

第3部

特集3 Software Defined Mediaコンソーシアム2016年活動報告書

SDM WGメンバー

第1章 はじめに

2020年の東京オリンピック・パラリンピックの開催に向けて、世界からの訪問者への高度なおもてなしを目指した社会インフラのスマート化が急がれている。中でも、オリンピックの理念である「スポーツを通して心身を向上させ、さらには文化・国籍など様々な差異を超え、友情、連帯感、フェアプレーの精神をもって理解し合うことで、平和でよりよい世界の実現に貢献する」という目標を支援するため、地球規模での一体感を生み出すような視聴メディアに関する研究開発の重要性が高まっている[1]。そのためには、グローバルなインターネットを前提とした映像・音響の視聴空間と視聴メディアの設計が必要である。

近年は、多くのスマートフォンに標準的に内蔵される収録機器を常時持ち歩くことが増え、収録された映像音声の情報はインターネット上で瞬時に伝達・共有・加工される状況が出来上がった。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としてそのまま伝送し、受信側でそのまま再生するだけでなく、空間に存在する収録対象を三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し、受信側ではこれらのオブジェクトを用いて空間を再合成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システムの構成に合わせた柔軟な三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組み合わせることで、今までにない表現が可能になる。

例えば、音響においては、Dolby Atmos[2]やAuroMax[3]

などの映画館やホームシアター、さらには個人向けモバイル機器を対象に、音のオブジェクトから三次元の音場を生成する立体音響システムが登場している。また、映像においては、複数の地点・角度から撮影された映像・動画から、撮影した空間に存在する三次元オブジェクトの抽出が可能であり、抽出した三次元オブジェクト情報を用いて、任意の視点(自由視点)からの映像の作成・再構築が可能となりつつある[4][5]。こうした流れに目を向けると、今後はインターネットで収録環境と再生環境を双方向で接続し、視聴オブジェクトを交換しながら、三次元表現を持つ情報空間をエッジヘビーコンピューティングまたはクラウドで計算処理することによって、映像音声を作り出されていくことになるであろう。

映像と音声のオブジェクト化が融合することで、従来の配布型コンテンツビジネスを超えた、新しい次元のインタラクティブなオリンピック・パラリンピックの視聴形態などこれまでにないビジネス領域や、これまでにないデジタルメディアを用いた表現方法などを開拓・開花させることが期待される。例えば、音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リアルタイムに応援スタンド側からの視聴や、選手の一人称視点からの視聴を切り替えることが可能になるだろう。さらに、コンサートの視聴においては、複数の歌手の中の一人や、演奏者のうちの一人をズームアップすると映像に追従して音も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制作が可能になる。

さらに映像・音声を作り出す空間に目を向ければ、例えば、劇場、ホームシアター、テレビ会議システム、講義室など、様々な用途に対して専用に設計されたシステムが存在している。これらの映像・音響システムには、收音システム、拡声システム、制作システム、BGMや館内放送

オブジェクト指向のデジタルメディアと、ネイティブ・デジタルなインターネット環境が前提の映像・音響空間を用いたビジネス創造を目指し、2014年1月にSoftware Defined Media(SDM)コンソーシアム^{*1}を設立した。

28 WIDE

SDMは図2.1左に示すように、視聴空間(実空間)に存在する収録対象を三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情報空間に取り込み、視聴オブジェクトの交換など様々な計算処理を行って、その結果を自在に視聴空間に投影することで視聴空間の再合成を行う仕組みである。上記の視聴空間と情報空間をつなぐ処理を行うSDMの基本単位をSDM拠点と呼ぶ。また、SDM拠点のシステムは、図2.1右に示す通り、設備層、サービス層、アプリケーション層に分けられる。

設備層には、収録する空間に存在する映像音声の収録対象を三次元的に解釈しオブジェクト化するためのセンサ、カメラ、マイクなどの設備と、再生空間において視聴オブジェクトから三次元の映像音声を作り出すためのスピーカ、HMD、ディスプレイなどの設備がある。設備例はこれに限らず、これから登場する革新的デバイスの機能も取り込めるよう設備層の差異を吸収できる形で、上位のサービス層を設計する。また、SDM拠点は収録専用・再生専用の空間だけでなく、収録環境と再生環境は同一の実空間上に配置される可能性も想定して設計する。

サービス層は、設備層の設備機能を抽象化しアプリケーションに対してApplication Programming Interface(API)を通じてサービスとして提供するための層である。視聴オブジェクトと管理するための基礎となる空間三次元モデルと地図を定義し、位置測位、音声オブジェクト化、映像オブジェクト化、立体映像再生などのサービスを提供する。APIは開発者が創造的なアプリケーションを実現できるよう柔軟かつ拡張性を持つ。

アプリケーション層は、APIを通じてSDMサービスを利用する任意のアプリケーションが動作する層である。この層にローカルレンダラと呼ぶアプリケーションがあり、インターネット上に存在するリアルタイムとアーカイブのコンテンツを取得してSDMサービスと組み合わせることでSDM拠点の視聴空間に対して視聴者の要望に合わせた映像音声を作り出す。レンダリングの機能は、ローカルだけではなく既存の放送システムやインターネット経由で取得したコンテンツを共同編集するなどの機能を持つパブリックレンダラにより前処理される。また、アプリケーション層に定義されるメディアエンコーダにより

SDM拠点を、他のSDM拠点からは放送局として機能させる。SDM拠点からの放送はリアルタイムに他のSDM拠点に転送されるだけでなく、視聴オブジェクトの蓄積・管理などのコンテンツ管理を行う。SDMでは、放送局から送信されるリアルタイムとアーカイブのコンテンツに対し、標準化されたメタデータや、タグを付与することで、ソフトウェアによって柔軟なコンテンツ管理や、状況に応じたコンテンツ選択を制御する。メディアデータとコンテンツ情報、位置情報などのメタデータを持つ視聴オブジェクトを基本単位として扱う。

第3章 三次元映像音声コンテンツの協調制作フレームワーク

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、カメラとマイクの置かれた位置と角度によって、視聴するコンテンツが大幅な制約を受け、視聴者の意思による自由な場所と角度、距離での視聴は困難である。こうした課題を解決するため、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と考えられる。

こうした機構により自由度の高いシステムが構築可能となり、汎用的な視聴オブジェクト管理によって様々なアプリケーション／コンテンツを様々な場所で提供することが可能となる。この結果、自由度の高いシステムによってより高い忠実度で音の場を収録したり、再生したり、共有したりすることが可能になるということが期待される。例えば、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3.1に示すとおり、収録を行う機器はイベント主催者が関わるプロ用の機器の他に、観客が持ち込む商用デバイスが考えられる。インタラクティブなコンテンツ制作には、両者の収録データを素材として利用できることが望ましい。こうした収録データを自由に流通させることで、第三者クリエイターが斬新な演出を加えたコンテンツを制

作する土壌が整うと考えられる。

したがって、三次元映像音声コンテンツの協調制作フレームワークでは、以下の機能要件を満たすことが望ましい。

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うことはできない。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、ファンの歌手の一人称視点での映像音声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。例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ーの手によるコンテンツが、第三者クリエイター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。

第4章 三次元映像音声メディアのオープンデータ化

4.1 素材データの収録

SDMコンソーシアムでは、試作システムとそのシステムを用いたアプリケーションを開発する一方、アプリケーション非依存な視聴オブジェクトの管理の研究のために、以下の収録を行いデータベースを構築する。本データベースは、実験素材としてコンソーシアムのメンバーに提供される。

4.1.1 慶應義塾大学コレgiumムジクム古楽アカデミーのコンサート収録

2016年01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記念ホールにて開催された慶應義塾大学コレgium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

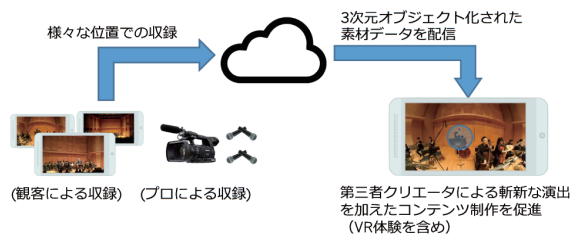


図3.1 三次元映像音声コンテンツの協調制作フレームワーク

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名の演奏家による全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[10]。

図4.1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をプロセニアム(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2メートルの位置に吊るした。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。これらは全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に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んでマイクロホンを設置、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拡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。ポイントア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共に、Eigenmike™を設置した。3Dおよび2Dマイクは、それぞれGoPro™を6台

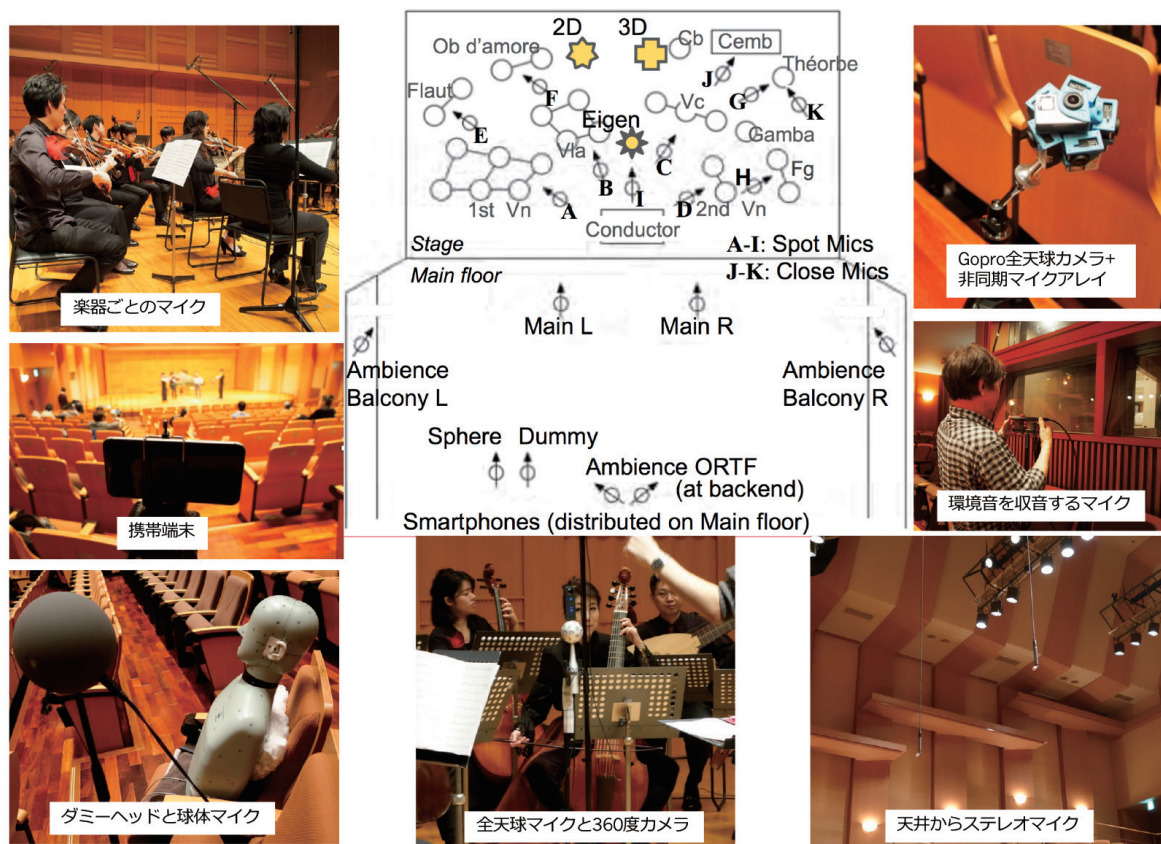


図4.1 オーケストラ収録のカメラとマイクの配置

組み合わせた三次元動画収録, およびRicoh Theta™ Sを使った二次元動画収録との組合せで収録された。また客席エリア内の代表点にはダミーヘッドマイクと球体マイクも設置されている。さらに主階席4点, バルコニー 2点の合計6点にはスタンド上に取り付けたスマートフォンを設置し, これにより録画と録音を行った。2Dと3Dアレイマイク, Eigenマイク, スマートフォンはローカルのストレージに記録し, その他の全てのマイク出力はDante付の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DAWに伝送された。また主階席後方に設置したカメラで4k Videoも同時に撮影している。

収録した映像音声データはSDMによる魅力的なコンテンツ制作とアプリケーション開発に利用するため, 視聴オブジェクトの概念を基礎とするデータベースを構築した。メディアデータとコンテンツ情報, 位置情報などのメタデータを持つ視聴オブジェクトを基本単位として扱う。オブジェクト型の概念を使うことで, マイクで集音した複数の音から, 音声オブジェクト化サービスによってより上位の抽出された音を生成でき, それらに同じオブジェクトの構造を使うことが可能になる。次にこれらのオブジェクトに意味情報を追加すれば, さらに上位層のオブジェクトを想定することができる。このようにこのデータベースには, 下位層から意味情報を付加することで一つ上の層を生成できる再起的な構造を持たせた。さらに例えば従来型録音で典型的に使われるマイク群をひとまとめにしたさらに上位層のオブジェクトなども作成している。このデータベースはこのように階層的にオブジェクトを構成することでその意味的(Semantic)な価値を向上させることを意図している。

今回作成したデータベースでは, GoPro™, RicohTheta™ S, 4Kビデオカメラ, スマートフォンと4種類の映像データを位置情報とコンテンツ情報を付与して管理しているため, コンテンツやアプリケーションを作成する際, 上記の音声データと組み合わせて利用することが容易である。音声データと同様, 映像でも映像オブジェクト化サービスを利用し, 収録対象を解釈したより上位のデータに変換し階層的にデータベースに格納することを想定するが, 上位のデータの作成は今後の課題である。

4.1.2 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BillboardLiveTokyoでムジログプロジェクトのコンサートを収録した。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360Squareアプリケーションを作成して, Live Music Hackasongで発表を行った(詳細は5章, 及び6章)。

編成はElectricPiano(Fender Rhodes), Electric Bass, Drumsのinstrumentalである。会場のリスナーへの表現手法としては全ての楽器にMic或いは電気信号を増幅する機材に接続し, SR(Sound Reinforcement)用MixingConsoleで調整されたのちに大規模なSpeakerで再生される。音圧レベルは場所にもよるがおおよそ100dBsplとなる。

映像データを収録するにあたりリコー社製品の全地球カメラ「RICOH THETA S」を8台配置した。

音声は全てSR(Sound Reinforcement)用Mixing-Consoleに纏められていて, ConsoleのHA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全てMixingConsoleのSettingに依存する為, 後に整音作業が必須となる。

回線は図4.2の通りである。

INPUT PLAN				
No.	INST	M C	STAND	Remarks
1	O/H L	C-414	ST-210	
2	O/H R	C-414	ST-210	
3	F TOM	e-904	-	
4	M TOM	e-904	-	
5	H TOM	e-904	-	
6	HH	PG81	ST-259	
7	SN Top	SM-57	ST-259	
8	SN Bottom	SM-57	ST-259	
9	Kick OFF	BETA52A	ST-259	
10	Kick ON	ATM-25	ST-259	
11	E. Bass Line	SM-58	ST-259	
12	E. Bass M.c	Type-85	-	
13	Rhodes L	Type-85	-	
14	Rhodes R	Type-85	-	
15	Ambience near L	4090	ST-259	
16	Ambience near R	4090	ST-259	
17	Ambience far L	4090	-	
18	Ambience far R	4090	-	

図4.2 マイク回線のリスト

舞台内のMicrophoneSettingについては楽器単体の他にAudienceマイクを4本設置する。音源より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DPA4090を2本つりさげでのSettingとした。

NearのMicはステージの角の位置、高さ1mくらいの所にsettingする。Farについてはステージから奥に5m、高さ5mのところにセンターを中心に約1mの幅で2本設置した。

録音したデータを編集用のDAWにてDrmsMix, BassMix, Piano(Rhose)Mix, Ambience near L, R Ambience far L, R, それぞれ楽器ごとの単体で纏め上げた音源に編集し、アプリケーションに組み込む作業を行う。

4.2 アプローチ

4.2.1 要求事項

SDMは、3次元映像音声メディアをオブジェクト指向で管理し、映像・音響の視聴空間と視聴メディアの設計を行うことを目的としている。その実現には、上述のコンサートホールでの収録のように多数のデバイスを用いて記録した視聴空間のデータを管理する方法が必要となる。SDMのデータベースの要求事項は以下の3つが挙げられる。

1. 3次元映像音声メディアを管理出来ること
2. アプリケーションに非依存な形式で保存されていること
3. 表現のレベルに応じてデータが取得できること

1つ目は、3次元映像音声メディアを管理するのに特化した形式、すなわちデータの性質を詳しく表せる形式で保存するということである。

2つ目は、様々なアプリ開発に利用可能な形式で保存するということである。SDMデータを用いたアプリには、ユーザとインタラクティブな挙動を示すアプリや、パッケージメディアのミキシング、半自動ミキシングなどが考えられる。どのようなアプリでもデータを利用できるよう、特定の用途にスペシフィックではなく汎用的な形式で保存することが必要となる。

3つ目は、目的に応じた表現のレベルが設定出来るということである。汎用と固有の情報を分離することで、様々なユースケースに対応可能となる。例えば「合唱コンクール」のようなイベント単位から「演奏者」のようなユースケース固有のレイヤまで表現出来る必要がある。

上記の条件を満たすデータベースの形式として、本研究ではRDF・LODによるデータの管理を行う。

4.2.2 RDF・LOD

インターネットが普及した現代において、web上に存在する無数のデータが検索によって容易に手に入るようになった。また、自分の持つデータを世界中の人々とシェアするために、オープンデータとしてインターネット上に公開することも可能となった。こうした状況において保有するデータをオープンデータとして広く利用してもらうには、誰もが容易に利用できる形式、または機械が自動的に処理可能な形式で公開することが望ましい。

データの公開方式の評価方法として、World Wide Webの考案者であるTimothy Berners-Leeは5 Star Open Data^{*2}というものを提案している。

これは、公開するデータの形式を5段階で評価したものであり、1段階目から3段階目までは次のような形式での公開を表している。

1. pdfやjpgでの公開: 文章データをjpgで公開するといった、自動処理が困難である形式での公開
2. xlsやdocでの公開: 特定のアプリケーション以外では前処理なしに扱うことのできない形式での公開
3. csvやtsvでの公開: データ解析において一般的に用いられている形式での公開

この指標の4段階目ではRDF(Resource Description Framework)^{*3}という形式を推奨している。

これによって記述されたデータはURIによって参照可能であり、機械による自動処理に適している。このRDFが外

*2 <http://5stardata.info/en/>

*3 <https://www.w3.org/RDF/>

部と連携したものはLOD(Linked Open Data)と呼ばれ、5 Star Open Dataの最高レベルである5段階目となっている。LODの形式でデータを公開することでweb上のデータ同士がリンクしあい、オープンデータの利用が推進される。

4.3 三次元視聴オブジェクトのデータ管理

三次元視聴オブジェクトのデータ管理はデータの制作、加工、利用において想定される機能要求を満たすため、以下の通り管理されるべきである。

4.3.1 三次元視聴オブジェクトに関するLinkedOpen Dataの定義

データの制作の利便性や、加工、利用における機能性や応用の自由度を高めるために、三次元視聴オブジェクトは記録対象の構成や作業フローの実態に即したデータ構造を持たせる必要がある。図4.3はコンサート全体を表現することをスコープとして設計した三次元視聴オブジェクトのクラス構造である。一番上位にあるコンサートオブジェクト(Concert)がコンサート全体を表現し、会場(Facility)とそのコンサートの演目(Program)に関連付けてある。演目には楽曲(Song)を関係づける。演目进行する側のオブジェクトとして演者(Performer)を定義した。そのサブクラスとして演奏パート(Part)、楽器(Instrument)、演奏者(Player)、オーケストラ(Orchestra)がある。三次元視聴オブジェクトを記録する行為及び記録物を定義するためのオブジェクトとしてレコーディングイベント(RecordingEvent)とシグナル(Signal)を定義した。レコーディングイベントは録音機器(Recorder)と録音トラック(Track)で構成され

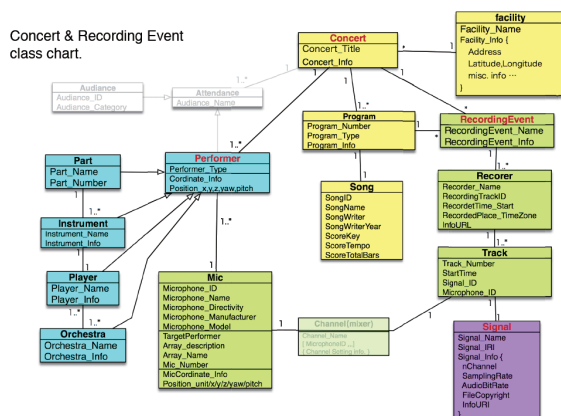


図4.3 三次元視聴オブジェクトのクラス構造

ており、録音トラックには音声入力のためのマイク(Mic)を関連付けてある。さらにマイクには録音対象の演者(Performer)が関連付けてあり、録音データの内容を表現する手がかりとしている。また、録音トラックにはシグナルオブジェクト(Signal)が関連付けてある。シグナルオブジェクトには録音データの格納場所のリンクが埋め込まれており、このリンクを辿ることでシグナルオブジェクトから録音データを取り出すことができる。図中の色分けは、データベースを作成する際に便宜上取り決めたグループを表している。また、グレースアウトした部分は将来拡張用とし今回は使用しなかったクラスである。以上のクラス構造は音声データの記録にフォーカスして設計したものである。映像データの記録を行う場合は、Micの代わりにVideoCameraを定義し、Recorder、Track、Signal各クラスの属性データを映像データの取扱に合わせて適切に変更することで表現が可能である。

4.3.2 データベース構成とアプリケーション利用

図4.4は三次元視聴オブジェクトを表現するLinked Open Dataデータベースの構成である。SDM LODServerには個々の三次元視聴オブジェクトを表現しているjsonld形式のjsonファイルを格納する。データ制作の便宜を考慮し、1つのjsonファイルには上記クラス構造をグループ化した上で記述している。jsonファイルが表現するそれぞれのオブジェクトファイルの記載内容は以下の通りである。

コンサートオブジェクトファイルはConcert、Facility、Program、Songクラスの属性情報を、パフォーマーオブジェクトファイルはInstrument等のPerformerクラスの属性情報を、レコーディングイベントオブジェクトファイルはRecordingEvent、Recorder、Track、Micクラスの属性情報

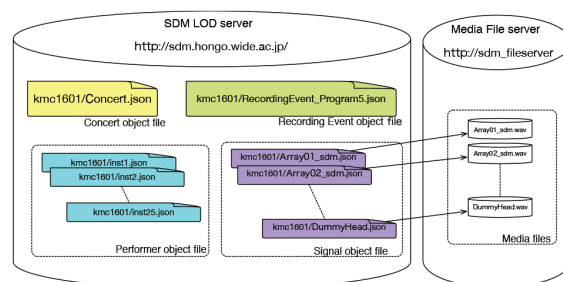


図4.4 データベース構成

報を一括して記録している。またシグナルオブジェクトファイルはSignalクラスを記述しており、そこからMedia FileServer上のメディアファイルがリンクされている。

データベースを利用するアプリケーションは任意のキーワードでデータベースクエリを行い、上記jsonファイルを取得する。jsonldの規定に従いデータ構造とその表現の意味を解釈・認識することで、リンクされたオブジェクトを次々と取り出すことができる。結果的にアプリケーションが必要とするメタデータやメディアファイルを取り出すことが可能となる。

第5章 360Square:インタラクティブな自由視聴点での三次元映像音声の再生

SDMコンソーシアムでは3章で述べた三次元映像音声コンテンツの協調制作フレームワークのコンセプトを元に、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「360Square」を開発した。

現在、360Square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(4.1節参照)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き、360Square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・無効化することができる。

5.1 設計

360Squareの全体の設計概要を図5-1に示し、映像音声の処理とユーザインターフェイスに分けて解説する。

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元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ことも想定する。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。映像にも移動や実際には収録されていない物など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 ユーザインターフェイス

「360Square」は、HMDで鑑賞に応用できるが、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5.1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。

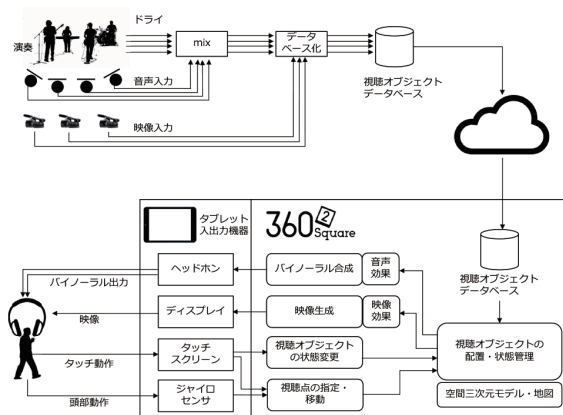


図5.1 360 Squareの設計

また、視聴点の移動に関しては、画面に常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ーを使い、タブレットの向いている方向を取得して、視点移動できるモードを用意する。

5.2 実装

360Squareは、Unity Version 5.5.0f3を用いて開発した。

本アプリケーションは起動すると360Square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(4.1節参照)である。

クラシックコンサートを選択すると図5.2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5.3に示すクリーンショットのような画面で行われる。図5.3は図中で音符を描画した立

表5.1 タッチ動作

動作	通常モード	位置固定モード
1 指ドラッグ	ドラッグ方向へ位置を移動	ドラッグ方向へ視線を移動
2 指回転	視線を水平方法に回転	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5.2 360 Squareのヘルプ画面

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時の映像である。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。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)で示される通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)に示す通り、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、視聴点が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5.4に示す。



図5.3 通常モードのスクリーンショット



図5.4 位置固定モードのスクリーンショット

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)に示す通り、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。図中(7)～(10)は、通常モードと位置固定モードの両方に常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5-3に示した。図中(9)、(10)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。

5.3 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macOS Sierra(Version 10.12.3)がインストールされたMacBook Pro(15-inch, 2016)である。プロセッサは2.9 GHz Intel Core i7、メモリは16 GB 2133 MHz LPDDR3であり、グラフィックチップはRadeon Pro 460 4096 MBと、Intel HD Graphics 530 1536 MBを搭載している。

実験は、Unity Version 5.5.0f3のエディタ上で360Squareを実行し、マウスで操作し、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。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。大きく影響すると思われる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5.5にビルボードでのジャズセッションのコンテンツ(8個の360度動画と、7個のオーディオ音源が含まる)を表示した際の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、全ての操作イベントの発生ではなく、視聴点が移動するイベントの影響が多かったため、図5-5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5.2～5.4で示した、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別れた。図5.5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、全体として概ね30フレーム/秒を維持していることがわかる。また、ヘルプ画面表示とイベント発生時は、最も処理が遅くなっているが、それ以外の時間には60フレーム/秒の性能を維持できている。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66.0%を占めている。図5.5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、11.4ms、11ms、9.6msの時間がかかってい

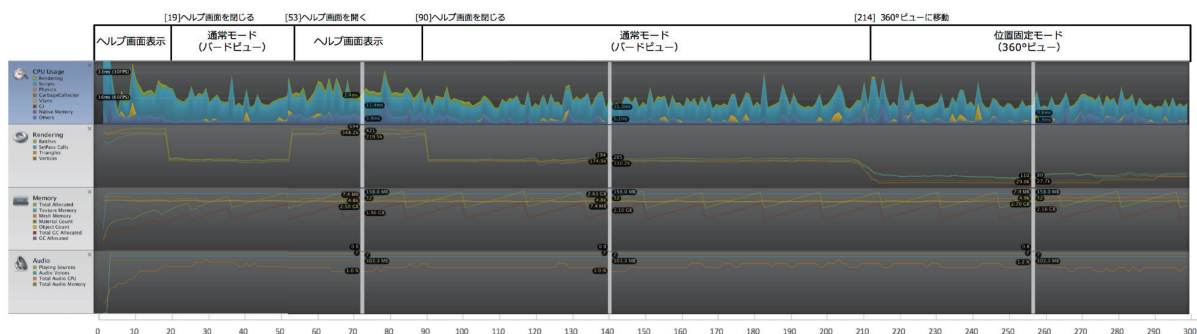


図5.5 360Squareの性能評価

る。ヘルプ画面では、その他、ステージの全景を描画するための時間が、5.9msかかっており、フレームレートが落ちる原因となっている。

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5.1に示すとおり、全ての値が、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わかる。

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2.5GB～2.7GB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り当て量(Unity)が420MB、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1.94GB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100MB、プロファイラーデータに使用されているメモリ量(Profiler)が約20MB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、有効化した時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300MB、CPU利用率は約1%であった。

本節では、360Squareを動作させたMacbook ProでのUnity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

表5.1 各モードにおけるレンダリングの数

	ヘルプ画面	通常モード	位置固定モード
Batches 数	594	294	110
SetPass calls 数	421	205	80
三角メッシュ数	348.2k	174.3k	29.0k
頂点数	219.5k	110.2k	27.7k

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。

第6章 SDM試作システムを利用したハッカソンの開催

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】にて360Squareのデモンストレーションを行った。審査員と約100名の一般の来場者は、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、360Square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、イベント運営に当たったビルボードジャパンによってLINE live^{*4}を利用して、スマートフォン及びPCでライブ視聴できるように放送されており、のべ3000人を超える視聴があった。

また、審査員にはipad Pro及び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360Square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、及び審査・受賞の様子は、様々なメディアで掲載された。数例を以下に挙げる。

- ・(公式イベントサイト)LIVE MUSIC HACKASONG | Special | Billboard JAPAN^{*5}
- ・“ライブ体験の拡張”を目指すハッカソン、ビルボードライブ東京にて最終審査が開催| Daily News | Billboard JAPAN^{*6}
- ・“ライブ体験の拡張”をテーマにしたハッカソン「Live Music HACKASONG」最終審査・表彰式を開催| Musicman-NET^{*7}

*4 <http://live.line.me>

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

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

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

図6.1は、最優秀賞と優秀賞のチームがステージ上で受賞している様子である。図6.2は、受賞後の懇談会で、SDMチームとスポンサー企業のNapsterの最高技術責任者(CTO)で行なった撮影である。

第7章 関連研究

RDF・LODを用いたデータ管理の研究は、データベースを構築するための語彙の研究や、実際にRDF・LODを用いたアプリケーションの作成に関する研究などがある。

語彙の定義に関する研究として、Y.Raimondら[11]によるThe Music Ontologyの研究がある。The Music Ontologyは、音楽に関するメタデータを詳細に記述することを目的として構築されたRDF Vocabularyである。The Music Ontologyは、3つのontologyを元に構築されている。1つ目のontologyはthe TimelineOntology^{*8}である。これは時刻(time instant)・時間間隔(time interval)・時間の流れ(timeline)を表現することを目的として構築されたオントロジーであり“this performance happened the 9th of March, 1984”といったデータをRDFで記述することが出来る。

2つ目はthe Event Ontology^{*9}である。“this musical event was performed in Suntory Hall in Tokyo”のようなイベントの情報を記述するためのontologyであり、場所

や公式ホームページといったイベント関係の情報を記述することが出来る。

3つ目はFriend-of-a-friend ontology^{*10}である。これは“the artist of this classic song is Johann Sebastian Bach”といった、人物に関係する情報や人間関係を記述するためのontologyであり、様々なLODにおいて広く用いられている。

以上のオントロジーを元に、音楽に関する情報を記述するための語彙を追加・再定義したものがThe Music Ontologyである。具体的には“performed in”, “conductor”, “recorded in”, “remix of”などといった、演奏や楽曲の情報を詳しく表すための語彙が定義されている。

また、RDF・LODで管理されているデータを用いたアプリケーションの研究例として、T. Di Noiaら[12]による映画のリコメンドシステムの研究を紹介する。この研究では、ユーザの嗜好を表すuser profileという値と、映画間の類似度を表すsimilaritiesという値を組み合わせて学習させている。

ユーザの嗜好を計算するのにNoiaらが用いたデータセットは、MovieLens^{*11}というものである。これはリコメンドシステム開発やベンチマークのために作成された、映画のレビューについてのデータセットである。Noiaらが用いたバージョンには、6,040のユーザによる、3,952の映画に対する、1,000,029のレビューが含まれて



図6.1 Live Music Hackasongでの授賞式の様子



図6.2 Live Music Hackasongでの授賞式の様子

*8 <http://purl.org/NET/c4dm/timeline.owl>

*9 <http://purl.org/NET/c4dm/event.owl>

*10 <http://xmlns.com/foaf/0.1/>

*11 <https://movielens.org/>

いる。このデータセットから、各ユーザの嗜好を計算する。また、映画間の類似度を計算するのに用いたデータは、DBpedia[13]、LinkedMDB[14]、Freebase[15]というLODから抽出された、MovieLensで扱っている映画に関するRDFのグラフである。類似度は、ある映画とある映画が「どれだけ同じ情報を共有しているか」によって計算される。例えば「主演が同じである」「ジャンルが同じである」といった内容がRDFで記述されている場合、高い類似度となる。

こうしてRDF・LODから計算された映画間の類似度と、ユーザの嗜好のデータを組み合わせることで、高精度なリコメンドシステムを構築することに成功している。

これにより、あるユーザがある映画に対してどのような評価を下すかを推定し、高評価を下すと推定された映画がリコメンドされることとなる。

第8章 まとめ

本報告書では、2014年より開始したSDMコンソーシアムで進めて来た、視聴空間サービスのソフトウェア制御による研究で、2016年度に行なった活動について報告した。その中で、以前より策定を進めて来たSDMアーキテクチャに基づき、三次元映像コンテンツの協調制作のフレームワークを構想し、その開発を進めて来た。実際に、オーケストラのコンサートの収録とジャスセッションの収録を行い、実験を進めるための素材データを取得した。さらに、三次元映像メディアは収録データをオープンに公開し、自由なコンテンツ制作を促進するため、オープンデータ化の仕様について策定を進めた。また、収録データを自由な視聴点においてインタラクティブに再生するアプリケーション「360Square」を設計、実装し、評価を行なった。さらに、ビルボードジャパンが開催したLive Music Hackasongにおいて360Squareのデモンストレーションを行った。このイベントで、審査員と一般の来場者の投票により、優秀賞を受賞した。三次元映像メディアのオープンデータ化を実施するにあたり、様々なオープンデータがどのように定義、利用されているか、Resource Description FrameworkとOpen Linked

Dataに関する動向調査を行い、関連研究としてまとめた。

SDMコンソーシアムでは、ソフトウェア処理による視聴空間の制御、映像音声を制御するネットワーク機器、インタラクティブ・ユーザインターフェイス、SDMプラットフォームを利用したコンテンツ作成など、共同研究活動に参加するパートナーを募集しています。

ご協力をよろしくお願いいたします。