

主観視点全方位映像と三次元音場による高臨場感シーン再現

近藤 一晃[†] 向川 康博^{††} 池田 雄介^{†††} 榎本 成悟^{†††} 伊勢 史郎^{†††}
中村 哲^{†††} 八木 康史^{††}

[†] 京都大学学術情報メディアセンター 〒606-8501 京都市左京区吉田本町

^{††} 大阪大学産業科学研究所 〒567-0047 大阪府茨木市美保が丘 8-1

^{†††} 独立行政法人情報通信研究機構 知識創成コミュニケーション研究センター
〒619-0289 京都府相楽郡精華町光台 2-2-2

^{††††} 京都大学大学院工学研究科 〒615-8540 京都市西京区京都大学桂 C1-4-386

E-mail: †kondo@ccm.media.kyoto-u.ac.jp, ††{mukaigaw,yagi}@am.sanken.osaka-u.ac.jp,

†††{yusuke.ikeda,seigo.enomoto,satoshi.nakamura}@nict.go.jp, ††††ise@archi.kyoto-u.ac.jp

あらまし 映像・音声技術の発展に伴い、高臨場感提供を目的とする研究が数多くなされてきた。我々は収録時にあける登場人物視点からの周囲 360 度の映像と登場人物位置の音場を同時に記録し、狭範囲の客観呈示ではなく全方位の主観呈示を実現させることで、さらなる高臨場感・高迫力感覚の提供を目指す。本研究では、装着型の収録機材を用いた実シーンのコンテンツ収録、カメラ校正に基づいた映像処理および境界音場制御の原理に基づいた音場再現、複数人への同時呈示システムの開発などの一連の要素技術を提案する。さらに作成したコンテンツを多数の一般視聴者に対して上映し、システムの実証実験を行った。結果、本システムの有効性および主観全方位メディアによる高臨場感提供の一般的ニーズが確認された。

キーワード 主観視点, 全方位映像, 音場再現, 境界音場制御の原理, バーチャルリアリティ

High Immersive Scene Reproduction with First-person Perspective Omnidirectional Movies and Three Dimensional Sound Field

Kazuaki KONDO[†], Yasuhiro MUKAIGAWA^{††}, Yusuke IKEDA^{†††}, Seigo ENOMOTO^{†††}, Shiro ISE^{††††}, Satoshi NAKAMURA^{†††}, and Yasushi YAGI^{††}

[†] Academic Center for Computing and Media studies, Kyoto University
Yoshida honmachi, Sakyo-ku, Kyoto, 606-8501 Japan

^{††} The Institute of Scientific and Industrial Research, Osaka University
Mihogaoka 8-1, Ibaraki-shi, Osaka, 567-0047 Japan

^{†††} National Institute of Information and Communications Technology
Hikari-dai 3-5, Seika-cho, Souraku-gun, Kyoto, 619-0289 Japan

^{††††} Graduate School of Engineering, Kyoto University
Yoshida honmachi, Sakyo-ku, Kyoto, 606-8501 Japan

E-mail: †kondo@ccm.media.kyoto-u.ac.jp, ††{mukaigaw,yagi}@am.sanken.osaka-u.ac.jp,

†††{yusuke.ikeda,seigo.enomoto,satoshi.nakamura}@nict.go.jp, ††††ise@archi.kyoto-u.ac.jp

Abstract Provision of high immersive sensation to audiences have proceeded with growing up of techniques about video and acoustic medias. Its application expands to virtual experiences and simulated practices. In our proposal, we record and reproduce omnidirectional movies captured at a perspective of an actor and three dimensional sound field around him, and try to reproduct more impressive feeling. We propose a sequence of techniques to archive it, including a recording equipment, video and acoustic processing, and a presentation system. Effectiveness and demand of our system had been demonstrated by ordinary people through evaluation experiments.

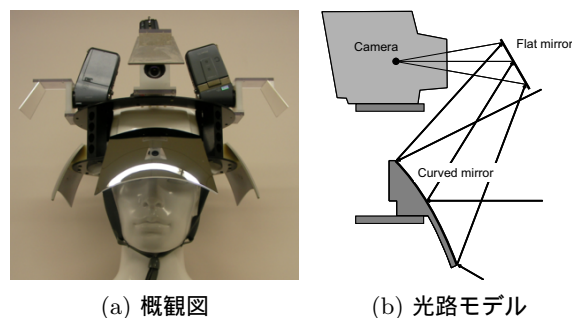
Key words First-person Perspective, Omnidirectional Vision, Three Dimensional Sound Reproduction

1. はじめに

高臨場感を伴ったシーンの再現は、仮想体験や疑似体験を通じ、アミューズメントや極地における体感シミュレータなど様々な用途に役立つ。映画館におけるコンテンツの上映がその代表例で、巨大スクリーンによる広視野呈示・サラウンドシステムによる立体音響再生が、視聴者にまるで対象シーンの中にいるかのような感覚を提供している。しかし、映像コンテンツの多くはシーンに対して客観的な第三者視点から撮影されたものである。客観的なカメラワークは視聴者がストーリーやシーンの構成を理解するのに効果的である一方、外側からシーンを俯瞰する形になるため臨場感や迫力感が損なわれている。それに対し、シーンに登場する人物の視点から映像を収録しそれを再現すれば、高い臨場感を視聴者に与えることができる。高い臨場感を与える手法としては自己の周囲 360° を覆う全方位映像の利用も効果的である。それら 2 つを組み合わせた主観視点全方位映像を利用することで、視聴者にまるでその場にいるかのような感覚を与えることができる。主観視点全方位映像を用いた仮想体験システムはこれまでにいくつか提案されてきた。山澤ら [10] は人物の頭部に全方位カメラを搭載することで、人物視点の全方位映像を獲得している。また尾上ら [7] や池田 [12] らは車両やジェットコースターといった動物体を用いて全方位撮影を行い、視聴者の四方を囲む全方位呈示系とともにテレプレゼンスシステムを開発している。近藤ら [15] は装着者視点と撮影系の視点が限りなく一致し、かつ追加の記録装置を必要としない独立型の収録装置を開発することで、主観視点全方位映像のより実用的な記録方法を提案している。

一方、立体音響再生も臨場感に対して大きな影響を及ぼす。シーンに対する音響コンテンツの再生においても、方向や距離など音源の位置情報も含めた忠実な 3 次元音場再現は臨場感を高める上で重要なファクターとなる。3 次元音場を忠実に再現する手法の一つとして、Berkhout らが提案する Wave Field Synthesis (WFS) [3] の研究開発が活発に行われており、既に実用段階に近づいている。しかし、WFS ではシステムを実現する段階で様々な近似と仮定を必要とし、実際の再現精度については十分に検証されていない。それに対して、榎本・池田らは伊勢の提案した境界音場制御の原理 [4] に基づく 3 次元音場再現システム (Bosc システム) を実現し、システムの性能を物理的・主観的に評価している [17] ~ [19]。

高臨場感を視聴者に与えるための上記 2 大要素はこれまで別々に扱われており、両者を統合したシステムの提案はほとんどなされていない。我々は主観視点の全方位映像と 3 次元音場再現を組み合わせることで、さらなる高臨場感呈示を目指す。収録機材の開発だけでなく、実際のコンテンツ収録、メディア処理、呈示システムの開発などの一連の要素技術の提案を通して、主観視点全方



(a) 概観図 (b) 光路モデル
図 1 装着型全方位カメラ FIPPO

位および 3 次元音場再現の有効性を検証する。特に評価実験では、約 1,100 名の一般視聴者を対象にデモンストレーションを行うことで構築したシステムの性能を実証した。

2. 装着型全方位収録機材

2.1 主観視点全方位映像撮影装置

本研究では、登場人物の主観視点における高品位な全方位映像を撮影するため、以下の性能を有する撮影装置が必要とされる。

- 高解像な全方位映像を実時間で撮影できること
- 装着者視点と撮影光学系視点が一致していること
- 可搬性が高く長時間撮影が簡単に行えること

第一項目を満たす、いわゆる全方位カメラに関する研究はこれまでも様々になされている [2], [5], [9], [11]。しかし、人物に装着することを目的とした全方位カメラはほとんど提案されておらず、人物視点からの全方位映像撮影は頭頂部に装着された一般的な全方位カメラで近似されることが多い。この場合、カメラ視点は装着者視点よりも上に位置することになり、第二項目を満たすことが困難である。また、全方位カメラ本体に加えて撮影映像の記録装置が別途必要であるため第三項目に対しても不適切である。向川ら [14] や東ら [16] が人物への装着を目的とした小型全方位カメラの提案を行っているが、上に述べた視点の不一致や解像度が低いといった問題が依然として残されている。一方、近藤らは主観視点の全方位映像を撮影を目的とした特殊な装着型全方位カメラである First-Person Perspective Omnidirectional camera (FIPPO) を提案している [15]。FIPPO はハンディタイプのデジタルビデオカメラと平面鏡・曲面鏡で構成された光学ユニットを 4 台組み合わせしており、高解像な全方位映像を装着者の視点から撮影することができる。また追加の記録装置や外部電源供給なしに長時間撮影が行える。これらの理由から、本研究における主観視点全方位映像の撮影に用いた FIPPO の概観およびスペックは図 1、表 1 に示すとおりである。

2.2 主観音場収録装置

主観音場を収録・再現する方法として、Head And Torso

表 1 FIPPO の性能表

視野角	360° × 俯角 27° ~ 仰角 37°
総解像度	3960 ~ 5020 × 584 ~ 682 pixels
装着との視点ズレ	50mm 未満
総重量	約 2.2kg
連続収録時間	約 85 分

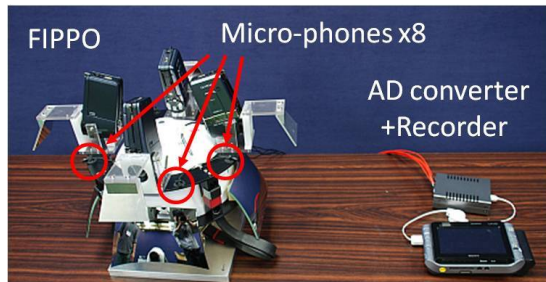


図 2 音場収録装置の構成

Simulator(HATS) や人(装着者)の耳の位置に付けたマイクロホンで音場を収録する方法がある。しかし、この方法では耳元 2 点の信号のみしか再現されないため、受聴時に自由に頭部や身体を動かすことが難しい。そこで本研究では、登場人物の位置からの全方位映像に合わせて登場人物の周囲の音場を受聴者が体験することが可能な Bosc システムを音場再現装置として用いる。

榎本らの開発した Bosc 収録システムは 70 チャンネルのマイクロホンアレーからなるため、装着者が自由に身体を動かしながら主観視点の音場を収録することは困難である。そこでチャンネル数を減らし構成を簡略化したシステムを「主観音場収録装置」として用いた。主観音場収録装置では、装着者の頭部の周囲の水平面内に 8 個の無指向性マイクロホン (ECM:DPA-4060) を設置して主観音場を収録する。マイクロホンは、装着者の両耳の高さに設置することが望ましいが、FIPPO のミラーとの干渉を防ぐため頭頂部のやや上の高さの水平面に設置した。主観音場収録装置と収録装置装着図をそれぞれ図 2、図 3 に示す。ECM は携帯型の小型 PC (SONY VAIO-typeU) の USB バスパワーで駆動する。また、音響信号は AD コンバーター (TD-BD-8CSUSB) を通して PC に記録されるため装着者の機動性への影響は非常に小さい。この小型 PC、及び AD コンバーターを使用することで、サンプリング周波数 48kHz、量子化ビット数 16bit の精度で 8 チャンネルの音響信号を同時に収録することが可能となる。

3. コンテンツ作成

3.1 収録内容

前章で述べた装着型全方位収録機材を用いて、大きく分けて 2 種のコンテンツ映像を収録した。まず日常生活で比較的良好に体験するシーンに着目した。視聴者はコンテンツ映像・音声による仮想体験と実体験とを容易に比

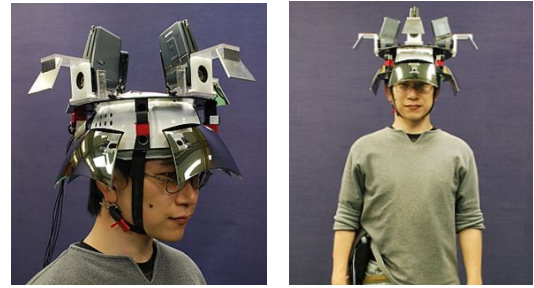
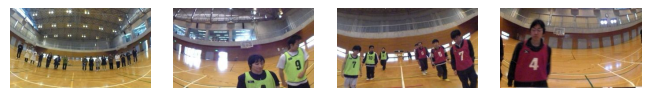


図 3 収録装置装着図



(a) 散策

(b) バスケットボール



(c) 後方

(d) 左方

(e) 前方

(f) 右方

図 4 コンテンツ収録の様子 (上段). 撮影された映像例 (下段).

較できるので、臨場感を相対的に評価するのに役立つ。収録は近郊の公園内で、紅葉・水辺・参道などの全方位映像や音の移動が効果的に現れるシーンを対象として行った。一方、普段あまり体感することのないシーンとしてスポーツゲームシーンを採り上げた。一般に、スポーツゲームはフィールドの外側に配置されたカメラの映像を用いて観戦される。本研究ではスポーツ選手が機材を装着して収録を行うことで、プレイヤー視点の仮想体験を試みた。具体的にはバスケットボールゲームにおける選手や審判の視点でコンテンツ素材を収録した。図 4 は実際に行った収録の様子および収録映像の例を示している。

3.2 映像処理

3.2.1 幾何歪みの補正

撮影された映像には幾何歪みが生じているため、視聴者に呈示するにはこの歪みを修正する必要がある。歪みは撮影機材である FIPPO が曲面鏡を用いて構成されているために生じているので、曲面鏡を含めた射影幾何を測定することで補正を行う。射影幾何は図 5 に示す専用のシーン構成を用いて測定した。FIPPO を大型の平面モニタの正面に設置し、モニタ上に表示された特別な幾何パターンを撮影する。幾何パターンをもちいることでモニタ上の二次元位置と画像上での画素位置が対応させることができる。誤対応を避けるため、幾何パターンには時系列で変化するグレイコードパターンを用いた。モニタと FIPPO との距離を変えて同様に測定を行うことで、各画素に対応するシーン中の光線が二次元平面上でどの位置を貫くかが決定するので、

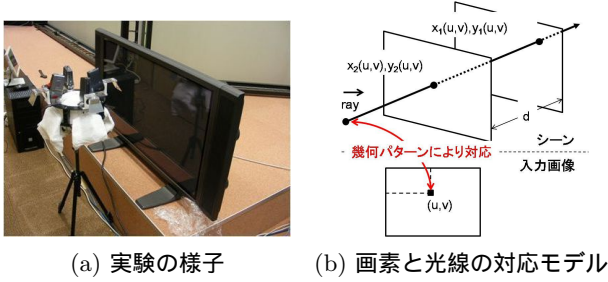


図 5 映像歪みの計測

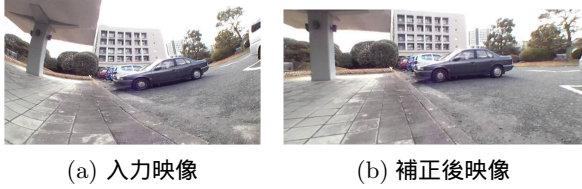


図 6 映像歪みの補正 (透視投影)



図 7 映像歪みの補正 (パノラマ). 図 4 の (c)-(f) を元に処理した結果.

$$\vec{\text{ray}}(u, v) = \begin{bmatrix} \vec{\text{p1}} - \vec{\text{p2}} \\ d \end{bmatrix} = \begin{bmatrix} x_1(u, v) - x_2(u, v) \\ y_1(u, v) - y_2(u, v) \\ d \end{bmatrix} \quad (1)$$

のように画素とシーン中の光線の対応を得ることができる． (u, v) は画像上の画素位置を， $\vec{\text{ray}} = [r_x, r_y, r_z]^t$ はシーン中の光線を， $\vec{\text{p1}} = [x_i, y_i]$ はモニタ上での三次元位置を示す．また d は FIPPO とモニタとの距離を変えて計測を行ったときの奥行き差である．この射影測定法では光線の向きは決定されるが，位置を決定することはできない．しかし FIPPO 全体がほぼ単一視点な構成であることから，近似的に FIPPO の中心を光線端点として位置決定を行った．画素とシーン中の光線の対応が決定すれば，シーン中に仮想スクリーンを作成して撮影された画像を逆投影することで歪み補正を行うことができる．逆投影は OpenGL を用いて画像を仮想二次元面へテクスチャマッピングすることで実装した．通常我々が目にする歪みのない平面 (透視投影画像) へ歪み補正を行った例を図 6 に示す．シーン中で直線だった部分が正しく直線に戻されていることが確認できる．また 4 面で構成される全方位映像を 1 枚のパノラマ映像に繋いだ例を図 7 に併せて示す．

3.2.2 色空間の補正

映像撮影機材で使われている 4 台のビデオカメラは全く同一の型番だが個体差により各々の収録映像は若干色味が異なる．そのためカメラ間の色補正も必要である．本研究では各ビデオカメラの色空間の間にアフィン変換が成立すると仮定して，カメラ間の色空間を変換する補



図 8 カメラ間の色補正. 異なるカメラで撮影されたカラーチャート (a) を元に，それらのカメラの映像境界部分 (b) で色補正を行った結果 (c).

正パラメータを算出した．全く異なる撮影機器間の色変換ではアフィン変換による近似精度は低い，今回は同一型番のカメラ間の補正であるので十分に高い近似精度が期待できる．アフィン変換の仮定に基づけば，シーン中で同一の色物体を異なるカメラで撮影した場合，各々の画像における RGB 値が 3×4 の行列によって線形に関係づけられる．すなわち

$$\begin{bmatrix} R_m \\ G_m \\ B_m \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} R_n \\ G_n \\ B_n \\ 1 \end{bmatrix} \quad (2)$$

が成り立つ．ここで $[R_k, G_k, B_k]^t$ はカメラ k 画像における RGB 色を， $p_{ij} (i = 1, 2, 3, j = 1, 2, 3, 4)$ はアフィン変換のパラメータを示す．1 つの色対応により 3 本の線形式を立式できるのに対して，求めたい未知数は 12 個であることから，すくなくとも 4 色以上の対応があれば補正パラメータを決定することができる．コンテンツ収録時には色画像処理で一般的に用いられているカラーチャート (Macbeth 社製) を撮影することで 24 色の色対応を得て最小二乗法にて最適補正パラメータ p_{ij} を算出した．図 8 に 2 台のカメラ間の色補正結果を示す．補正前は明らかに異なっていた色味が，ほとんど区別できなくなる程度まで補正されているのがわかる．

3.3 Bosc システムによる音場再現

3.3.1 境界音場制御の原理

Kirchhoff-Helmholtz 積分方程式によると，空間内のある仮想的な境界面上の音圧と音圧勾配を制御することにより境界に囲まれた領域の音圧を制御することができる．しかし，この場合境界面上に monopole 音源と dipole 音源を並べる必要があり，実際のスピーカでは近似的にしか実現することができない．また，再現音場が自由音場でなければならないという制約がある．境界音場制御の原理 [4] は Kirchhoff-Helmholtz 積分方程式と多チャンネル逆システムを組み合わせることにより，これらの近似や制約を取り除くことができることを示す原理である．三次元音場再現システムに応用する場合には，3 次元音場内の任意の点にマイクロホンを設置し，その点で計測された音圧を別の場所で再現することによりマイクロホ

ンによって囲まれた領域内部の音場を厳密に再現することができる。したがって、一般的なトランスオーラルシステムやバイノーラルシステムとは異なり、領域内部では受聴者が身体を動かしても元の音場と矛盾しない音場を体験することができる。

3.3.2 逆システムの設計

境界面上で計測された音圧を再現する際に必要となる逆システムの設計方法について述べる。ここで、音圧を制御するスピーカを2次音源、再現すべきマイクロホン制御点という。

2次音源数を M 、制御点数を N とし、 i 番目の音源から j 番目の制御点までの周波数伝達特性を $G_{ji}(\omega)$ 、原音場で収録された信号を $X_j(\omega)$ 、音源から出力される信号を $Y_i(\omega)$ 、制御点で観測される信号を $Z_j(\omega)$ とすると音場再現システムの入出力の関係は式 (3) のように書ける。

$$\mathbf{Z}(\omega) = [\mathbf{G}(\omega)]\mathbf{Y}(\omega) = [\mathbf{G}(\omega)][\mathbf{H}(\omega)]\mathbf{X}(\omega) \quad (3)$$

ただし、

$$\begin{aligned} \mathbf{X}(\omega) &= [X_1(\omega), \dots, X_N(\omega)]^T \\ \mathbf{Y}(\omega) &= [Y_1(\omega), \dots, Y_M(\omega)]^T \\ \mathbf{Z}(\omega) &= [Z_1(\omega), \dots, Z_N(\omega)]^T \\ [\mathbf{G}(\omega)] &= \begin{bmatrix} G_{11}(\omega) & \cdots & G_{1M}(\omega) \\ \vdots & \ddots & \vdots \\ G_{N1}(\omega) & \cdots & G_{NM}(\omega) \end{bmatrix} \\ [\mathbf{H}(\omega)] &= \begin{bmatrix} H_{11}(\omega) & \cdots & H_{1N}(\omega) \\ \vdots & \ddots & \vdots \\ H_{M1}(\omega) & \cdots & H_{MN}(\omega) \end{bmatrix} \end{aligned}$$

である。音場再現システムにおける逆フィルタ設計の目的は $\mathbf{Z}(\omega) = \mathbf{X}(\omega)$ となる $[\mathbf{H}(\omega)]$ 、すなわち $[\mathbf{G}(\omega)]$ の逆フィルタ $[\mathbf{H}(\omega)]$ を定めることである。 $\mathbf{X}(\omega)$ に含まれる微小な誤差や伝達系 $[\mathbf{G}]$ の変動に対して $\mathbf{Z}(\omega)$ の値を大きく変動させる逆フィルタ $[\mathbf{H}(\omega)]$ を不安定であるという。本稿では、不安定性を緩和するパラメータを連続的に変化させることが可能な正則化を用いて逆フィルタの設計を行った。正則化を用いた場合、逆フィルタ $[\mathbf{H}(\omega)]$ の推定値 $[\hat{\mathbf{H}}(\omega)]$ は

$$[\hat{\mathbf{H}}(\omega)] = ([\mathbf{G}(\omega)]^\dagger [\mathbf{G}(\omega)] + \beta \mathbf{I}_M)^{-1} [\mathbf{G}(\omega)]^\dagger \quad (4)$$

と書くことができる。ただし、 $\dagger$ は共役転置、 β は正則化パラメータ、 $\mathbf{I}_M$ は M 次の単位行列を表す。

4. 全方位呈示装置

4.1 映像呈示装置

視聴者の前面に設置されたモニタに映像を呈示する方法では、全方位映像が持つ折角の臨場感が損なわれてしまう。そこで視聴者に対して広視野の映像を呈示することで、撮影者主観の高臨場感再現を試みた。広視野映像

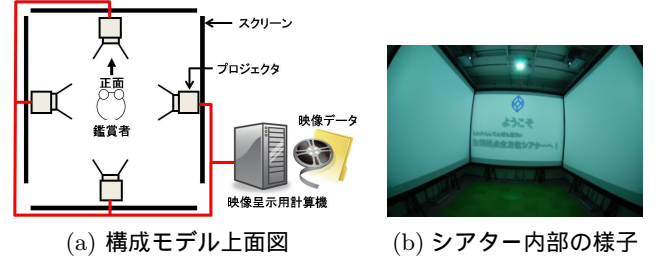


図 9 全方位映像呈示装置

を呈示することで臨場感を向上させたり、作業を効率化させるアプローチはこれまでも数多くなされてきた。長原ら [13] や橋本ら [8] は個人観賞を目的とした全方位映像呈示装置を提案している。また尾上ら [7] や池田ら [12] は、移動ロボットの誘導や作業を効率よく行うための広視野プレゼンスを行っている。本研究では全方位映像呈示でも複数視聴者の同時体験を目的とし、CAVE [1] のように平面スクリーンで四方が囲われるような全方位呈示装置を試作した。図 9 に試作した呈示装置である主観視点全方位シアターのモデル図および内部の様子を示す。映像投影用のスクリーン (横 3.0m、高さ 2.0m) は視聴者の前後左右に 1 枚ずつ系 4 枚配置されており、後部スクリーンは出入のために左右スクリーンから間隔を空けて配置されている。スクリーンはアルミニウム製のフレームで構成された部屋状の構造 (横 3.6m、長さ 4.8m、高さ 3.4m) に取り付けすることで固定されている。内部スペースはおよそ 3.0m 四方となっているため、数人の視聴者が同時に全方位映像を体験することができる。シアター天井部には 1 枚のスクリーンに対して 1 台のプロジェクタが設置されており、各々が正面投影で 1024×768 pixel の映像を呈示する。背面投影で構成することも可能だが、音場再現のためのスピーカ (4.2 節) が影になってしまうので本研究では正面投影を採用した。プロジェクタの投影角や視聴者の身長などを考慮し、人物の影が映像に干渉しないように設計されている。斜め投射によって生じる映像ひずみは台形補正を用いて対応した。また単一の計算機により映像投影を制御することで 4 面映像の同期呈示を単純化させている。映像と音声との同期再生法については 4.3 節にて述べる。

4.2 音場再生装置

主観音場収録装置で収録した音場の再現は視聴者の周囲に並べた複数のスピーカにより行う。本シアターでは、8 台のスピーカを使用することとした。スピーカの配置を図 10 に示す。穴あき加工を施したサウンドスクリーンを使用し、スピーカは各スクリーンの背後に 2 台ずつ設置している。3.3.2 節で述べた逆システムを設計するために、実際に装着者が身につけたマイクロホンアレーと同じ配置のアレーをシアター内に設置してスピーカからマイクロホンまでのインパルス応答を測定した。インパルス応答長は 1024 点とし、 2^{16} 点の Swept-sine 信号

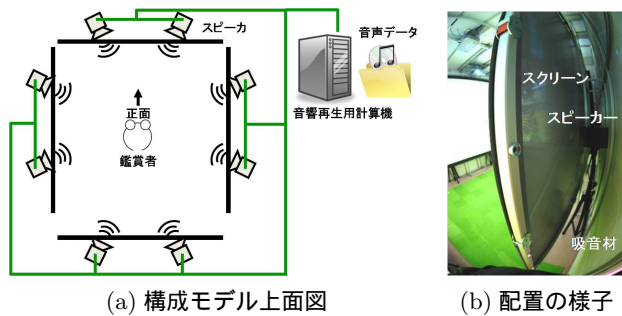


図 10 三次元音響再現装置

を 8 回の同期加算を行い測定した．サンプリング周波数は 48 kHz，量子化ビット数は 24 bit である．逆フィルタ長は 4096 点として式 (4) に従い計算した．計算された逆フィルタを畳み込んだ信号を各スピーカから出力することによりマイクロホン内部の領域の音場が再現される．

Bosc システムでは，厳密には，マイクロホンアレーに囲まれた領域の内部のみが再現エリアとなる．しかしながら，マイクロホンを近接して配置することにより，その周辺を含む広い範囲において原音場に近い音場が生成されることが期待される [18]．また，逆フィルタの計算はサウンドスクリーンによる減衰を補正するのみではなくシアター内部の音響特性も補正しているため，より忠実な音場の再現が可能であると考えられる．なお，逆システムの計算を容易にするため，本シアターでは天井面とスクリーン外側の壁面に吸音パネルを設置し，床面にはカーペットを敷いた．

4.3 映像と音場の同期再生

全方位映像撮影機材における 4 台のカメラ間の同期，及び全方位映像と音場収録装置の同期はコンテンツ収録時に行うことができないため，収録終了後に同期処理を行う必要がある．映像コンテンツと音響コンテンツの同期にはカメラに内蔵されたマイクロホンで収録された音響信号との相互相関値を計算することによって行った．収録時に同期を取ることができるよう，SMPTE タイムコード等の外部信号を入力することも検討したが，専用の機材が必要となり装着者の機動性を阻害すると考え，本システムでは採用しなかった．

映像・音場の再生システムの構成を図 11 に示す．図のように，音響再生装置はスピーカから再生する 8 トラックの音響信号に加え，LTC 信号を出力する．映像再生装置は LTC 信号を受け，音響データに同期して映像のデコード，及び再生を行う．このような構成にすることにより，コンピュータに過度の処理能力がなくとも全方位映像，及び音場の再生が可能となる．さらに，再生する映像投影系の数が増えた場合にも LTC 信号に同期可能な装置をもう一式用意するだけで良く拡張性も高い．

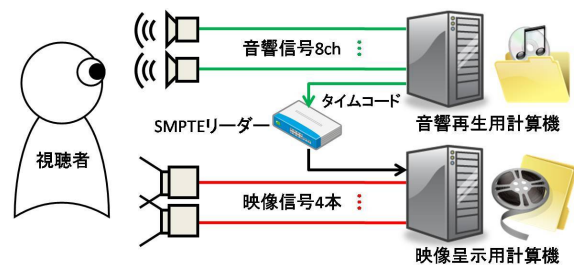


図 11 映像と音声の同期再生

5. 実 験

5.1 実験の目的

開発した主観視点全方位シアターで収録した映像・音声コンテンツを体験し，主観視点の全方位映像と三次元音場再現の客観的な評価を行う．特に一般視聴者を対象に実験を行うことで，提案システムが新映像技術として社会的かつ一般的なニーズを満たしているかを検証する．

5.2 実験内容

本システムは「新映像技術ダイブイントゥザムービーの研究」の研究成果の一部として，一般人を対象としたアウトリーチ活動 (平成 21 年 3 月 20～22 日，日本科学未来館にて) で展示・公開された．評価実験はその時に来場した一般視聴者を対象に実施した．三日間のイベント期間中に 1100 人を超える来場者に対してコンテンツを上映し，およそ 750 人分の有効アンケート結果を得た．友人・カップル・家族連れなど幅広い年齢層・客層であったことからより一般的で信頼性の高い評価が期待できる．

コンテンツの体験前に，視聴者には本研究の目的・研究機材・収録方法・シアター構成などについて簡単に解説を行った．その後，会場に設営したシアター内で 3.1 節で述べたコンテンツを上映した．本研究についての解説とコンテンツの体験は来場したグループ単位で行った．1 グループあたりの解説時間および上映時間はどちらもおよそ 3 分である．上映後，アンケートを実施し，その結果を元に本研究の有効性を評価した．アンケート内容の詳細を表 2 に示す．アンケートは各項目の五段階批評と自由筆記形式の大きくわけて 2 つで，前者は主に臨場感に関する設問，後者は本研究が持つ課題や応用例に関する設問である．

5.3 結果と考察

臨場感や迫力に関する五段階批評結果を図 12 に示す．図中では評価値が 5 に近いほど高評価であることに対応する．大多数の体験者が高い主観の臨場感を体感しており，本研究の目的である，迫力のある新映像技術の提案の有効性が実証された．一方で画質の項目は評価が低い．評価の低かった原因として大きく二つの理由が考えられる．1 つ目の原因は撮影機材である FIPPO の問題で，鏡

表 2 一般視聴者に対して実施したアンケート調査の設問内容

■ 臨場感・迫力感に関する設問 (五段階批評)	
A. 自分がその場にいるかのような感覚 (臨場感) を得られましたか？	1. 全く得られない 2. 得られない 3. 普段通り 4. 得られた 5. とても得られた
B. 前方カメラのみの映像と比べ、迫力のある映像でしたか？	1. 全く迫力がない 2. 迫力がない 3. 普段と同じ 4. 迫力がある 5. とても迫力がある
C. 映像の画質はどうでしたか？	1. 非常に低い 2. 低い 3. 普通 4. 高い 5. 非常に高い
D. 前方スピーカのみのお音響と比べ、迫力のある音響でしたか？	1. 全く迫力がない 2. 迫力がない 3. 普段と同じ 4. 迫力がある 5. とても迫力がある
E. 音質はどうでしたか？	1. 非常に低い 2. 低い 3. 普通 4. 高い 5. 非常に高い
■ 自由筆記形式の設問	
F. 体感してみたいシーンがあればお聞かせください。	
G. より臨場感を高めるために必要だと思われる要素や機能があればお聞かせください	

による二回反射で光量が落ちるために画質が多少劣化していることである。これは鏡面を高反射率素材で作成することで解決できる。2つ目の原因は対面して設置されたスクリーン間の相互反射が引き起こしている全体のコントラスト低下である。グレー色のスクリーンや再帰性反射素材を用いたスクリーンを使用するなどの解決方法が考えられる。

自由筆記形式の設問に記述されていた主な回答例を表3に示す。第一の設問に対しては、スポーツ・アトラクション・日常生活といったジャンルに分類される数多くのシーンが回答として寄せられた。特に普段体験することができない(体験することが難しい)シーンに対する要望が多かったことから、主観かつ全方位の仮想体験ニーズが一般視聴者からも求められていることが確認された。第二の設問では、やはり揺れに対する補正が圧倒的に多い結果となった。「少し気分が悪くなった」「酔った」等の感想を視聴者から直接聞くこともしばしばであった。映像揺れが視聴者に与える悪影響については我々も事前から認識しており、できるだけ頭部を揺らさないように収録を行った。しかしそれでは不十分であったことが今回の評価から明らかになった。主観視点の映像を(揺れも含めて)忠実に収録・再生している、という点で正しい評価を得られたと言えなくもないが、視聴者への呈示を考慮するとやはり適切な揺れ補正が必要である。対策としては、別途ジャイロセンサなどを用いることで装着者の頭部姿勢も記録したり、全方位映像の周期性を活用して揺れを補正する[6]などが考えられる。また、スクリーンを前後左右の4面構成ではなく、円筒形にすることを望む声も多かった。理由としては主に「4面スクリーンでは映像中と実際の遠近感が異なる」「スクリーンどおしの繋ぎ目が気になる」などが挙げられていた。特に遠近感に関しては「立体感が欲しい」「正しい距離感を感じたい」等の声も多く、高臨場感の仮想体験を実現するにはやはり奥行き知覚の知覚が大きな要素となることが確認された。

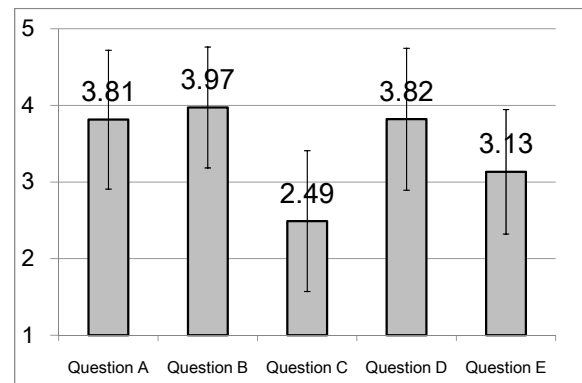


図 12 五段階批評の結果。各設問で得た評価値の平均を標準偏差とともに示している。評価値は5に近いほど高評価となる。

6. ま と め

本研究では、シーンにおける登場人物視点(主観視点)からの周囲360度の全方位映像および三次元音場を収録・再現することで、高臨場感の仮想体験システムの提案を行った。映像および音声コンテンツは、主観視点の全方位映像と三次元音場を記録できる装着型の特殊機材を用いて収録した。さらに視聴者を囲むように4面のスクリーンと8台のスピーカを配置し、収録コンテンツを再生する主観視点全方位シアターを開発した。多数の一般人視聴者がイベント展示を通して提案システムを体験し、そのアンケート結果から高い臨場感と迫力を与えるシステムであることが確認された。一方で、映像揺れの補正・画質と音質の向上・シーンの奥行き感を出す等の課題も明らかになった。今後は指摘された問題点について改良を加え、さらに高品質な主観視点全方位の仮想体験システムを構築する計画である。

謝 辞

本研究は、文部科学省科学技術振興調整費「重要課題解決型研究等の推進」プログラムにおける平成18年度

表 3 自由筆記形式の設問で得られた主な回答例．各カテゴリは我々の研究グループが独自にまとめたもの．

設問 F に関する主な回答例	
a. 日常生活	- 人混みの中, バーゲンセール, 街中や森林の散歩
b. スポーツ	- サッカー, 野球 (バッター・ピッチャー), テニス - ボクシング, スキー, サーフィン, バンジージャンプ - プロレス, スカイダイビング, スキューバダイビング - アメリカンフットボール, カヌー
c. 乗り物	- 自転車, ジェットコースター, オートバイ, 観覧車
d. 極地体験	- 海中, 空中, 宇宙空間
e. エンターテインメント	- ホラー映画, お化け屋敷, 洞窟探検 - コンサート歌手視点, テーマパークキャラクター視点
f. 動物	- 犬の視点, ライオンに食べられる
設問 G に関する主な回答例	
a. スクリーン構成	- 円筒スクリーン, 8 面スクリーン, ドームスクリーン - 4 面映像間の隙間をなくす, 天井部スクリーンの追加
b. 奥行き感の再現	- 正しい距離感を感じたい, 立体感が欲しい
c. 画質・音質	- 画質・解像度・音質の向上 - スクリーンが明るくてコントラストが低下している
d. 揺れの補正	- 酔ってしまった, 少し頭が痛くなった, フラフラした - 頭を動かさないようにして収録してほしい

採択課題「新映像技術ダイブイントゥザムービーの研究」により行われた．

文 献

- [1] C. Cruz-Neria, D. J. Sandin, and T. A. DeFanti, "Surround-Screen Projectorion-Based Virtual Reality: The Design and Implementation of the CAVE", Proc. of International Conference on Computer Graphics and Interactive Techniques(SIGGRAPH) 1993, pp. 135-142, AnaHeim, USA 1993.
- [2] K. Yamazawa, Y. Yagi, and M. Yachida, "Omnidirectional Imaging with Hyperboloidal Projection", Proc. of IEEE/RSJ Int. Conf. on Intelligent Robotics and Systems(IROS), Vol. 2, pp. 1029-1034, Tokyo, Japan, July 1993.
- [3] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis", J. Acoust. Soc. Am. Vol. 93, Issue 5, pp. 2764-2778, May 1993.
- [4] 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(9), pp. 706-713, 1997.
- [5] S. K. Nayar, "Catadioptric Omnidirectional Camera", Proc. of Int. Conf. on Computer Vision and Pattern Recognition(CVPR), pp. 482-488, San Juan, Puerto Rico, June 1997.
- [6] J. gluckman and S. K. Nayar, "Egomotion and Omnidirectional Cameras", Proc. of Int. Conf. on Computer Vision (ICCV1998), pp. 999-1005, Bombay, India, Jan. 1998.
- [7] Y. Onoe, K. Yamazawa, and N. Yokoya, "Telepresence by real-time view-dependent image generation from omnidirectional video streams", Computer Vision and Image Understanding(CVIU), Vol. 71, No. 2, pp. 154-165, Aug. 1998.
- [8] W. Hashimoto and H. Iwata, "Ensphered vision: Spherical immersive display using convex mirror", Trans. of the Virtual Reality Society of Japan, Vol. 4, No. 3, pp. 479-486, March 1999.
- [9] H. Tanahashi, K. Yamamoto, C. Wang, and Y. Niwa, "Development of a Stereo Omni-directional Imaging System (SOS)", Proc. of IEEE Int. Conf. on Industrial Electronics, Control and Instrumentation, pp. 289-294, Nagoya, Japan, Oct. 2000.
- [10] K. Yamazawa, and H. Takemura, and N. Yokoya, "Telepresence system with an omnidirectional HD camera", Proc. of Fifth Asian Conference on Computer Vision(ACCV2002), Vol. II, pp. 533-538, Melbourne, Australia, Jan. 2002.
- [11] Y. Yagi, and M. Yachida, "Real-Time Omnidirectional Image Sensors", International Journal of Computer Vision(IJCV), Vol. 58, No. 3, pp. 173-207, July 2004.
- [12] S. Ikeda, T. Sato, M. Kanbara, and N. Yokoya, "Immersive telepresence system with a locomotion interface using high-resolution omnidirectional videos", Proc. IAPR Conf. on Machine Vision Applications(MVA), pp. 602-605, May 2005.
- [13] 長原一, 八木康史, 谷内田正彦, "屈性反射光学系を用いた広視野角ヘッドマウントディスプレイ", 電子情報通信学会論文誌, Vol. J88-DII, No. 1, pp. 95-104, Jan. 2005.
- [14] 向川康博, 西岡潔郁, 八木康史, "複眼全方位センサを用いた装着型防犯アラームの開発", 第 9 回, 画像の認識・理解シンポジウム (MIRU2006), pp. 722-727, 淡路島, July 2006.
- [15] 近藤一晃, 向川康博, 八木康史, "主観視点全方位映像撮影のための装着型システムの開発", 第 11 回, 画像の認識・理解シンポジウム (MIRU2008), pp. 1656-1661, 軽井沢, July 2008.
- [16] H. Azuma, Y. Mukaigawa, and Y. Yagi, "Spatio-Temporal Lifelog Using a Wearable Compound Omnidirectional Sensor", Proc. of the Eighth Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras (ONIVIS2008), Marseille, France, Oct. 2008.
- [17] S. Enomoto, Y. Ikeda, S. Ise, and S. Nakamura, "Three-dimensional sound field reproduction and recording system based on the boundary surface control principle", The 14th Int. Conf. on Auditory Display, pp. o-16, Jun. 2008.
- [18] 榎本成悟, 池田雄介, 伊勢史郎, 中村哲, "境界音場制御の原理を用いた音場再現システムにおける再現領域の物理的な評価", 日本音響学会講演論文集, pp.855-856, March 2008.
- [19] 池田雄介, 榎本成悟, 伊勢史郎, 中村哲, "多チャンネル三次元音場再現システムの再生方式の違いによる再現性能評価", 電子情報通信学会技術研究報告, Vol. 108(333), pp.77-82, Dec. 2008.