LiVRation: VRによる自由視聴点映像音声のインタラクティブ再生

粕谷 貴司[1,2]　塚田 学[1]　菰原 裕[1]　高坂 茂樹[3]　水野 拓宏[4]　野村 譲誉[4]　上田 雄太[5]　江崎 浩[1]

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，中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．本研究では，クラシックコンサートとジャス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「SDM360[2]」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 にて 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．

LiVRation: An interactive audio-visual service by using Virtual Reality

Takashi Kasuya[1,2]　Manabu Tsukada[1]　Yu Komohara[1]　Shigeki Takasaka[3]　Takuhiro Mizuno[4]
Yoshitaka Nomura[4]　Uta Ueda[5]　Hiroshi Esaki[1]

1. はじめに

インターネットを前提とした視聴サービスが登場し，中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している．そこで著者らは 2014 年から Software Defined Media (SDM) コンソーシアムを設立し，オブジェクトベースのディジタルメディアと，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].

近年，Youtube やニコニコ動画などの動画配信サービスや HMD のような VR デバイスの発展により，4 K 映像や 360 度動画なども容易に再生できるようになってきた．5G のような大容量の通信技術や，MPEG4-ALS[2] などの音声のロスレス配信技術も一般化してきており，今後はより高品質でリッチなストリーミングメディアが普及していくだろう．一方で，コンサートやライブの需要や人気が増大しているにもかかわらず, それらを記録した Blu-ray などのパッケージメディアでは収録機器の設置位置に制約を受けるため，視聴者の意思による自由な角度，距離による視聴は困難である．一般家庭においても VR コンテンツの視聴が可能になってきたが，配置されている音源は静的なものがほとんどで，高品質なストリーミング再生は困難であった．こうした課題を解決するため，我々は VR 技術とロスレス配信技術，SDM オントロジを使って，自由視点でのライブ体験を実現する「LiVRation」を試作・検証した．

本論文ではまず，第 2 章で本研究の目的を述べる．第 3 章で SDM によるアプリケーションのデータ構造・意味表現である SDM オントロジについて述べる．第 4.3 章でシステムの設計，実装およびコンテンツの収録についての述べ，第 5 章でそれらの評価と各所で実施したデモンストレーションの報告を行う．第 6 章で関連研究について述べ，第 7 章で本論文の結論と今後の課題について述べる．

2. 本研究の目的

本研究の目的は，音楽イベントの遠隔配信を対象とし，自由視点映像音声のインタラクティブ再生を行うことである．具体的には，収録したライブを自由視点で視聴するとともに，収録された音声を自由にコントロールし，かつ演奏者や楽器のメタデータや，SNS などで発信されるコンテンツについての情報についても動的に扱うことができる環境である．そのために，我々は LiVRation を試作し評価を行った．システム要件として以下を想定した．

自由な視聴体験：　利用者が視聴位置を自身で決めることができる．位置に基づいて，収録音声がリアルタイム

1　東京大学
The University of Tokyo
2　株式会社 竹中工務店
Takenaka Cooperation
3　エスイーディー 株式会社
Sophisticated Engineers Department (SED)
4　株式会社 アルファコード
5　株式会社 CRI・ミドルウェア

© 2018 Information Processing Society of Japan

で自動合成され，ヘッドホン等で視聴することができる．

遠隔配信： ライブをリアルタイム配信することが可能であり，再生環境を整えれば，ネットワークを介してどこでも視聴することができる．収録済みの音源に対しても，同様に配信することができる．

高臨場感・没入感： ４Ｋやロスレスなどの高音質な映像・音声を扱うことができる．より臨場感の高いライブ体験をするために，VR や振動伝達（ハプティクス）を利用する．

インタラクティブな体験： コントローラを使って特定の音を強調したり，不要な音を消したりすることができる．ソーシャルメディアと連携して，コメントなどを共有することができる．

上記を実現するため，我々は収録した音源をオブジェクトオーディオ化（SDM オブジェクト化）し，多拠点映像とともに配信する仕組みを構築するとともに，SDM オントロジを使って，それぞれの SDM オブジェクトの情報を定義した．それらを受信し，VR デバイスで再生する LiVRation の詳細について，以下で述べる．

3. SDM オントロジ

SDM コンソーシアムでは，データを記録・再現するためのシステム構築だけでなく，実際にライブやコンサートのデータを記録し，活用するためのアプリケーション開発も行っている [3]．記録されたデータは，インターネットで公開され，外部からも利用されることが望ましいといえるが，当該データには対象の映像や音だけでなく，位置情報や楽器の向き，演奏された曲目や会場情報，録音プロセスといった情報も記録される．それらのデータは相互に関連するが粒度が異なるため，構造化し相互にも外部にも連携可能とする必要がある．

インターネット上でのデータ構造の公開する手法としては，RDF(Resource Description Framework) が推奨されている．これによって記述されたデータは IRI(Internationalized Resource Identifier) によって参照可能であり，自動処理に適している．この RDF が外部と連携したものは LOD(Linked Open Data) と呼ばれる．LOD の形式でデータを公開することで web 上のデータがリンクしあい，オープンデータの利用が推進される．一般的には SPARQL Endpoint と呼ばれるサーバをインターネット上に配置し，SPARQL クエリを処理するサービスとして公開される．

我々は SDM が保有しているデータを LOD に変換するために，SDM 用の RDF Vocabulary である”SDM オントロジ”を定義した [4]**?**．SDM オントロジでは，イベント全体の情報である SDMEvent，イベント内のプログラム情報である Context，収録対象の情報である Target，収録機器情報である Recorder，生成されるメディア情報である MediaObject の５要素によって大きく構成される．SDM のアプリケーションでは，各要素のアノテーションを使って仮想空間における楽器や演奏者の位置情報などの自動セットアップを行う．

4. LiVRation

本節では LiVRation で用いた収録コンテンツの詳細と，システムの設計指針，実装の詳細について述べる．なお，LiVRation は SIP 協議会が主催しているハッカソンイベントである「Billboard LIVE MUSIC HACKASONG 2017」のために作成されたシステムで，「Live で Vibration を伝える VR 配信」というコンセプトをもとに開発されている．

4.1 システム概要

LiVRation は HMD を使った没入環境において立体音響環境を体験することができる．また，MPEG4-ALS を使ってハイレゾ音声のストリーミング再生を実現している．視聴者は，HMD を用いて仮想空間を動き回るとともに，視聴位置での音響空間を仮想的に体験することができる．加えて，特定の楽器の音のみの抽出や，Twitter の情報表示，振動伝達デバイスからの振動を感じることもできる．図 1 に実際の動作画面を示す．画面上に表示されている球状のオブジェクトに収録された 360 度動画や音声がマッピングされている．また，視聴者を囲うように Twitter のメッセージが表示されている．



図 1 LiVRation の動作画面

以下では，コンテンツの収録環境，システムの設計指針，実装について詳細を述べる．

4.2 収録環境

LiVRation で利用するコンテンツは，2018 年 1 月 30 日に Billboard Live Tokyo で行われた「仮谷せいら」のライブリハーサルである．リハーサル中に収録した映像，音声を配信コンテンツとして完成させ，収録から約 4 時間後にはイベントでの最終発表を行った．

© 2018 Information Processing Society of Japan

図 2 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．

ライブの構成は，ボーカルにシンガーソングライターの「仮谷せいら」で，バックバンドにドラムを担当する松浦と，キーボード，ベースを担当する川原という構成でであった．会場のリスナーへの表現手法としては全ての楽器にマイク或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 100dBspl となる．

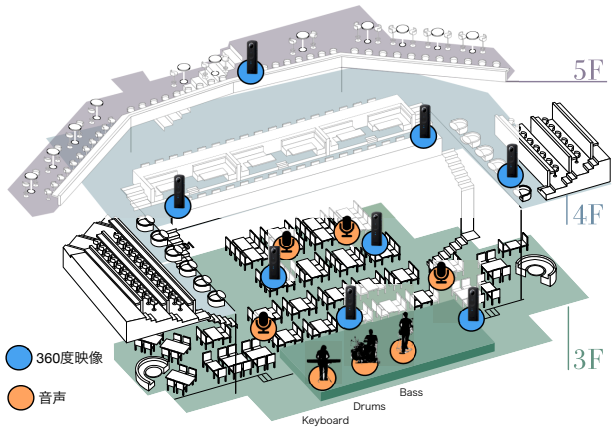


図 2　ビルボードライブ収録 (2018) のカメラとマイクの配置

舞台内のマイク構成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マイク DPA4090 を 2 本つりさげての構成とした．音源より近いマイクはステージの角の位置，高さ 1m くらいの所に設置する．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 のところにセンターを中心に約 1m の幅で 2 本設置した．音声は全て SR 用ミキシングコンソールに纏められおり，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全てミキシングコンソールの設定に依存する為，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 にて VocalMix，DrumsMix，BassMix，KeyboardMix それぞれボーカルと楽器ごとの単体で纏め上げた音源に編集し，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．

映像を収録する 360 度カメラは，「Insta 360 Pro」1 台，「Garmin VIRB 360」4 台，「Ricoh Theta V」2 台を利用した．全てのカメラで 3840 x1920@30fps(4K) での撮影を行った．Insta 360 Pro は，6 枚の F2.4 魚眼レンズを備えたハイエンド 360 度カメラであり，ライブ会場の全景を撮影するため，5 階席に設置した．Garmin VIRB 360 は，2 枚のレンズを備えた 360 度カメラであり，設置場所は，ハッカソンの審査員席に利用された 4 階正面，ボーカルの真前，ドラムの真前，ステージ右横の 4 カ所を選択した．Ricoh Theta V は，2 枚のレンズを備えた 360 度カメラであり，もっとも軽いため，ライブ会場の天井に吊ってあるバトンに 1 台設置した．また，もう一台を 4 階席右手に設置し撮影を行った．

4.3　設計

LiVRation は図 3 に示す SDM アーキテクチャに基づいて設計を行った．SDM アーキテクチャは，ネットワーク上に存在するコンテンツおよびその配信システムと，実空間内の各種情報を SDM オブジェクトとして情報空間に取り組み，その結果を自在に視聴空間に投影する SDM 拠点から構成される．

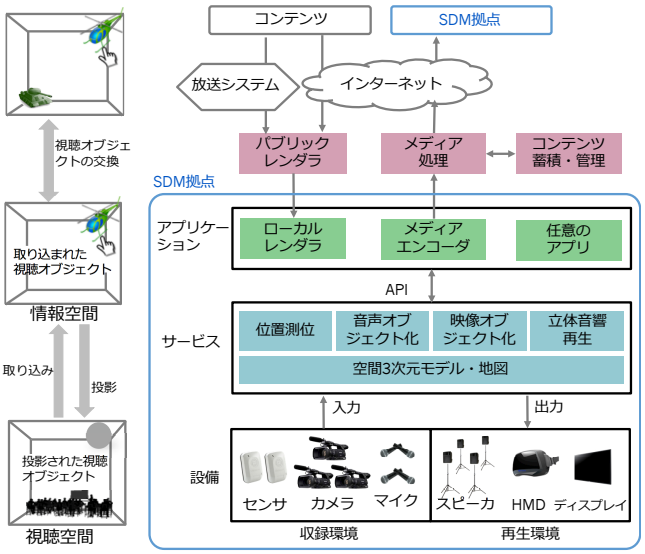


図 3　SDM アーキテクチャ

LiVRation の設計指針も SDM アーキテクチャに準拠している．図 4 に LiVRation の設計概要を示す．

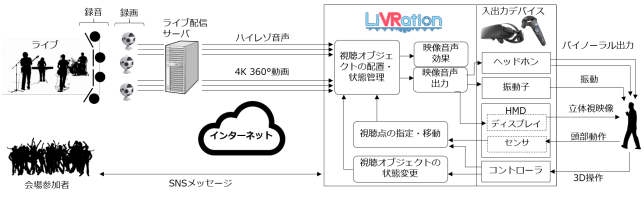


図 4　LiVRation 設計概要

前章に述べた映像音声は，対象となる 1 曲を切り出して編集され，ライブ配信サーバに格納する．ライブ配信サーバからネットワークを通じて，ハイレゾ音声と 4K 360 度動画を SDM 拠点である LiVRation クライアントへとストリーミング配信する．クライアントでは，これらの映像音

© 2018 Information Processing Society of Japan

声ストリームを音声，動画を表す SDM オブジェクトとして扱い，図 5 のようにあらかじめ製作しておいた仮想空間内に配置し，全てのストリームのバッファが一定量蓄積したのちに再生を行う．

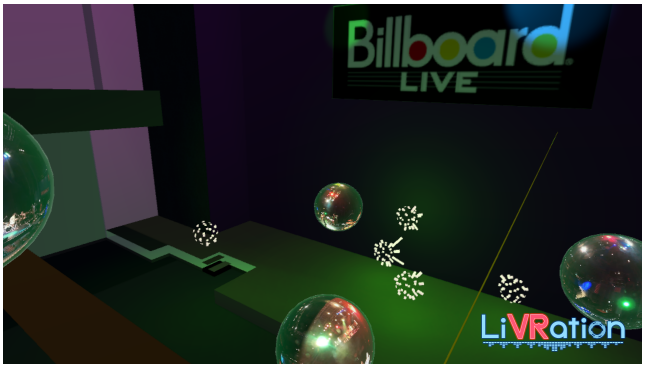


図 5　仮想空間への SDM オブジェクトの配置

音声は，音声オブジェクトと視聴位置との距離関係から，音量と聞こえる方向を決定し，ヘッドホンからバイノーラル出力される．また，ヘッドホンジャックからローパスフィルタを通して低音成分に反応する振動子を接続し，身体に装着することでより臨場感の高い体験を提供する．さらに，音声は周波数成分の音量ごとに異なるバーの長さが変化する映像効果を付与することで，可視化を行う．

360 度動画は球体の両面に貼り付け再生を行う．これらの動画や映像効果は HMD の両眼映像を通じてユーザに立体的な映像を提示する．ユーザは，HMD のセンサを通じて頭部の動きを LiVRation クライアントに伝える．これにより自由に視聴したい方向を向く事が可能であり，3 次元の動きを取得できるコントローラを 3D 操作することにより，様々な意図をインタラクティブにシステムに伝える事ができる．

ユーザはコントローラを使って，音声の制御や視聴位置の移動などを行うことができる．コントローラで音声を掴んだ状態で，自分の方に引く操作により，その音声だけをソロで視聴すること可能であり，歌声をアカペラで聞いたり，楽器の音をインストルメンタルで聞いたりすることができる．反対に音声を掴んだ状態で，向こう側に押し込む動作で，全ての音声を有効化して通常通り音声が混ざった状態で音声を聞く事ができる．加えて，音声を掴んだ状態で手首をひねると音量の強弱を調整できる（図 6）．また，コントローラから照射されるポインタを映像を提示している球体に当てた状態で，選択するとその場所への移動が可能である．

最近では，ライブ放送において一体感を味わうため，Twitter などの SNS を通じて，他の視聴者とインタラクションを行う事で，ライブを楽しむ事が一般化している．LiVRation では，Twitter の関連ツイートを仮想空間内に提示する事で，この他の視聴者とのインタラクションを実現する．



図 6　コントローラによる音声オブジェクト操作

4.4　実装

上記の設計指針に基づき，表 1 に示す開発環境，フレームワークを用いて LiVRation を構築した．図 7 にシステムの実装概要を示す．

表 1　LiVRation 実行環境

実行マシン	CPU： メモリ： グラフィックス： SSD：	Intel Core-i7-8700K 32.0 GB GeoForce GTX 1080 Ti 480GB
HMD		Samsung HMD Odyssey
クライアント開発・実行環境		Unity 2017.2.0f3(64-bit)
配信サーバ		Wowza Streaming Engine 4.7.5
サービス		NodeRED v0.17.5
SPARQL Endpoint		graphdb-free-8.3.0
時系列データベース		Elasticsearch v5.6.2

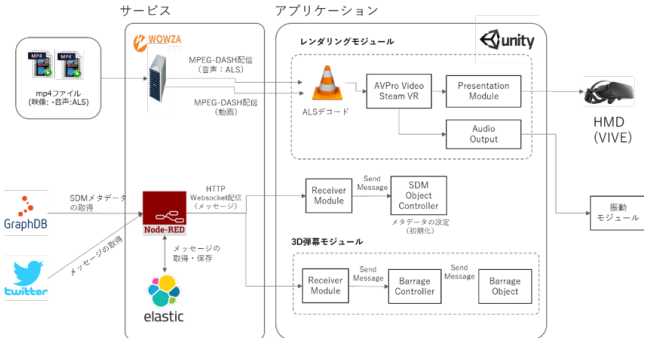


図 7　LiVRation 実装概要

コンテンツ配信の仕組みとしては，配信サーバである Wowza に mp4 にエンコードした収録動画，音声を配置することで実現している．それぞれの配信コンテンツには URL が自動的に付与され，MPEG-DASH 等で配信することができる．なお，Wowza は MPEG4-ALS にも対応している．これらのストリーミングデータを Unity で実装した LiVRation クライアントが受信する構成となっている．構築環境としては，今回の実装では Wowza はローカルネットワーク内に配置し，そこからクライアントがストリーミング配信を受ける構成とした．技術的にはインターネット

© 2018 Information Processing Society of Japan

上からでも可能ではあったが，今回デモンストレーションを行ったライブ会場等では一定品質の通信環境の調達が難しかったためにそのように構成している．インターネット上から配信する場合，コンテンツ調達のための認証等の仕組み構築が課題になるといえるが，今回はそこまで検討されていない．LiVRation クライアントからストリーミングデータ以外を取得するサービスは，NodeRED を使って HTTP によるサービスとして実装した．具体的には，NodeRED の Twitter API を使って配信コンテンツに関係する Tweet を収集するモジュールと，クライアントの初期化時に SDM オントロジを格納している GraphDB に対して SPARQL クエリを投げ，その結果を LiVRation クライアントに送信する機能を持つ．前者については，例えば「仮谷せいら」といったキーワードを予め設定しておくと，それに対応したメッセージが websocket で継続的にクライアントに送信される．過去データについては，時系列データベース（Elasticsearch）に保存され，初期化時に送られる実装となっている．

LiVRation クライアントには，コンテンツの受信や映像・音声のエンコード，レンダリングするモジュールをや各種サービスの情報を受けて画面内に反映，表示する機能モジュールを配置している．具体的には，以下のモジュールを有する．

レンダリングモジュール:Wowza からのストリーミングデータは Unity のプラグインによって処理されて表示される．動画は AVPro Video プラグインによって球体オブジェクトのテクスチャにマッピングされ，３６０度動画として再生される．この際，360 度カメラで収録した音声も同時に取得できるが，ミュートしている．音声は UniversalMediaPlayer プラグインを使って音声オブジェクトにマッピングされ，周波数成分の音量によって変化するエフェクトが加えられる．なお，UniversalMediaPlayer は VLC と FFmpeg をベースとしたプラグインであり，MPEG 4-ALS の再生も可能である． エフェクトについては Rhythm Visualizator Pro プラグインを用いて実現している．前節で述べた距離による音声ミックスは，Unity によって自動計算され，出力された音声はヘッドホンや振動デバイス [5] で再生することができる．動画・音声それぞれの再生に対しては，現状では特定の同期処理を行っていない．動画はそれぞれの配信コンテンツがプラグイン内ですべて再生可能となった時点で再生される．音声は音声オブジェクトが個別に再生可能と判断された時点で再生が自動再生される．再生のタイミングについては，UniversalMediaPlayer プラグインのバッファに関するパラメータによって多少は調整することができるが，Wowza からの配信タイミングに依存するので，映像との完全な同期は Unity だけでははは実装が難しい．これらの高度な同期処理については今後の課題といえるが，ローカル環境での再生においては，再生のズレが気になるようなことはなかった．

上記に加えて，HMD のセンサやコントローラの操作に応じて，視聴位置である球状オブジェクトの中を旋回・移動したり，音声オブジェクトへの操作によって音量を調節する機能を有する．

SDM モジュール:SDM オブジェクトのメタデータ（音源，演奏者，位置情報など）は，クラウド上に構築された SPARQL endpoint である GraphDB から，サービス層を介して LiVRation クライアントに取得される．クライアントの初期化時に Unity から，サービスプラットフォームに対して HTTP でアクセスすると，予め用意された SPARQL クエリで GraphDB に問合せを行い，その結果がクライアントに送られる．そこで取得したメタデータをレンダリングモジュール等に渡す実装となっている．今回実装した LiVRation のために定義したオントロジの抜粋を以下に示す．

```
@prefix sdmo: <http://sdm.hongo.wide.ad.jp/
    sdmo/> .
@prefix sdm: <http://sdm.hongo.wide.ad.jp/
    resource/> .

### イベントの定義
sdm:hackasong rdf:type owl:NamedIndividual
    ,
                        sdmo:SDMEvent ;
        s:contentLocation sdm:billboard ;
        s:name "Billboard LIVE HACKASONG
            2017"@ja ;
        s:startDate "2018/1/30 17:30" ;
        sdmo:has sdm:song1 ;
        sdmo:recording sdm:bass_mic ,
                       sdm:drum_mic ,
                       sdm:far_left_mic ,
                       sdm:far_right_mic ,
                       sdm:keyboard_mic ,
                       sdm:near_left_mic ,
                       sdm:near_right_mic ,
                       sdm:vocal_mic .

### コンテキストの定義
sdm:song1 rdf:type owl:NamedIndividual ,
                    sdmo:Song ;
        s:name "Colorful World" ;
        sdmo:performed_by sdm:bass ,
                          sdm:drum ,
                          sdm:keyboard ,
                          sdm:vocal ;
        sdmo:program_number "1" ;
        sdmo:recorded_by sdm:bass_mix ,
                         sdm:drum_mix ,
                         sdm:far_left_mix ,
                         sdm:far_right_mix
                             ,
                         sdm:keyboard_mix ,
```

© 2018 Information Processing Society of Japan

```
                                sdm:near_left_mix
                                    ,
                                sdm:near_right_mix
                                    ,
                                sdm:vocal_mix  .

### ターゲットの定義
sdm:vocal  rdf:type  owl:NamedIndividual  ,
                        sdmo:Musician  ;
           s:name  仮谷せいら""@ja  ;
           sdmo:attend_to  sdm:hackasong  ;
           sdmo:localX  "0.0"^^xsd:float  ;
           sdmo:localY  "0.5"^^xsd:float  ;
           sdmo:localZ  "4.5"^^xsd:float  ;
           sdmo:perform  sdm:song1  .
```

SDMEvent である Billboard Hackasong のインタンスや Context である sdm:song1 を中心に performed by でリンクされる演奏者（Musican クラス），recorded by でリンクされる音源（AudioRecorder クラス）がある．演奏者である sdm:vocal には音源を設置するそれぞれの座標位置 (sdm:localX〜Z) がアノテーションとして記述されている．同様に音源に対しては，メディア配信用の URL が記述される．

3D 段幕モジュール:Twitter 等のソーシャルメディアから取得したライブに関するコメントなど，３ D コンテンツ上に表示するためのモジュールで，Twitter からのタグ付きコメントやプロフィールイメージなどを取得・取得することができる．SDM モジュールと同様に，アプリケーションの初期化時にサービスプラットフォームの特定の URI にアクセスすることで，websocket によって継続的にデータが送られてくる実装となっている．視聴位置の周りをアイコン化された Twitter メッセージが，回りながら次々と表示される演出となっている．なお，弾幕の表示はコントローラで制御（表示／非表示）することが可能である．

5. 評価

LiVRation についてコンテンツ配信含めたネットワーク評価と主観評価を行った．加えて，Billboard Live Hackasong におけるデモンストレーションの様子について述べる．

5.1 ネットワーク性能評価

今回のコンテンツでは，8 つの 360 度動画と 7 つの音声をそれぞれストリーム配信しており，コンテンツ内で同時再生している．動画と音声の再生時間は 2 分 14 秒で，データ容量はそれぞれ，160M バイト、5M バイト程度であった．コンテンツ開始時からキャプチャしたネットワークの状態を図 8 に示す．

ローカルネットワークでの配信環境では，概ね 300Mbsp

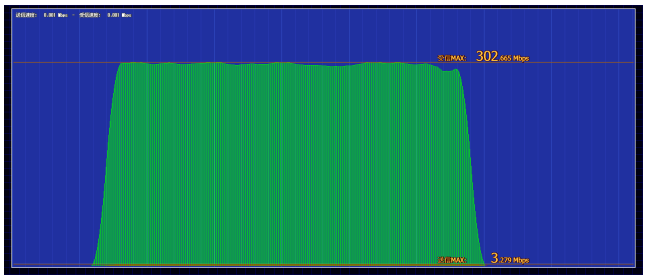


図 8 起動時のトラフィック計測

程度の受信速度であった．それぞれの Unity プラグインの仕様だと考えられるが，動画については最初にコンテンツを全てダウンロードしており，ダウンロードが終わると，音声ストリームのみの 15Mbps 程度となる．音声ストリームについては，コンテンツの終わりまで継続的にデータを受信しているのが確認できる．なお，インターネット上の配信サーバを指定した場合，最大で 100Mbps 程度の速度であり，動画の再生に 30 秒程度かかった．一般家庭への導入を考えると，ローカルネットワーク環境内にキャッシュ用のサーバや専用のミドルウェアなどが必要であるといえる．

5.2 主観評価

LiVRation の体験者〇人を対象にアンケートを使った主観評価を実施した．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？

Q2 音声の立体感は感じられましたか？

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？

Q4 映像が動いた時，音声も追従して動いたと感じられましたか？

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？

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？

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？

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り，設問 Q3 と Q4 はその組み合わ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 では動いた時の映像と音声の追従性についての問いを設定した．設問 Q5 と Q6 はコントローラによるインタラクティブな視聴に関する問いであり，Q5 は全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

© 2018 Information Processing Society of Japan

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個別の音声オブジェクトの音色が聞けたかを問う．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．

結果を**??**に示す．映像や音声の立体感に関する設問 Q1，Q2 については概ね高評価であった．同様に映像・音声の追従性についても非常に良い結果が得られている．一方で，システムの操作については，体験者によって評価が分かれた．コントローラで音声を掴んだ状態での押す，引くといったジェスチャーによってインタラクティブに個別の音声オブジェクトの音色を聴くことはできているが，その動作が直感的でない，または反応が悪いといった意見も多く，HMD のコントローラを使った操作方法については改善の余地があるといえる．アンケートにおいては「より自由な視点でライブを追体験できた」「生で見る以上の体験ができる」といった評価がある一方で，ヘッドフォンの性能，360 度動画の画質向上や軽量化がなされれば，より没入感が高まったという指摘もあった．

5.3 デモンストレーション

2018 年 1 月 30 日にビルボードライブ東京で行われた「Billboard LIVE MUSIC HACKASONG 2017」の審査において，LiVRation のデモンストレーションを行った．4.2 で述べたように，事前に行ったリハーサルにて収録を行い，そのデータを用いてデモンストレーションを行った．デモンストレーションでは，HMD を装着したステージ上の操作者の視聴音声・映像が，会場のプロジェクターやステージの左右に設置されているのスピーカーで再生されるというものであったが，Line Live によるライブ中継により 3600 人以上に視聴され，その後も様々なメディアで記事が発表されるなど，非常に反響が大きかった．審査員のコメントとしては，「完成度が高い」「すぐにでも使いたい」などのコメントがあり，結果として審査員による最優秀賞，および観客の投票数で決まる会場賞をダブル受賞した．今回の受賞によって，ビジネス的なポテンシャルについても確認することができたと考えている．

6. 関連研究

著者らの先行研究 [3][6] は，LiVRation と同様に仮想空間内での視聴位置に応じた音声ミックスを行う，タブレット端末での視聴を前提としたアプリケーションである．アンケートによる主観評価によると，仮想空間上に配置された音声オブジェクトの操作により，90 %以上の被験者が特定の音のみを聞くといった体験の有用性や，音声の立体感が感じられるという結果を得た．一方で映像の立体感やインタラクティブ性にかかわる設問については，比較的低い結果が出ており，LiVRation ではそれらが改善されたといえる．

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 [7]．

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 [8] では，22.2 マルチチャンネルの立体音響システム [9] 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．例えば，映画館やホームシアターでの採用が進む Dolby Atmos[10] や，AuroMax[11] 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[12], [13] 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て，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）[14] の収音では，アンビソニックマイクという特殊なマイクを利用し，収音したデータは 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）[15] の研究も行われている．

映像音声の遠隔配信は IP ネットワークの高速化に伴い，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

© 2018 Information Processing Society of Japan

や情感など，臨場感を高めることができる [16]．さらに，同様の技術を利用してディ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ー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 [17]．

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ュ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付け，同室感コミュニケーションシステム「t-Room」のを開発した [18][19]．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[20] や多面型全天周ディスプレイ (CABIN)[21] を始めとする没入型多面ディスプレイの研究が行われて来た．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来た [22]．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 [23] 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 [24]．

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，現在例えば，人文社会系大規模データベース [25] や，ノックアウトマウスの表現型のデータベース [26] 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[27] として定義されて，英国放送協会（BBC）などで広く利用されている．

7. まとめと今後の課題

本稿では，収録したライブを HMD を利用して自由視点で視聴するとともに，収録された音声を自由にコントロールすることが可能な LiVRation について述べた．システム要件として設定した自由な視聴体験，遠隔配信，高臨場感・没入感，インタラクティブな体験のそれぞれについて実現することができたと考えるが，　録音環境，再生環境，HMD によるインターフェースについて課題が顕在化した．4.2 章で述べた録音環境においては，スポットライトが当たった部分の映像が白飛びする現象が確認された．一般的にライブ会場は暗く，撮影中にコントラストの調整が困難であったためといえるが，こうした環境での撮影手法の確立も課題だろう．音声についても，高品質なコンテンツとするためには，プロフェッショナルによるミックスが必須である．今回の試作環境では，音声・映像それぞれの再生タイミングは Unity に依存する構成となっていたが，ネットワーク環境によっては，再生タイミングが著しくずれ，それによって没入感が阻害されるという結果になった．対策としては，音声・映像の同期のためのミドルウェアや，映像キャッシュ用のプロキシ・サーバなどをローカル環境に構築することが考えられる．しかしながら，実際のライブ環境においても距離によっては音声に遅れが発生するため，多少のずれであれば，没入感の阻害にならない可能性はあるため，SDM オントロジなどを参照しながら，距離によって柔軟な補正を行う機構の導入が必要といえる．

今後は，5.2 節で述べた，ユーザインターフェースや没入感の改良に加えて，SDM アプリケーションの BIM（Building Information Modeling）との連動について検討を行っていく計画である．LiVRation では，3D モデリング用のソフトウェアである SketchUp で作ったモデルを Unity に取り込んだが，近年の建設プロジェクトでは，3 次元形状や属性情報を含む BIM を使った設計が一般化している [28]．それらの属性情報や寸法などの三次元形状データを，SDM オントロジなどと連携させることで，より簡便なアプリケーション制作が可能になると考えている．加えて，今後は LiVRation のビジネス展開も視野に入れて活動を続けていきたい．

参考文献

[1] 塚田学, 小川景子, 池田雅弘, 曽根卓朗, 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, September 2017.

[2] T. Liebchen and Y. A. Reznik. Mpeg-4 als: an emerging standard for lossless audio coding. In *Data Compression Conference, 2004. Proceedings. DCC 2004*, pp. 439–448, March 2004.

[3] 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. SDM360^2: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.

[4] 菰原裕, 塚田学, 江崎浩, 曽根卓朗, 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.

[5] 仲谷正史, 筧康明, 南澤孝太, 三原聡一郎, 舘 [ススム]. 触感表現の一般普及に向けた方法論とテクタイルワークショップを通したその実践 (lt; 特集 gt; ハプティックコン

© 2018 Information Processing Society of Japan

テンツ). 日本バーチャルリアリティ学会論文誌, Vol. 19, No. 4, pp. 593–603, 2014.
[6] 塚田学, 菰原裕, 粕谷貴司, 新居英明, 高坂茂樹, 小川景子, 江崎浩. SDM360^2: インタラクティブ 3D コンテンツの自由視聴点再生. 情報処理学会論文誌 : デジタルコンテンツ (DCON), August 2018. (投稿中).
[7] ITUR Rec. Itu-r bs. 2051-0 (02/2014) advanced sound system for programme production. *Int. Telecommun. Union, Geneva, Switzerland*, 2014.
[8] E. Nakasu. 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, April 2012.
[9] 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, September 2005.
[10] Dolby Atmos© Specifications. Technical Report Issue 3, Dolby Laboratories, 2015.
[11] AUROMAX© Next generation Immersive Sound system. November 2015.
[12] J. Herre, J. Hilpert, A. Kuntz, and J. Plogsties. MPEG-h 3D audio —the new standard for coding of immersive spatial audio. Vol. 9, No. 5, pp. 770–779, August 2015.
[13] R. Bleidt, A. Borsum, H. Fuchs, and S. M. Weiss. Object-based audio: Opportunities for improved listening experience and increased listener involvement. *SMPTE Motion Imaging Journal*, Vol. 124, No. 5, pp. 1–13, July 2015.
[14] Michael A Gerzon. Periphony: With-height sound reproduction. *Journal of the Audio Engineering Society*, Vol. 21, No. 1, pp. 2–10, 1973.
[15] Mark A Poletti. Three-dimensional surround sound systems based on spherical harmonics. *Journal of the Audio Engineering Society*, Vol. 53, No. 11, pp. 1004–1025, 2005.
[16] 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, April 2012.
[17] 藤井竜也, 藤井哲郎, 小野定康, 白川千洋, 白井大介. ディ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・境界ソサイエティ Fundamentals Review, Vol. 5, No. 1, pp. 80–89, 2011.
[18] K. Hirata, Y. Harada, T. Takada, S. Aoyagi, Y. Shirai, N. Yamashita, K. Kaji, J. Yamato, and K. Nakazawa. t-room: Next generation video communication system. In *2008 IEEE Global Telecommunications Conference (GLOBECOM 2008)*, pp. 1–4, November 2008.
[19] 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.
[20] Carolina Cruz-Neira, Daniel J. Sandin, and Thomas A. DeFanti. Surround-screen projection-based virtual reality: The design and implementation of the cave. In *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH 1993, pp. 135–142, New York, NY, USA, 1993. ACM.
[21] 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, May 1998.
[22] 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, March 2003.
[23] 大久保洋幸, 大谷眞道, 小野一穂, 正岡顕一郎, 池沢龍, 小宮山摂, 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, September 2000.
[24] 大久保洋幸, 中山靖茂, 池永敏和, 小宮山摂. 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.
[25] 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.
[26] Mary E Dickinson, Ann M Flenniken, Xiao Ji, Lydia Teboul, Michael D Wong, Jacqueline K White, Terrence F Meehan, Wolfgang J Weninger, Henrik Westerberg, Hibret Adissu, et al. High-throughput discovery of novel developmental phenotypes. *Nature*, 2016.
[27] Yves Raimond, Samer A Abdallah, Mark B Sandler, and Frederick Giasson. The Music Ontology. *Proceedings of the International Conference on Music Information Retrieval*, pp. pp. 417–422, 2007.
[28] 仲間祐貴, 大西康伸, 位寄和久. 継続的利用と情報共有を可能にする建物維持管理支援のための bim を活用したウェブシステムの開発. 日本建築学会技術報告集, Vol. 22, No. 50, pp. 359–364, 2016.

© 2018 Information Processing Society of Japan