、位置固定モードの代表的な値それぞれで、8.0ms, 11.2ms, 8.2msの時間がかかっている。

レンダリングでは、最終的なDrawCall数を意味するBatches数、マテリアルへのDrawCall数を意味するSetPass calls数、三角メッシュ数、頂点数を示している。表3に示すとおり、すべての値が、ヘルプ画面、通常モード、位置固定モードの順に減少していることが分かる。

メモリ使用量は、全体を通して小刻みな増減を繰り返すものの、表示しているモードによる違いは見られなかった。メモリ使用量は、実験を通して全体で155MBほどであった。内訳としては、ネイティブコードへのメモリ割当て量(Unity)が約82MB、全体のヒープサイズとマネージドコードが使用しているヒープサイズ(Mono)が1.2MBほ

どであった。また、ドライバがTexture、レンダリングのターゲット、Shader、Meshデータに使用している推定メモリ量(GfxDriver)が、約69MB、プロファイラデータに使用されているメモリ量(Profiler)が約3.7MBであった。

オーディオは、実験を通してほぼ変動しなかった。実験を通じて7つの音源が再生されていた。これは、ユーザ操作によって無効化された音源も、有効化したときに演奏のずれがないように停止することなくミュートしていたためである。オーディオドライバが使用している推定メモリ量(FMOD)が約185MB、CPU利用率は約0.5%であった。

本節では、SDM360²を動作させたiPad ProでのCPU使用量、メモリ使用量、レンダリング、オーディオの統計を調べた。360度動画の映像8つ、オーディオ7つのコンテンツをインタラクティブに自由視聴点において、問題なく再生を行うことができた。

6.2 アプリケーションの主観評価

6.2.1 実験方法

主観評価では、被験者がSDM360²を利用することで、インタラクティブな3次元コンテンツの視聴体験を得られたかを確認することを目的に、アンケート調査を行った。アンケートを利用し、ソフトウェアで演出された3次元の映像音声から立体感を得られたか、映像と音声の追従性は十分だったか、インタラクティブ性は十分だったか、今回取り入れた拡張的な演出(特定の音声を有効・無効化できる機能、音声可視化機能)が意図していたとおりに受け取られていたか、を調査した。

アンケート調査は、2017年11月3日、4日に2日間にわたって開催されたKMDフォーラムでデモンストレーション展示を行った際、来場者がSDM360²を体験した後、アンケートに回答してもらうことで行った。展示は、

SDM360²をインストールしたタブレット (Apple iPad Pro 10.5 (2017)) に、ヘッドホン (Sony WH-1000XM2) を有線接続する構成で、それぞれ3台用意して行った。

来場者に対し、SDM360²のボタンUIやタッチ操作に対する利用の仕方を説明し、実際にやってみせうえで、体験者に自由に操作させ、体験者の意思で止めるまで1つのiPodを占有して視聴してもらった。その後、アンケートへ回答をお願いすることでデータを取得した。

6.2.2 被験者

評価実験はKMDフォーラム2017の来場者からSDM360²を体験した被験者で、男性55人、女性34人、無回答6人の合計95人で行った。また、年齢構成は、10代が5人、20代が46人、30代が23人、40代が11人、50代・60代がそれぞれ4人ずつ、無回答2人という内訳であった。さらに社会人が44人、学生が36人、教員が8人であった。

6.2.3 アンケートの質問項目

アンケート項目は、以下に示す設問Q1からQ7までの7つを、それぞれ1から7までの7段階のリッカート尺度を用いて設定した。それぞれの回答に関して、最低の1、中間の4、最高の7の回答の目安を括弧内に記載した。

- Q1** 映像の立体感は感じられましたか? (1:まったく感じられない, 4:感じられる, 7:とても感じられる)
- Q2** 音声の立体感は感じられましたか? (1:まったく感じられない, 4:感じられる, 7:とても感じられる)
- Q3** 音声は映像の方角と比べて正しい位置で鳴っているように聞こえましたか? (1:まったく聞こえない, 4:違和感に気づくが許せる, 7:聞こえる)
- Q4** 映像が動いたとき、音声も追従して動いたと感じられましたか? (1:まったく追従していない, 4:違和感に気づくが許せる, 7:追従している)
- Q5** インタラクティブな視聴体験の操作は簡単にできましたか? (1:操作性が悪い, 4:操作によりインタラクティブな視聴体験ができた, 7:思いどおり操作して視聴できた)
- Q6** 音量可視化による音声オブジェクトの有効化・無効化は直感的でしたか? (1:説明を聞いても理解できない, 4:説明を聞くと違和感なく理解できる, 7:直感的で理解できる)
- Q7** 音声オブジェクトを有効化・無効化することで個別の楽器の音色を聞くことができましたか? (1:個別に聞くことができなかつた, 4:個別に聞いた, 7:自由に楽器を選んで聞くことができた)

設問Q1とQ2は映像と音声の基本的な立体感を問う設問である。設問Q3とQ4は映像と音声の組合せが正しく一致しているように知覚されるかを問う設問である。設問Q3では静止時の音声の聞こえる方角と位置を問い、Q4では動いたときの映像と音声の追従性についての問いを設定

した。また、設問Q5とQ6はタブレットのタッチ動作によるインタラクティブな視聴に関する問いであり、Q5は一般的な操作の容易さについての問いで、Q6では音声オブジェクトの可視化および操作についての問いを設定した。設問Q7は音声オブジェクトに分解されたオーケストラの個別の楽器の音色が聞いたかを問う。

さらに、アンケートの末尾には、「感想・要望・その他」という自由記載項目を用意して、視聴体験のコメントを得た。

6.2.4 主観評価結果

結果は図9に示すとおりである。グラフはリッカート尺度のグラフデザインの助言[33]に従って作成した。X軸は1から7までの7段階の回答の比率をパーセントで表し、棒グラフの位置は、尺度の中間である4をX軸の0の中央に配置し、左に行くほど低評価、右に行くほど高評価という配置で描画した。また、Y軸は各設問を表示し、それぞれの回答には有効回答数を記載した。

映像の立体感に関する設問Q1では、(4:感じられる)以上の評価とした被験者は、86%以上であり、良好な結果が得られた。通常モードでの映像は、コンサートホールの立体モデルの中を自由に移動できるため立体感が感じられたと考えられる。また、位置固定モードでは、360度動画を球体の内側に貼り付けて表示することで再生しているため、厳密には立体ではないが、ジャイロで任意の方向を見回すことができるため、立体感が感じられたと考えられる。

音声の立体感に関する設問Q2では、(4:感じられる)以上の評価とした被験者は、約97%であり、映像の立体感Q1と比較しても、非常に良好な結果が得られた。その他のコメントの中で、音声の立体感はコンピュータグラフィックスと合わせたときのほうが、360度動画と合わせたときより感じられたという感想があった。

映像と音声の位置の一致に関する設問Q3では、(4:違和感に気づくが許せる)以上の評価とした被験者は94%以上であり、良好な結果が得られた。(4:違和感に気づくが許せる)と回答した被験者は約20%であったが、違和感の原因と考えられるのは、楽器ごとの音源の分離が不十分であったためだと考えられる。アコースティックの楽器では、完全な音源分離は難しく、音源に最も近いマイクにはその音源の音が最も強く入るが他の音も混ざってしまう。そのため、かすかな他の音が本来の方向ではない場所から聞こえてくるため、楽器の音の方向が曖昧になってしまうことが考えられる。コメントでは、「どれがどのパートなのか素人には分かりにくい」との指摘があった。映像に映っている楽器から、どのような音色が鳴るのか推定するのは、難しい可能性がある。被験者からは、アイコンを楽器ごとに変えるという助言を得たが、対応は今後の課題である。その他のコメントでは、「左右の音源定位は左右は分かるが前後は分かりにくかつた」というものがあつたが、耳が

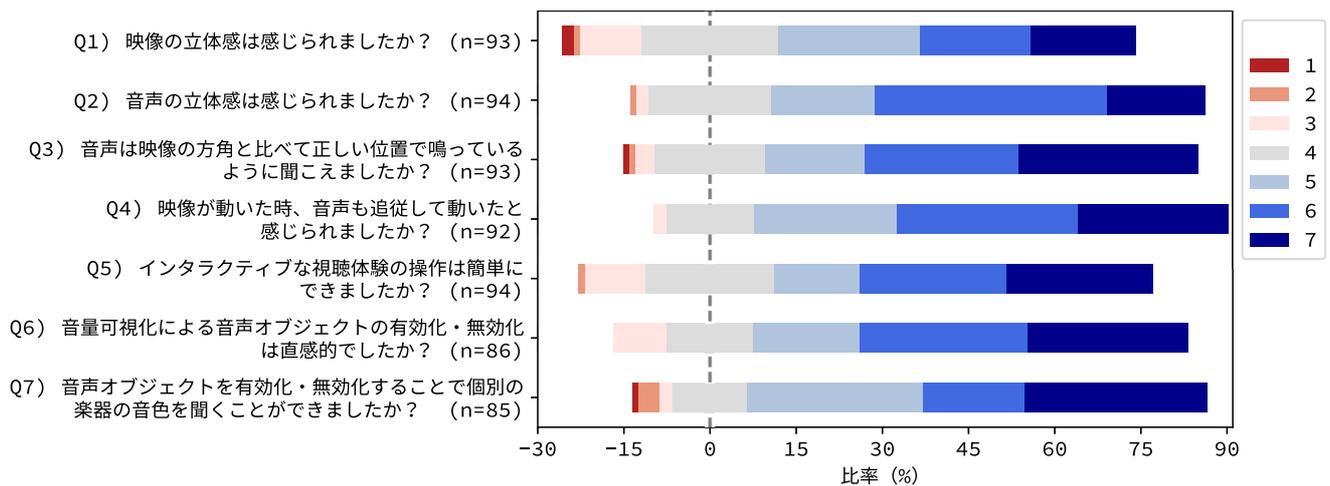


図 9 アンケートの回答比率
Fig. 9 Answer ratio for questionnaire.

左右についている特性上、前後の音は両耳に同じ音量で届いてしまうため分かりにくいという、音源定位の特徴が確認できた。

映像と音声の追従を問う設問 Q4 では、(4:違和感に気づくが許せる) 以上の評価とした被験者は約 98%であり、92 人の被験者の中で、尺度 1 または 2 と回答した者はいなかった。タブレットを左右に動かすことで楽器の音が聞こえる方角が変わり、左右の耳に到達する音の音量と、遅延の差から方向感が割り出せるため、音の追従を認知することができやすいためと思われる。また、その他の欄では、連続的に回転しながら音を聞くのが楽しかったとのコメントを得た。

インタラクティブな視聴体験の操作に関する設問 Q5 では、(4:操作によりインタラクティブな視聴体験ができた) 以上の評価を行った者は 88%以上を占め、良好な結果を得た。しかし、(1:操作性が悪い) という評価をした被験者はいなかったが、2 と評価した者が約 1%、3 が約 10%となり、(4:操作によりインタラクティブな視聴体験ができた) に満たない評価をしたものが 10%以上いた。その場での聞き取りで、ジャイロを利用して視線を移動している最中に音声オブジェクトのタッチ動作を行うと、画面上でオブジェクトが動くため、タッチ動作ではなくドラッグ動作として判定されてしまう場合が見つかった。この場合に音声オブジェクトの有効化・無効化できず、視線移動など意図しない動作をしてしまうことが確認され、操作に対する低評価の要因となった。

音声オブジェクト操作の直感性への設問 Q6 では、尺度 1 と 2 と回答した被験者はおらず、(4:説明を聞くと違和感なく理解できる) 以上の回答が 91%を占めた。音量可視化では、有効化されているときは回転し、無効化すると止まる立体を表示しており、さらに音量によって変化する白い輪を付与していたので、音声オブジェクトの状態が直感

的に理解できたものと思われる。

個別の楽器の音色を聞くことができたかという設問 Q7 では、約 93%が (4:個別に聞けた) 以上の評価を行い、良好な結果を得た。その他のコメントでは、すべての音をいったん無効化してから、聞きたい楽器パートの音声オブジェクトだけ有効化するとそのパートの音が特によく分かるというものがあった。また、特定の音だけ聞くという体験はふだんは難しいので、音声オブジェクトに分解してソフトウェアで再生するというコンセプトの利点がよく分かるというコメントがあった。

主観評価によって、ソフトウェア制御の映像音声で立体感が得られたか、映像と音声の追従性は十分か、インタラクティブに視聴できたか、拡張演出が意図どおり受け取られたかを評価した。7つの設問のうち、5つの設問では 90%以上の被験者が、SDM360² の設計意図どおりに視聴できたことを示す 4 以上の回答を得た。またその他の 2つの設問でも 80%以上の被験者が、4 以上の回答を行った。これにより、おおむね良好に SDM360² の設計意図どおりに 3D コンテンツのインタラクティブ視聴が行えることが確認できた。

7. SDM360² のデモンストレーション

2017 年 1 月 26 日にビルボードライブ東京で「ライブ体験の拡張」をテーマとして行われたハッカソン【LIVE MUSIC HACKASONG】で SDM360² のデモンストレーションを行った。審査員と約 100 人の一般の来場者は、Musilogue Band の実際のジャズ演奏を聴いてから、事前にリハーサル収録してあったデータから、SDM360² で視聴できるというデモンストレーションを行った。また、本イベントは、イベント運営にあたったビルボードジャパンによって LINE live を利用して、スマートフォンおよび PC でライブ視聴できるように放送されており、のべ 3,000

人を超える視聴があった。また、審査員には iPad Pro および iPad air2 を渡して手元で鑑賞できるようにしたほか、一般の来場者には Unity から出力されるステレオ音声出力を舞台の左右に設置されているスピーカから放送し、スクリーンには Unity の Play 画面を表示することで鑑賞できるようにした。

また、最終審査では SDM360² が自由な視聴点から聞きたい音にフォーカスできる点が評価され、審査員と一般の来場者からなる投票で、10 組の中から優秀賞を受賞した。デモンストレーション、および審査・受賞の様子は、様々なメディア^{*2, *3, *4, *5}で掲載された。

8. まとめと今後の課題

本論文では、2014 年より、SDM コンソーシアムで取り組んできた視聴空間サービスのソフトウェア制御に関する研究の中でも、音楽イベントのための自由視聴点映像音声のインタラクティブな再生を行う SDM360² に関して報告した。まず、慶應義塾大学コレギウムムジクム古楽アカデミーのコンサートと Musilogue Band ジャズセッションを収録し、その収録データを利用した 3 次元映像音声を利用した様々な演出を可能とするプラットフォームの要求事項を考察した。要求事項に基づき SDM360² を設計、実装し、性能を評価した。主観評価では、95 人の被験者に SDM360² を視聴後、アンケートに回答してもらい、インタラクティブ 3D コンテンツの自由視聴点再生が意図どおり行われたことを確認した。さらにビルボードジャパンが開催した Live Music Hackasong で SDM360² のデモンストレーションを行い、審査員と一般の来場者の投票により、優秀賞を受賞した。

今後の課題として、本論文ではタブレットを対象デバイスとして選択したが、HMD などの没入感を高めた視聴サービスを検討する。この場合には、適したユーザインタフェースを開発する必要がある。また、3 次元映像音声を利用した様々な演出を可能とするプラットフォームとして、収録データを高い再利用性を持った方式で公開・共有する必要がある。現在、オープンデータのメタデータ記述手法 RDF を利用した LOD として収録データを公開、共有することを検討している [34]。また、収録データをインターネットを利用して配信する、オンデマンド型の方式や、ライブ配信の方式を検討する。今回収録したデータを配信する場合、収録データは 1 分間あたり 330 MB 程度であるため、約 44 Mbps のネットワーク帯域が必要となる。

参考文献

- [1] 塚田 学, 小川景子, 池田雅弘, 曾根卓朗, 丹羽健太, 齊藤翔一郎, 粕谷貴司, 砂原秀樹, 江崎 浩: Software Defined Media: 視聴空間サービスのソフトウェア制御, 日本ソフトウェア科学会学会誌『コンピュータソフトウェア』「ネットワーク技術」特集 (Sep. 2017).
- [2] Tsukada, M., Ogawa, K., Ikeda, M., Sone, T., Niwa, K., Saito, S., Kasuya, T., Sunahara, H. and Esaki, H.: Software Defined Media: Virtualization of Audio-Visual Services, *IEEE International Conference on Communications (ICC2017)*, May 2017, Paris, France (2017).
- [3] ITUR Rec. Itu-r bs. 2051-0 (02/2014) advanced sound system for programme production, Int. Telecommun. Union, Geneva, Switzerland (2014).
- [4] Nakasu, E.: Super hi-vision on the horizon: A future TV system that conveys an enhanced sense of reality and presence, *IEEE Consumer Electronics Magazine*, Vol.1, No.2, pp.36-42 (2012).
- [5] 濱崎公男, 火山浩一郎: 22.2 マルチチャンネル音響システム, 平成 17 年電気学会電子・情報・システム部門大会講演論文集, pp.3-7 (Sep. 2005).
- [6] Dolby Atmos© Specifications, Technical Report Issue 3, Dolby Laboratories (2015).
- [7] AUROMAX© Next generation Immersive Sound system (Nov. 2015).
- [8] Herre, J., Hilpert, J., Kuntz, A. and Plogsties, J.: MPEG-h 3D Audio – the new standard for coding of immersive spatial audio, Vol.9, No.5, pp.770-779 (Aug. 2015).
- [9] Bleidt, R., Borsum, A., Fuchs, H. and Weiss, S.M.: Object-Based audio: Opportunities for improved listening experience and increased listener involvement, *SMPTE Motion Imaging Journal*, Vol.124, No.5, pp.1-13 (2015).
- [10] Gerzon, M.A.: Periphony: With-height sound reproduction, *Journal of the Audio Engineering Society*, Vol.21, No.1, pp.2-10 (1973).
- [11] Poletti, M.A.: Three-dimensional surround sound systems based on spherical harmonics, *Journal of the Audio Engineering Society*, Vol.53, No.11, pp.1004-1025 (2005).
- [12] Berkhout, A.J., de Vries, D. and Vogel, P.: Acoustic control by wave field synthesis, *The Journal of the Acoustical Society of America*, Vol.93, No.5, pp.2764-2778 (1993).
- [13] 伊勢史郎: キルヒホッフ-ヘルムホルツ積分方程式と逆システム理論に基づく音場制御の原理, 日本音響学会誌, Vol.53, No.9, pp.706-713 (1997).
- [14] Ise, S.: A principle of sound field control based on the kirchhoff-helmholtz integral equation and the theory of inverse systems, *Acta Acustica united with Acustica*, Vol.85, No.1, pp.78-87 (1999).
- [15] Kanade, T., Rander, P. and Narayanan, P.J.: Virtualized reality: Constructing virtual worlds from real scenes, *IEEE Multimedia, Immersive Telepresence*, Vol.4, No.1, pp.34-47 (1997).
- [16] 三功浩嗣, 石川彰夫, 内藤 整, 酒澤茂之: 被写体 3 次元形状モデル投影型背景分離方式, 映像情報メディア学会誌, Vol.64, No.11, pp.1685-1697 (2010).
- [17] Koyama, T., Kitahara, I. and Ohta, Y.: Live mixed-reality 3D video in soccer stadium, *The 2nd IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp.178-186 (2003).
- [18] Iwase, S. and Saito, H.: Parallel tracking of all soccer players by integrating detected positions in multiple view

*2 http://www.billboard-japan.com/d_news/detail/46826/2

*3 <https://wirelesswire.jp/2017/02/58944/>

*4 <https://tvfan.kyodo.co.jp/music/news-music/1089676>

*5 <http://www.musicman-net.com/business/64896.html>

images, *Proc. 17th International Conference on Pattern Recognition, ICPR 2004*, Vol.4, pp.751-754 (2004).

[19] 金 順暎, 仲地孝之, 江村 暁, 藤井竜也, 羽田陽一: 4K マルチ映像と 6 チャンネルエコーキャンセラを用いた超高臨場遠隔コラボレーションシステム, 電子情報通信学会技術研究報告, CQ, コミュニケーションクオリティ, Vol.112, No.10, pp.87-92 (2012).

[20] 藤井竜也, 藤井哲郎, 小野定康, 白川千洋, 白井大介: デジタルシネマ劇場へのライブ配信 (ODS) 技術, 電子情報通信学会基礎・境界サイエティ Fundamentals Review, Vol.5, No.1, pp.80-89 (2011).

[21] Hirata, K., Harada, Y., Takada, T., Aoyagi, S., Shirai, Y., Yamashita, N., Kaji, K., Yamato, J. and Nakazawa, K.: t-Room: Next generation video communication system, *2008 IEEE Global Telecommunications Conference (GLOBECOM 2008)*, pp.1-4 (Nov. 2008).

[22] 平田圭二: 未来の電話を考える-遠隔コミュニケーションシステム t-Room (特集コミュニケーション環境の未来に向けた研究最前線), NTT 技術ジャーナル, Vol.19, No.6, pp.10-12 (2007).

[23] Cruz-Neira, C., Sandin, D.J. and DeFanti, T.A.: Surround-screen projection-based virtual reality: The design and implementation of the cave, *Proc. 20th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1993*, New York, NY, USA, pp.135-142, ACM (1993).

[24] 廣瀬通孝, 小木哲朗, 石綿昌平, 山田俊郎: 多面型全天周ディスプレイ (CABIN) の開発とその特性評価, 電子情報通信学会論文誌 D-II, 情報・システム, II-情報処理, Vol.81, No.5, pp.888-896 (1998).

[25] 小木哲朗, 茅原拓朗, 加藤允文, 浅山 宏, 廣瀬通孝: 没入型多面ディスプレイのためのインタラクティブ高臨場感音場提示手法, 日本バーチャルリアリティ学会論文誌, Vol.8, No.1, pp.75-83 (2003).

[26] 大久保洋幸, 大谷眞道, 小野一穂, 正岡顕一郎, 池沢 龍, 小宮山撰, 浅山 宏, 湯山一郎: CG 同期したインタラクティブ音場再生システムについて, 日本バーチャルリアリティ学会論文誌, Vol.5, No.3, pp.965-973 (2000).

[27] 大久保洋幸, 中山靖茂, 池永敏和, 小宮山撰: インタラクティブ 3D 映像音響再生システム, NHK 技研 R&D ('04 [NHK] 技研公開講演・研究発表特集号 (1)), No.86, pp.72-79 (July 2004).

[28] 後藤 真: 人文社会系大規模データベースへの linked data の適用-推論による知識処理, 情報知識学会誌, Vol.25, No.4, pp.291-298 (2015).

[29] Dickinson, M.E., Flenikien, A.M., Ji, X., Teboul, L., Wong, M.D., White, J.K., Meehan, T.F., Weninger, W.J., Westerberg, H., Adissu, H., et al.: High-throughput discovery of novel developmental phenotypes, *Nature* (2016).

[30] Raimond, Y., Abdallah, S.A., Sandler, M.B. and Giasson, F.: The Music Ontology, *Proc. International Conference on Music Information Retrieval*, pp.417-422 (2007).

[31] Ikeda, M., Sone, T., Niwa, K., Saito, S., Tsukada, M. and Esaki, H.: New recording application for software defined media, *Audio Engineering Society Convention Paper, 141st AES Convention*, Los Angeles, USA (Sep. 2016).

[32] 塚田 学, 菰原 裕, 新居英明, 粕谷貴司, 高坂茂樹, 小川景子, 江崎 浩: SDM360²: 音楽イベントのための自由視聴点映像音声のインタラクティブ再生, マルチメディア, 分散, 協調とモバイル (DICOMO2017) シンポジウム (June 2017).

[33] Heiberger, R. and Robbins, N.: Design of diverging

stacked bar charts for likert scales and other applications, *Journal of Statistical Software, Articles*, Vol.57, No.5, pp.1-32 (2014).

[34] 菰原 裕, 塚田 学, 江崎 浩, 曾根卓朗, 池田雅弘, 高坂茂樹, 新 麗, 新 善文: SDM Ontology: Software Defined Media のメタデータ管理のための Ontology, マルチメディア, 分散, 協調とモバイル (DICOMO2017) シンポジウム (June 2017).



塚田 学

2005 年慶應義塾大学環境情報学部卒業。2007 年慶應義塾大学政策・メディア研究科修士取得。2007 年よりフランス・パリ国立高等鉱業学校 (Mines ParisTech) ロボット工学センター博士課程在籍および、フランス国立情報学自動制御研究所 (INRIA) の IMARA チームで研究員として勤務。2011 年博士号取得。現在は、東京大学大学院情報理工学系研究科の特任助教。2014 年より WIDE プロジェクトのボードメンバおよび SDM コンソーシアム・チェア。自動車の情報化等、次世代インターネット IPv6 における移動体通信, SDM に取り組む。



菰原 裕

2016 年東京大学工学部卒業。2018 年東京大学大学院情報理工学系研究科修士課程修了。研究内容は、Software Defined Media データを LOD 化するためのオントロジの構築。学業の傍らマジシャンとしても活動。韓国・中国・台湾・タイ・イタリアといった諸外国の大会で受賞。



粕谷 貴司

2005 年東京農工大学情報コミュニケーション工学科卒業。2007 年慶應義塾大学大学院政策・メディア研究科修了。2008 年竹中工務店入社。ワークプレイスプロデュース本部を経て、2015 年より情報エンジニアリング本部。建物設備に対する情報エンジニアリングの業務や研究に従事。



新居 英明

2003年電気通信大学大学院電気通信学研究科博士後期課程入学。2006年東京大学大学院情報理工学系研究科助手，2007年同学科助教，2009年慶應義塾大学KMD特別研究講師，2010年シンガポール国立大学研究員，2012年(株)IIJイノベーションインスティテュート，2017年よりVIVITA(株)，現在に至る。ネットワークおよびプロジェクトカメラ系インタフェースの研究に従事。博士(工学)。



江崎 浩

1987年九州大学大学院工学研究科電子工学専攻修士課程修了。同年(株)東芝入社。1990年米国ニュージャージー州ベルコア社。1994年コロンビア大学客員研究員。1998年東京大学大型計算機センター助教授。2001年同大学大学院情報理工学系研究科助教授。2015年同大学院同研究科教授，現在に至る。博士(工学，東京大学)。MPLS-JAPAN代表，IPv6普及・高度化推進協議会専務理事，WIDEプロジェクト代表，JPNIC副理事長。



高坂 茂樹

1998年ヤマハ音楽院ミキサー科卒業。1998年ミキサーズ・ラボ入社。長瀬剛，吉田拓郎，浜崎あゆみ，ジャニーズ等のレコーディング事業に従事。2005年エスイーディー有限会社設立，代表取締役就任。IT事業部として通信系，車載系，医療系のソフトウェアテスト，品質管理を行いながら，音響事業部ではコンサートのレコーディングや音声コンテンツ制作等を行う。2017年一般社団法人IT検証産業協会理事就任。2017年エスイーディー有限会社を株式会社へ変更。



小川 景子

2013年慶應義塾大学経済学部卒業。2016年同大学大学院メディアデザイン研究科修士課程修了，修士論文で研究科委員長表彰。グローバルイノベーションデザイン(GID)プログラムで，ロンドンのロイヤル・カレッジ・オブ・アート(RCA)とインペリアル・カレッジ・ロンドン(Imperial)，そしてニューヨークのプラット・インスティテュート(Pratt)へ留学。現在は，慶應義塾大学メディアデザイン研究科研究員。