基礎論文

視覚・聴覚の現実と仮想を融合する 2×2 方式複合現実感システムの実現

比嘉 恭太[*1], 西浦 敬信[*1],
木村 朝子[*2], 柴田 史久[*1], 田村 秀行[*1]

A Two-by-Two Mixed Reality System That Merges Real and Virtual Worlds in Both Audio and Visual Senses

Kyota Higa[*1], Takanobu Nishiura[*1],
Asako Kimura[*2], Fumihisa Shibata[*1], and Hideyuki Tamura[*1]

Abstract － In virtual reality, there have been many implementations using both audio and visual senses. However, mixed reality (MR), which merges the real and virtual worlds in real time, has thus far dealt with these implementations using only the visual sense. In this study, we developed an MR system that merged the real and virtual worlds in both the audio and visual senses, and in which the geometric consistencies of the audio and visual senses were coordinated. We also tried two approaches for merging real and virtual worlds in the audio sense, one using open-air headphones, and the other using closed-air headphones. The former corresponded to an optical see-through method and the latter, in the case of the visual sense, to a video see-through method.

Keywords: Mixed Reality, Audio and Visual Senses, Geometric Consistency, Open-Air Headphones, Closed-Air Headphones

1. はじめに

現実世界と仮想世界を実時間で融合する複合現実感 (Mixed Reality; MR) は急速な発展を遂げ, 人工現実感 (Virtual Reality; VR) 研究の中でも活発な分野の 1 つとなっている[1][2]. 人工的に生成した仮想環境だけを扱う従来の VR に対して，MR では目の前の現実世界に電子的なデータを融合させて提示できることから，新しい情報提示技術としての期待が大きい．また，実物の道具，体験者の手足，共同作業者の姿が視認できることから，これまで（狭義の）VR で実現された対象も, MR で実現することにより実用性が増している．

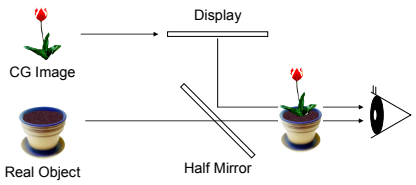
VR 分野では既に，3D-CG による視覚提示と 3 次元音場による聴覚提示を同時に行う研究が多数存在する[3][4][5]．これに対して，現存する MR システムの大半は視覚的な融合に主眼が置かれていて，音や触覚は付加的な演出効果として利用されているだけであった[6][7]. 最近では, 拡張現実感 (Augmented Reality; AR) システムや MR システムに 3 次元音場を導入する研究事例[8][9][10]も存在するが，これらは視覚的 MR＋聴覚的 VR を提示しているにすぎない．即ち，現実世界の音刺激と仮想世界で生成された音が実時間で矛盾無く融合する聴覚的 MR を達成しているわけではない．

我々が目指し，かつ他に先駆けて達成したのは，視覚的 MR と聴覚的 MR を扱い，かつそれらが同時提示される MR システムである. このシステムは「現実と仮想の融合」と「視覚と聴覚の併用」の両方を達成していることから，「2×2（ツー・バイ・ツー）方式の複合現実」と呼ぶことにした．そのようなシステムを実現するためには，

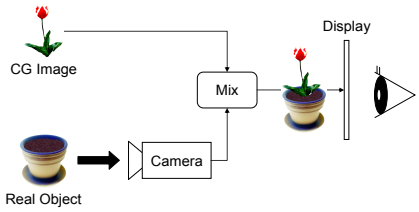
(1) 視覚的には現実世界の光景と電子的に生成した CG 映像の重畳合成
(2) 聴覚的には現実世界で発した実音とコンピュータで生成した人工音（生成音）の混合
(3) 上記の映像と音像の複合現実空間における実時間幾何学的整合性

を達成する必要がある．

*1 立命館大学大学院理工学研究科
*1 Graduate School of Science and Engineering, Ritsumeikan University
*2 科学技術振興機構 さきがけ
*2 PRESTO, Japan Science and Technology Agency



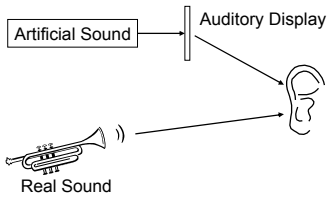
(a) 光学シースルー



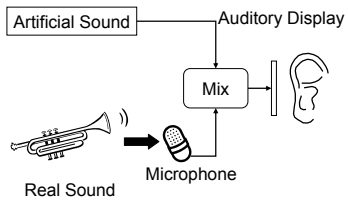
(b) ビデオシースルー

図 1 2 つのシースルー方式

Fig. 1 Two see-through methods



(a) 開放型音独立提示方式



(b) 密閉型混合音提示方式

図 2 2 つの音像提示方式

Fig. 2 Two methods of sound image presentation

(1) は既に数多くの研究や応用事例があり，筆者らのグループでも過去 10 数例の開発実績がある．このため，本研究では (3) の達成を第一義に考え，(1) と調和がとれた実装形態と精度での (2) の方式を検討し，全体として違和感なく体験できる「視覚・聴覚併用型 MR システム」を目指すことにした．

(1) は体験者がシースルーHMD (Head Mounted Display) を装着して MR 空間を移動体験するのが一般的であることから，(2) の混合結果は据置のスピーカでなく，ヘッドホンを装着して音像提示する方式を採用する．この場合でも，現実（実音）と仮想（人工音）は，開放型ヘッドホンを利用する場合と密閉型ヘッドホンを利用する場合の 2 つの場合が考えられる．これは視覚的融合における「光学シースルー方式」と「ビデオシースルー方式」に相等する[11]．本研究では，聴覚における両方式を実際に試し，その結果を考察した．

以下本論文では，2 章で視覚的 MR，聴覚的 MR それぞれの提示方法と両者をバランス良く同時提示できるシステムの構成を，3 章でヘッドホンの特性の違いによる音響的整合性の実験とその結果を，4 章で視聴覚での 2×2 方式幾何学的整合性の達成方法を，5 章で視聴覚を併用した MR システムの具体的な実現とその運用結果に関して述べる．

2. 視覚と聴覚の複合現実感

2.1 視覚の複合現実感

HMD を用いた視覚的な複合現実感の提示方法としては，図 1 に示した 2 つの方式に大別できる．図 1 (a) はハーフミラー機能を有した光学的映像合成で，いわば網膜上で 2 つの映像が合成される．図 1 (b) は，両眼を代行する一対のビデオカメラを遮蔽型 HMD に付加（もしくは内蔵）した構成で，カメラでキャプチャした現実世界の映像にコンピュータで生成した仮想物体を重畳描画し，眼前のディスプレイに表示する．それぞれに長短所があるが，光学的整合性（画質合わせ）には図 1 (b) が勝っているので「ビデオシースルー方式」が用いられることが多く，本研究でも同方式の HMD を用いる．

2.2 聴覚の複合現実感

コンピュータで生成した 3 次元音場の提示には，ヘッドホンを利用するバイノーラル方式と複数のスピーカを利用するトランスオーラル方式がある．我々は同時に複数人が MR 空間体験することを想定しており，トランスオーラル方式ではその実現が容易ではない．本研究では，視覚的 MR との調和と簡便に実現できることを重視して，ヘッドホンを用いて各人に 3 次元音場を提示するバイノーラル方式を採用する．ヘッドホンに磁気センサを装着し，頭部位置姿勢の変化に応じて聴覚的 VR を精度よく提示できることは既に報告されている[12]．

現実と仮想の聴覚的な融合にも，視覚における HMD と同様，図 2 に示すような 2 方式が考えられる．図 2 (a) は開放型ヘッドホンを用いる方式で，人工音はヘッドホンから提示されるが，ヘッドホンと耳の間の隙間から実音を聞くことができる．これは視覚における光学シースルー方式に相等する．図 2 (b) は，密閉型ヘッドホンの外側にマイクロホンを装着する方法で，マイクロホンで受音した実音と人工音の混合結果がヘッドホンから提示される．これは視覚におけるビデオシースルー方式に相等する．

本研究では，この両方式をそれぞれ「開放型音独

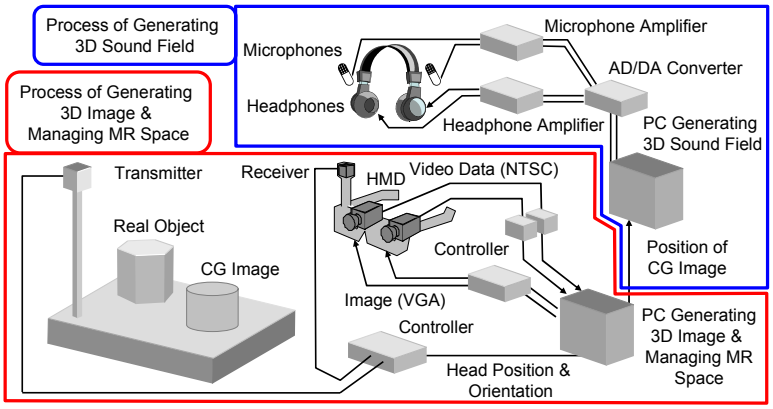
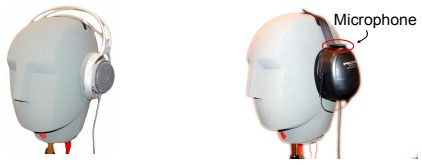


図 3 システム構成

Fig. 3 System configuration



(a) 開放型ヘッドホン (SONY MDR-F1)　(b) 密閉型ヘッドホン (PELTOR HTM79B-S)

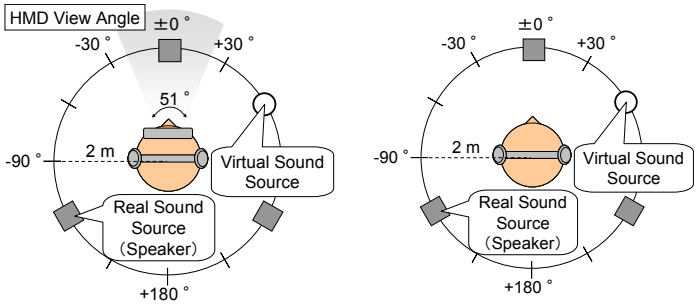
図 4 ヘッドホンの外観

Fig. 4 Exteriors of headphones

表 1 通信データ

Table 1 Transmittable data

Send Data	The information of the data
Object ID	ID number of CG Image
Sound ID	ID number of sound data
Relative Distance [m]	Relative distance between user and CG Image
Relative Direction [deg]	Relative direction between user and CG Image



(a) 視聴覚併用　(b) 聴覚のみ

図 5 実音と人工音の聴取

Fig. 5 Reception of real and artificial sound in Ex. 1

立提示方式」「密閉型混合音提示方式」と呼ぶことにした．

2.3 視聴覚の調和がとれた複合現実感

人が外界から受ける全ての情報のうち，視覚を通じて受ける情報は全ての感覚器官から受ける情報の 87%を占めると言われている[13]．このため，「視覚・聴覚併用型 MR システム」を考える場合も，体験者の主感覚は視覚であり，聴覚はそれを補う副感覚とするのは自然な考えである．即ち，視覚的 MR と聴覚的 MR を時間的，幾何学的に矛盾なく提示することを目指すが，聴覚的 MR が視覚的 MR の精度を超える必要はなく，両者調和のとれた方式とする．

視覚的 MR には，筆者の 1 人（研究リーダー）が既に開発した方式（[14][15]など）を踏襲するため，本論文ではその内容に詳しくは触れない．要点は，シースルー型 HMD を装着し，幾何学的整合性に関しては，物理的位置姿勢センサの精度不足を，必要に応じてマーカ認識法で補うという方法であり，他の類似の方法でも目的は達成できる．

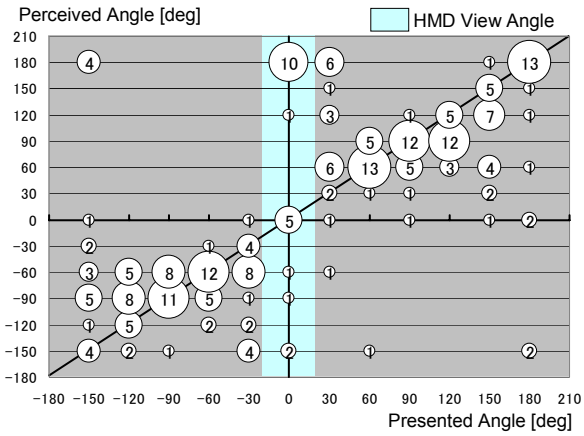
一方，聴覚的 MR の実現には，これまでの聴覚的 VR の研究成果を参考にする．文献[12]では，頭部伝達関数 (Head-Related Transfer Function; HRTF) を水平方向 5 度間隔で測定したものを用いて，補間により任意の位置の HRTF を 0.1 度間隔で算出している．本研究では，視覚的 MR にもそこまでの検出精度は求めていないので，HRTF の補間処理は必要なく，「水平方向 5 度間隔」のみを採用する．HRTF としては，名古屋大学武田研究室が公開しているデータ[16]を用いて実現することにした．

2.4 視聴覚併用 2×2 方式 MR システムの構成

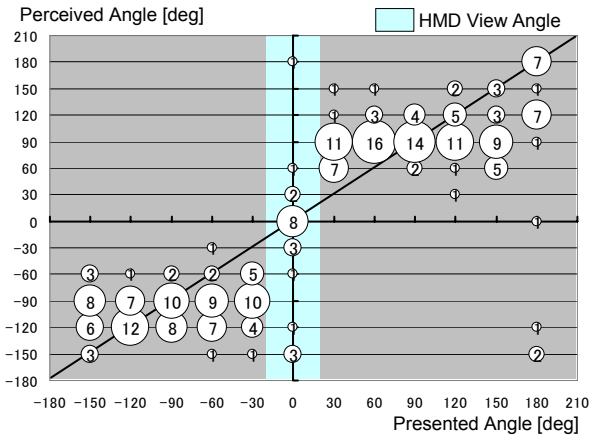
体験者が存在する現実世界をもとに世界座標系を考え，CG 映像空間と 3 次元音場がそれぞれ幾何学的整合性を保って存在し，体験者の移動・対話的操作に対して応答できれば，視覚・聴覚を併用した MR 空間が構築できる．即ち，現実の映像と音，仮想の映像と音の 4 種類が融合する「2×2（ツー・バイ・ツー）方式」ができるわけである．本研究では図 3 に示すようなシステム構成でこれを実現した．

本システムの構成は，3 次元映像生成兼 MR 空間管理処理と 3 次元音場生成処理の 2 つに大別できる．前者では，文献[17]のシステムをもとに構成されていて，HMD に内蔵のビデオカメラでキャプチャした現実世界の映像にコンピュータで生成した CG 映像を重畳描画し，眼前のディスプレイに表示する．このとき体験者の頭部の位置・姿勢は，HMD に取り付けた磁気センサで検出する．磁気センサの精度が不十分な場合は，位置が既知のマーカを画像認識する方法で，幾何学的整合度を高める．

後者では，3 次元映像生成兼 MR 空間管理処理用 PC から送られてくる仮想物体の情報（表 1）をもと



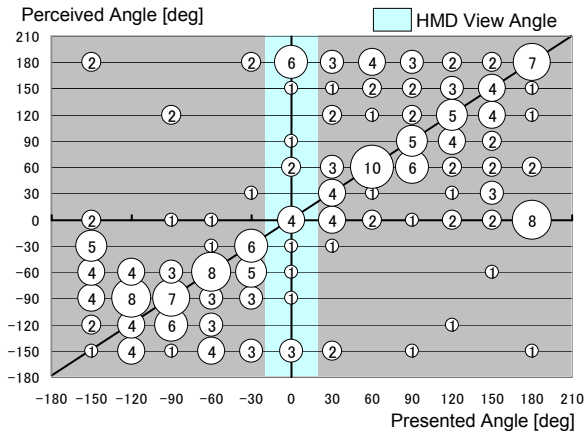
(a) 実音



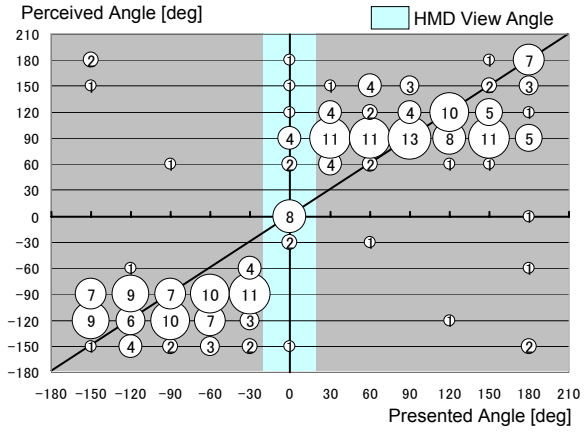
(b) 人工音

図 6 実験 1・開放型方式の結果（視聴覚併用）

Fig. 6 Results of physical mixing method in Ex. 1 (with audio and visual sensations)



(a) 実音



(b) 人工音

図 7 実験 1・密閉型方式の結果（視聴覚併用）

Fig. 7 Results of electronic mixing method in Ex. 1 (with audio and visual sensations)

に 3 次元音場生成用 PC が 3 次元音場を生成する.「開放型音独立提示方式」では, DA 変換器を通してヘッドホンより人工音を提示する.「密閉型混合音提示方式」では実音をマイクロホンで受音し，AD 変換器を通して 3 次元音場生成用 PC に取り込み，人工音と混合した結果を提示する.

3. 実音と人工音の音響的整合性実験

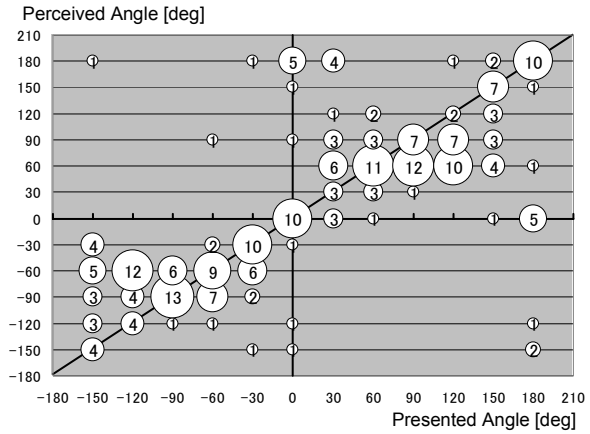
3.1 実験の目的と準備

我々が考案した「開放型音独立提示方式」「密閉型混合音提示方式」（以下の一部では，「開放型方式」「密閉型方式」と略す）のいずれが視聴覚併用型の MR 空間に適しているか，音響的整合性の観点から 2 つの実験で評価する.

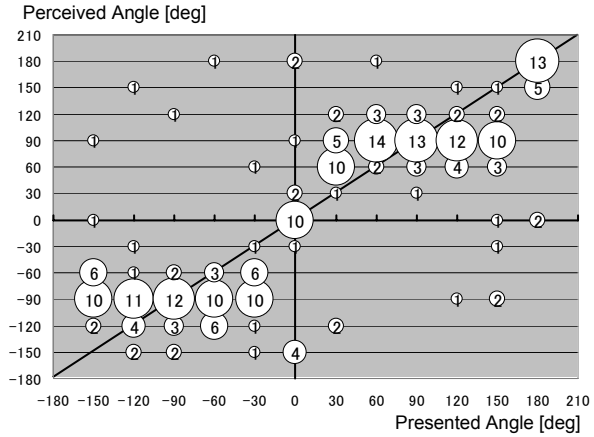
密閉型方式では，マイクロホンを使用し実音を受音するため，受音した実音に HRTF を畳み込む必要がある．本実験では，無指向性のマイクロホンと名古屋大学武田研究室が公開している HRTF を使用し，HRTF の畳み込みあり／なし両場合の比較も行う．即ち実音聴取において，開放型方式では被験者自身の HRTF で聴取しているのに対し，密閉型方式では他人の HRTF で聴取する．人工音聴取においては，両方式ともに他人の HRTF を用いて再現した人工音を聴取する．また，使用するヘッドホン（**図 4**）によって特性が異なるため，事前に擬似頭を用いて特性を測定し, 振幅補正を行い頭部周りの補償を行う.

本実験では，スピーカ (MITSUBISHI DIATONE DS-7) から再生された音を現実空間から伝わってくる実音とし，人工的に 3 次元音場を再現しヘッドホンより提示する音を人工音とみなす．実音源は床面から 115cm の高さに，仮想音源は被験者の目の高さに配置する．実験中，被験者はイスに座り，常に正面方向を向くよう指示する.

実験で使用したシステムのサンプリング周波数は 16kHz, HRTF データ長は 64ms, フレーム周期は 32ms, 映像の処理速度は約 12.5fps である. 実験環境は暗騒音 48dBA のオフィス環境で, 被験者は正常な聴力を



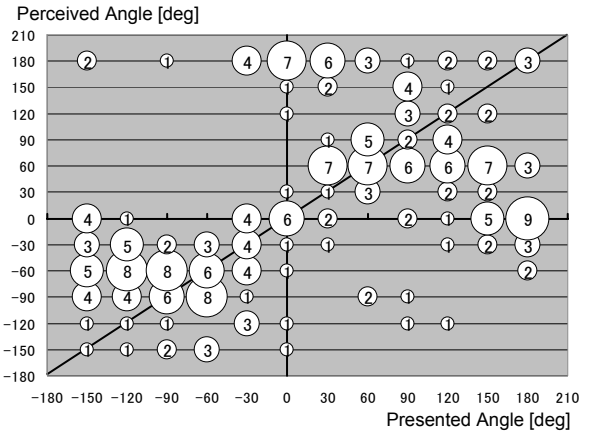
(a) 実音



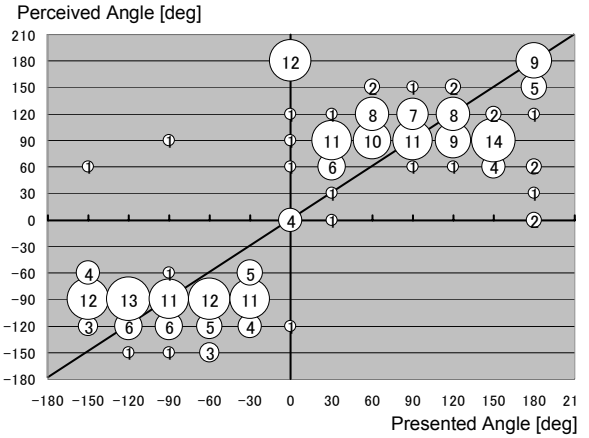
(b) 人工音

図 8 実験 1・開放型方式の結果（聴覚のみ）

Fig. 8 Results of physical mixing method in Ex. 1 (only with auditory sensation)



(a) 実音



(b) 人工音

図 9 実験 1・密閉型方式の結果（聴覚のみ）

Fig. 9 Results of electronic mixing method in Ex. 1 (only with auditory sensation)

持つ学生 10 名である．

3.2 実験 1：固定音源による混合音評価実験

3.2.1 実験内容

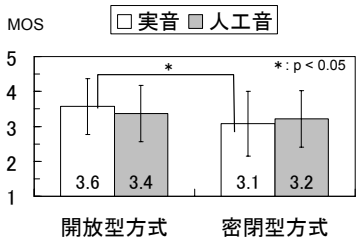
実験 1 では固定音源を用いて開放型方式，密閉型方式のいずれが視聴覚併用型の MR 空間に適しているのか評価する．前述したように，密閉型方式は実音をマイクで受音するため，人工音の音質と差異が生じる．開放型方式はヘッドホンと耳の間の隙間から実音が聞こえるが，ヘッドホンが耳を覆うため音源から被験者の外耳道入り口までの伝達特性が変化する．そこで，実験 1 では実音，人工音ともに固定音源とし，それらを密閉型方式（HRTF あり／なし），開放型方式のそれぞれで同時に聴取したときの定位精度や音質を評価する．また，聴覚的 MR の確認のため，聴覚情報のみの実験も行う．音質評価は被験者による 5 段階（1: bad ～ 5: excellent）の主観評価とする．

図 5 に実験条件を示す．実音再生用のスピーカは 3 台のみ配置し被験者の座るイスを回転することで 12 方位からの実音提示を実現する．実音には周波数帯域が広い楽曲（クラシック）を，人工音にはヘリコプタのプロペラ音を使用する．視聴覚情報を併用する実験ではヘッドホンと HMD を装着し，聴覚情報のみ提示する実験では被験者がヘッドホンとアイマスクを装着する．前者では，仮想音源の位置に CG のヘリコプタを重畳描画する．

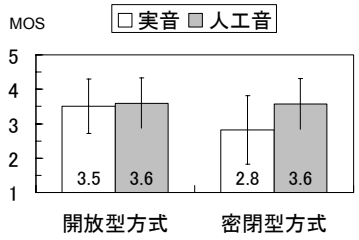
3.2.2 実験手順

被験者ごとに以下の手順を各提示方式で行う．

(1) 実音と人工音をそれぞれ 12 方位よりランダムに選択し 3 秒間同時提示
(2) 実音と人工音の定位方位（12 方位）を回答
※被験者の要求に応じて繰り返し提示
(3) 実音と人工音をもう 1 度 3 秒間同時提示
(4) 実音と人工音の音質を評価（5 段階）
※被験者の要求に応じて繰り返し提示
(5) コメントの自由回答



(a) 視聴覚併用



(b) 聴覚のみ

図 10 音質評価結果

Fig. 10 Evaluation of sound quality in Ex. 1

(6) 全方位から 2 回ずつ提示されるまで手順 (1) ～ (5) を繰り返す

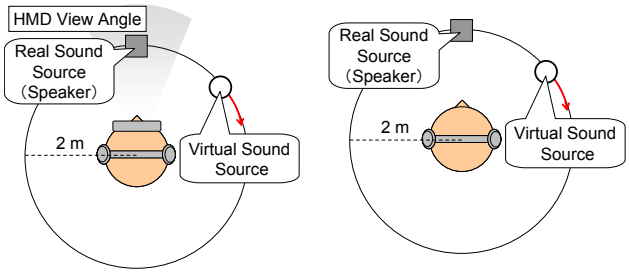
3.2.3 実験結果と考察

実音と人工音の定位精度に関する評価結果を**図 6** ～ **図 9** に，音質に関する評価を**図 10** に示す．密閉型方式では受音した実音に HRTF の畳み込みあり／なし両場合で実験を行ったが，傾向が同じであったため，ここでは受音した実音に HRTF を畳み込んだ場合の結果のみ掲載する．定位精度の評価結果の横軸は音の提示方位を，縦軸は被験者の回答方位を，円の大きさは回答頻度を示している．また右肩上がりの直線は提示方位と回答方位が等しいことを示している．

視聴覚情報併用における定位精度の評価（**図 6**, **図 7**）において，開放型方式で直接実音を聴取しているにも関わらず前後誤りがあるのは，ヘッドホンが耳を覆うため被験者自身の HRTF が変化することが原因と考えられる．人工音聴取は両方式で同等な結果となっており，大きく分けて左右の区別しかできていない．これは，2 音の同時聴取であること，実験システムで利用している HRTF が被験者自身のデータではないことが原因と考えられる．

聴覚情報のみの定位精度の評価（**図 8**, **図 9**）は，視聴覚情報併用時と同様の傾向であった．

音質に関する評価（**図 10**）では，t 検定の結果，視聴覚情報併用において実音の聴取に有意差 ($p < 0.05$) がみられたが，人工音の聴取に関して有意差はみられなかった．また，聴覚情報のみにおいても両方式に有意差はみられなかった．



(a) 視聴覚併用　　(b) 聴覚のみ

図 11 実音と人工音の聴取（実験 2）

Fig. 11 Reception of real and artificial sounds in Ex. 2

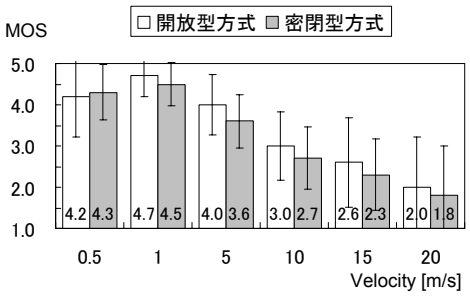
また，10 名中 8 名の被験者より「密閉型方式では自分自身が発話すると不快感がある」という意見があった．被験者自身の声は，体を伝わり直接耳に届くと同時に，マイクロホンでも受音されヘッドホンから提示されることが原因であると考えられる．さらに，密閉型方式では，実験対象である実音以外の環境音（PC のキーボードを叩く音や周りの雑音）がかなり気になるという声が多かった．とりわけ，ヘッドホンやマイクのケーブルが擦れる音が悪影響を及ぼす最大の要因であった．また，視覚では見たいものに注視できるが，密閉型方式のように実音を無指向性のマイクロホンで受音しヘッドホンから提示する場合，耳の構造と無指向性マイクロホンの構造の違いから，普段なら気にならない方向の音が気になるという現象があった．特に，ヘッドホン近くで発生した音の影響が顕著で，これは視覚の HMD にはない予期せざる結果であり，視聴覚併用型の MR システムにおける密閉型方式の欠点といえる．

3.3 実験 2：移動音源による混合音評価実験

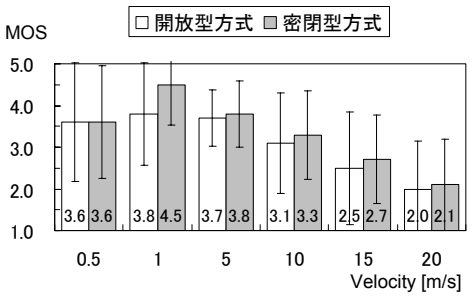
3.3.1 実験内容

実験 2 では移動音源を用いて開放型方式，密閉型方式（HRTF あり／なし）のいずれが視聴覚併用型の MR 空間に適しているか評価する．本実験では，スピーカ再生した音を実音としているため，それを移動することは困難である．そこで，実音は固定音源とし人工音のみ移動音源とした上で，それらを同時に聴取したとき人工音の定位感がどの程度であるか評価する．実験 1 同様，聴覚的 MR の確認のため，聴覚情報のみの実験も行う．定位感の評価は被験者による 5 段階（1: bad ～ 5: excellent）の主観評価とする．

図 11 に実験条件を示す．提示音は実験 1 と同じで，被験者から実音源までの距離は 2m，仮想音源であるヘリコプタは被験者を中心に半径 2m の円周上を旋回する．また仮想音源の旋回速度は 0.5, 1, 5, 10, 15, 20m/s の 6 種類とする．実験 1 同様，視聴覚情報併用の実験ではヘッドホンと HMD を装着し，



(a) 視聴覚併用



(b) 聴覚のみ

図 12 定位感の評価

Fig. 12 Evaluation of sound source localization

仮想音源の位置に CG のヘリコプタを重畳描画する．聴覚情報のみ提示する実験では被験者がヘッドホンとアイマスクを装着する．移動音は HRTF を 5 度間隔で切り替えながら生成，提示する．

3.3.2 実験手順

被験者ごとに以下の手順を各提示方式で行う．

(1) 実音と人工音を同時提示
(2) 仮想音源の定位感を評価（5 段階）
※被験者の要求に応じて繰り返し提示
(3) コメントの自由回答
(4) 仮想音源の速度を変えて，手順 (1) 〜 (3) を繰り返す

3.3.3 実験結果と考察

人工音の定位感に関して，視聴覚情報併用の場合の評価結果を**図 12** (a) に，聴覚情報のみの場合の評価結果を**図 12** (b) に示す．密閉型方式で HRTF あり／なし両提示方式の実験を行ったが，本実験でも傾向が同じであったため，受音した実音に HRTF を畳み込んだ場合の結果のみ掲載する．図の横軸はヘリコプタの旋回速度，縦軸は MOS (Mean Opinion Score) 値を示している．**図 12** から，視聴覚情報併用，聴覚情報のみどちらの場合も，ともにヘリコプタの速度が 1m/s の場合に最も評価が高く，速度が上がるにつれ評価が下がる傾向にあることがわかる．これは，ヘリコプタの速度が上がると，現システムの映像処理速度（約 12.5fps）では，フレーム間のヘリコプタの方位変化量が大きくなり，即ちフレーム間の 3 次元音場の変化量が大きくなるため，自然な音源定位をできないことが原因であると考えられる．また，t 検定の結果，どの速度でも提示方式の違いによる有意差はみられなかった．

一方，**図 12** (a) (b) を比較すると，旋回速度が 0.5 m/s，1m/s と低速な場合は，視覚による補正効果があったことがわかる．しかし，その他の速度で視覚による補正効果が見られないのは，ヘリコプタの速度が上がると，CG 映像が HMD 視野内で描画されるフレーム数が少なくなり，CG 映像を十分確認できないことが原因であると考えられる．

以上をまとめると，視聴覚情報併用または聴覚情報のみにおいて，人工音の定位感に関して両方式の優劣に大きな差は無く，仮想音源が被験者から 2m の位置で，速度 0.5 〜 5m/s 程度で移動させた場合，MOS 値は 3.5 以上で，十分な定位感を提示できているといえる．また，速度が低速の場合には，視覚が加わることで MOS 値が上がり，視聴覚併用の効果があることが確認された．

3.4 方式選択の総合判断

本実験の前に立てた予測は，実音単独では原理的に「開放型方式」が若干優れていても，混合音の提示では「密閉型方式」の方が勝り，現実と仮想の融合提示方式として適しているというものであった．これは，視覚における「ビデオシースルー方式」の優位性から想定した仮説である．しかし，3.2 節の実験で遭遇した「密閉型方式」の大きな欠点

- 自分自身の声が直接体を伝わって聞こえるためヘッドホンと 2 重提示になってしまう．
- 注視（意識）していない方向からの音が受音，提示されて大きな違和感となる．特に，ヘッドホン近くのコード類の摩擦音が悪影響を及ぼす．

は，3.3 節の実験でも続き，特に前者は全く予期せざる結果であり，「密閉型方式」の優位性では補えないレベルであった．後者は，マイクロホンをワイヤレス化することで幾分回避できると考えられるが，「密閉型方式」の実現に制約を与えることになる．一方，3.3 節の実験結果は，両方式で人工音の定位感に大差がないことを示している．であれば，当初の予測に反して「開放型方式」の採用を否定する要因はない．

以上の理由から，実音を使用したり体験者自身が音を発する MR 空間では，開放型ヘッドホンによる「開放型音独立提示方式」を 2×2 方式複合現実感システムに採用することにした．

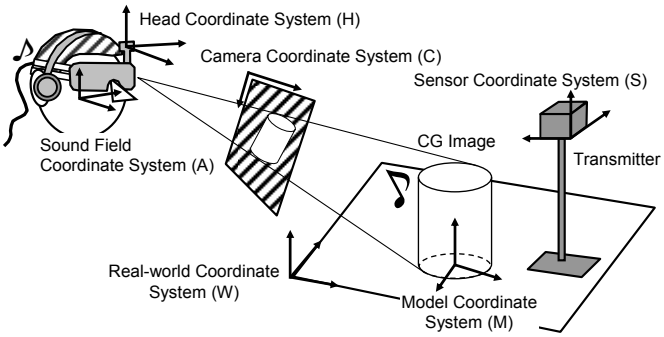


図 13 幾何学的位置合わせを行う座標系

Fig. 13 Coordinate systems for geometric registration

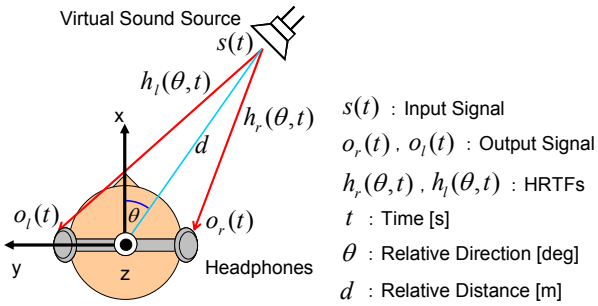


図 14 人工音の聴取

Fig. 14 Reception of artificial sound

4. 2×2 方式での幾何学的整合性の達成

4.1 視覚における現実と仮想の位置あわせ

現実空間と仮想空間の視覚的な幾何学的整合性を達成するには，**図 13** の実世界座標系 (W)，仮想物体のモデル座標系 (M), HMD 内蔵カメラ座標系 (C) の 3 つの座標系の幾何学的整合性をとる必要がある．実世界座標系 (W)，センサ座標系 (S)，モデル座標系 (M) の 3 者の相対位置，頭部座標系 (H)，カメラ座標系 (C) の相対位置は，いずれも一定である．

これらの座標系の位置合わせは，センサとカメラ間の事前キャリブレーションによって行う．また HMD に取り付けた磁気センサの位置・姿勢はセンサ座標系 (S) 上で常に計測されているので，磁気センサを原点とする頭部座標系 (H) とセンサ座標系 (S) の対応付けが可能である．以上により，実世界座標系 (W)，モデル座標系 (M)，カメラ座標系 (C) の 3 つの座標系の幾何学的対応付けが可能となる．

4.2 聴覚における現実と仮想の位置あわせ

同様に，現実空間と仮想空間の聴覚的な幾何学的整合性の達成には，**図 13** の実世界座標系 (W)，モデル座標系 (M), 音場座標系 (A) の 3 つの座標系の幾何学的整合性をとる必要がある．

音場座標系 (A) は体験者の頭部中心を原点としており，体験者の頭の大きさがわかれば，HMD 上のセンサを原点とした頭部座標系 (H) との対応付けが可能である．頭部座標系 (H), 実世界座標系 (W), モデル座標系 (M) の対応付けは 4.1 節で述べたように既知であるため，実世界座標系 (W)，仮想物体のモデル座標系 (M), 音場座標系 (A) の 3 つの座標系の対応付けが可能である．

4.3 3 次元音場の生成

体験者が提示音を聴取する様子を**図 14** に模式的に示す．**図 14** において，仮想音源を $s(t)$，仮想音源から体験者までの相対距離を d，仮想音源と体験者の相対方位を θ，左右の耳に対する HRTF をそれぞれ，$h_l(\theta,t)$，$h_r(\theta,t)$，左右の耳で提示される提示音をそれぞれ，$o_l(t)$，$o_r(t)$とすると，

$$o_l(t) = (s(t) * h_l(\theta,t)) / d$$
$$o_r(t) = (s(t) * h_r(\theta,t)) / d$$

（*は畳み込み演算）

となる．本来，距離減衰を厳密に再現するためには，温度や湿度，風速などを考慮して距離減衰量を算出しなければならない．しかし，音の定位感を重視するのであれば，相対距離だけでも十分に距離減衰を表現できるといえるため，本システムでは，畳み込み演算で得られた結果を相対距離で割り，簡易的に距離減衰を表現する．

表 2 構成機器

Table 2 Hardware configuration

Constituent Elements	Model Number/Names
PC Generating 3-D Image and Managing MR Space	Canon MR Platform System
PC Generating 3D Sound Field	Dell Precision 670
Magnetic Sensor	Polhemus 3SPACE FASTRAK
Head Mounted Display	Canon VH-2002
DA Converter	Thinknet DF-2032D
Open-air Headphones	SONY MDR-F1

5. 視聴覚 2×2 方式複合現実感システムの実現

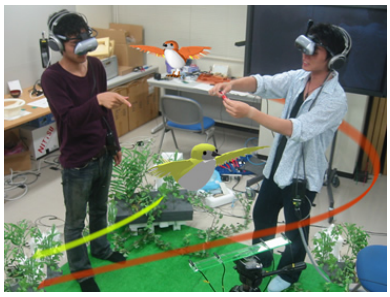
5.1 試作システムでの融合臨場感の確認

3 章での実験と考察により「開放型音独立提示方式」を採用し，4 章の内容を実装することで，視聴覚 2×2 方式の MR システムを試作した．

図 3 のシステム構成の実現に，**表 2** に示す機器を用いた．本システムはサンプリング周波数 16kHz, HRTF データ長 64ms，フレーム周期を 32ms で処理している．また，実時間で動作させて機能検証することが第一であるため，処理時間のかかる部屋の壁などによる反射音は考慮していない．



(a) ヘリコプタ



(b) Watch the Birdie!

図 15 視覚・聴覚を併用した MR 空間

Fig. 15 Scene of the MR space using audio-visual senses

本試作システムでは，現実空間（幅：約 8.8m，奥行き：約 6.2m，高さ：約 2.7m の部屋）に CG 製のヘリコプタを重畳描画することを試みた（図 15 (a)）. ヘリコプタが体験者に向かって近づくにつれ，音が徐々に大きくなり，逆に体験者から遠ざかるにつれ, 音が徐々に小さくなることを確認した. また，ヘリコプタが体験者の左方向から右方向に移動すると音も同じように変化することも確認した．ここでは，特に意図的な実音は発していないが，現実世界から様々な環境音が聞こえ，かつ現実の光景に溶け込んで違和感はなかったので，本試作システムが目標とする 2×2 方式の視聴覚 MR を達成しているものと判断した.

5.2 複数人体験できる MR アトラクションへの実装

前述の試作システムでは，「体験者は 1 人」「仮想音源は 1 つ」といった制約を設けていたが，所望の機能検証を達成できたので，この制約を排除した複数人同時体験可能な本格システムへと移行した.

本システムの視聴覚併用効果を最大限に活かした MR アトラクションとして，音を利用して鳥を呼び寄せ，鳥の姿と鳴き声を楽しむ「Watch the Birdie!」を制作した（図 15 (b)）. そのコンテンツとしての設計方針は文献[18]に譲り，本論文に関わる技術的概要のみを述べる.

このアトラクションでは，実音と人工音を同時に提示するだけでなく，現実空間で発した音刺激をイベントとして検出し，MR 空間への入力として利用することを必要とした．このため，試作システムに AD 変換器 (Thinknet DF-2X16-2) と小規模直線型マイクロホンアレーを追加して，音刺激の発生位置や方向を検出する機能を持たせた．また，その検出精度向上との兼ね合いで, サンプリング周波数 32kHz, HRTF データ長 128ms, フレーム周期 64ms で処理している.

この MR アトラクションは，複数人が同時に MR 空間を体験することができる．飛び回る複数の CG 製の鳥の姿が，どの体験者に対しても目の前の現実空間の光景に重畳描画され，鳥のさえずり（人工音）が周りの実音に溶け込んでヘッドホンから聞こえてくる．いずれも 3D 空間内で視聴覚ともに幾何学的違和感がなく, 合成映像や混合音提示が達成できた.

本 MR アトラクションは，2006 年 9 月 7 日から 3 日間，仙台市青少年文化センターにおいて開催された日本バーチャルリアリティ学会第 11 回大会でデモ展示を行った. 1 回あたりの体験時間が約 10 分と参加体験型アトラクションとしては長いながらも，開催期間中約 200 人が体験するところとなった．体験者の中でも，これまで視覚的 MR 体験が多数ある参加者には，狭視野の HMD では見ることができない側面や背面の様子を音によって補完できることが好評であった．また，3 次元音場による聴覚的 VR しか経験のない参加者には，視覚と整合性がとれた体験が驚きをもって受け入れられた.

また，本 MR アトラクションの体験者からは，視聴覚刺激，現実と仮想の間に時間遅延を感じるという声は全くなかった．文献[12]の聴覚的 VR では，システム遅延時間が 100ms の場合，約 60%の割合で遅延を知覚したと報告されている．我々の視聴覚 MR システムでは，聴覚的 MR 部に 100ms 以上のシステム遅延時間があるが，視覚刺激を伴った場合にはこの遅延を感じないということが，初期実験から指摘されていた．このことが，被験者という意識のない一般参加者にトータルシステムとして意義のある MR アトラクションの体験により，改めて確認されたわけである.

上記の現象は，この条件下では視聴覚間での優位性が働いたことによるものだと考えられる．このことからも本システムは視覚と聴覚を時間的，幾何学的に違和感無く融合しており，我々が目指した視聴覚の調和がとれた MR システムを実現したといえる.

6. むすび

本論文では，社会的にも注目を集める複合現実感 (MR) の発展形として，まだ例のない視聴覚併用型の MR システムの実現を目指し，その実現方式の検

討と達成したシステムの結果を記した．視覚的 MR に関しては既に開発済みで既発表の方式を踏襲したので，紙幅の多くは聴覚的 MR に関する記述に割いたが，本論文は聴覚・音響分野での新規性を主張するものではなく，視覚・聴覚の調和がとれ，現実と仮想の融合が同時に達成できるシステムの提案が主眼である．

視覚的 MR とのアナロジーから，聴覚的 MR の提示方式として，「開放型音独立提示方式」「密閉型混合音提示方式」の 2 方式を考案し，その優劣を比較実験した．当初の予想では，視覚の「ビデオシースルー方式」と併用するには「密閉型混合音提示方式」が適しているかと思われたが，結果は開放型ヘッドホンを用いる前者の方が優れていることが判明し，以後「開放型音独立提示方式」を採用した．

続いて，視覚的 MR と聴覚的 MR の幾何学的整合性を達成し，具体的に稼動するシステムを開発して，その機能検証を行った．これは，単に 2×2 視聴覚 MR の基本システムを実装するだけでなく，具体的に一般参加者が体験できる MR アトラクションを構築することで，目指した機能が確実に作動していることを確認した．

本研究が当初目指した視聴覚併用 MR の基本機能は達成できたので，今後は，仮想世界での生成音が現実の物体や壁面に反射する場合，現実世界の実音が仮想物体に反射する場合等を考慮した発展的な研究を進めて行く計画である．

謝辞

本研究の一部は，科研費・基盤研究 A「三感融合型複合現実空間の構成法に関する研究」による．

参考文献

[1] Y. Ohta and H. Tamura (eds.): *Mixed Reality—Merging Real and Virtual Worlds*, Ohm-sha & Springer-Verlag, 1999.

[2] “複合現実感特集 1～3”，日本バーチャルリアリティ学会論文誌, 1999, 2002, 2005.

[3] 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 画像と同期したインタラクティブ音場再生システムについて”，同上, Vol. 5, No. 3, pp. 965 – 974, 2000.

[4] W. W. Gaver, R. B. Smith, and T. O'Sheak: “Effective sounds in complex systems: The ARKola simulation,” Proc. CHI’91, pp. 85 – 90, 1991.

[5] P. Flanagan, K. I. McAnally, R. L. Martin, J. W. Meehan, and S. R. Oldfield: “Aurally and visually guided visual search in a virtual environment,” *Hum. Factors*, Vol. 40, pp. 461 – 468, 1998.

[6] 大島登志一, 佐藤清秀, 山本裕之, 田村秀行: “AR^2 ホッケー：協調型複合現実感システムの実現”，日本バーチャルリアリティ学会論文誌, Vol. 3, No. 2, pp. 55 – 60, 1998.

[7] C. E. Hughes, C. B. Stapleton, D. E. Hughes, and E. M. Smith: “Mixed reality in education, entertainment, and training,” *IEEE Compt. Graph. & App*, Vol. 25, No. 6, pp. 24 – 30, 2005.

[8] J. Sodnik, S. Tomazic, R. Grasset, A. Duenser, and M. Billinghurst: “Spatial sound localization in an augmented reality environment,” Proc. OZCHI 2006, pp. 111 – 117, 2006.

[9] D. Dobler, P. Stampfl, and M. Haller: “ASR—Augmented sound reality,” SIGGRAPH 2002 Sketches and Applications, p. 148, 2002.

[10] Z. Y. Zhou, A. D. Cheok, Y. Qiu, and X. Yang: “The role of 3-D sound in human reaction and performance in augmented reality environments,” *IEEE Trans. on Syst., Man & Cybern.*, Part A, Vol. 37, No. 2, pp. 262 – 272, 2007.

[11] H. Fuchs and J. Ackerman: “Displays for augmented reality: Historical remarks and future prospects,” in [1], pp. 31 – 40, 1999.

[12] 矢入聡, 岩谷幸雄, 鈴木陽一: “頭部運動感応型ソフトウェア聴覚ディスプレイの開発”，日本バーチャルリアリティ学会論文誌, Vol. 11, No. 3, pp. 437 – 446, 2006.

[13] K. Boff: *Handbook of Perception and Human Performance: Cognitive Processes and Performance*, Wiley-Interscience, 1986.

[14] 佐藤清秀, 山本裕之, 田村秀行: “カメラと３次元センサの組み合わせによる現実空間と仮想空間の位置合わせ手法”，日本バーチャルリアリティ学会論文誌, Vol. 4, No .1, pp. 295 – 302, 1999.

[15] 内山晋二, 山本裕之, 田村秀行: “複合現実感のためのハイブリッド位置合わせ手法−6 自由度センサとビジョン手法の併用−”，同上, Vol. 8, No. 1, pp. 119 – 125, 2003.

[16] 名古屋大学武田（旧板倉）研究室：http://www.sp.m.is.nagoya-u.ac.jp/HRTF/

[17] S. Uchiyama, K. Takemoto, K. Satoh, H. Yamamoto, and H. Tamura: “MR Platform: A basic body on which mixed reality applications are built,” Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR 2002), pp. 246 – 253, 2002.

[18] 石黒祥生, 大槻麻衣, 比嘉恭太, 木村朝子, 柴田史久, 田村秀行: “Watch the Birdie!—三感融合型複合現実感アトラクション”，日本バーチャルリアリティ学会論文誌, Vol. 12, No. 3, pp. 259 – 268, 2007.

(2007 年 12 月 4 日受付)

［著者紹介］

比嘉 恭太　（学生会員）



2006 年立命館大学理工学部情報学科卒．2008 年同 大学院理工学研究科博士前期課程修了．同年 4 月，日本電気（株）入社．2005 年より 2007 年まで視聴覚併用型複合現実感システムの研究に従事．電子情報通信学会学生会員．

西浦 敬信　（正会員）



1999 年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了．2001 年同 研究科博士後期課程修了．和歌山大学システム工学部助手を経て，2004 年 4 月より立命館大学情報理工学部メディア情報学科助教授．現在，同 准教授．博士（工学）．音響信号処理，主として音環境の理解・生成に関する研究に従事．電子情報通信学会，日本音響学会，ISCA の会員．

木村 朝子　（正会員）



1996 年大阪大学基礎工学部卒．1998 年同 大学院基礎工学研究科博士前期課程修了．同 大学助手，2003 年立命館大学理工学部助教授，2004 年同情報理工学部助教授を経て，2007 年 4 月より科学技術振興機構さきがけ研究員，立命館大学総合理工学研究機構客員教授．博士（工学）．実世界指向インタフェース，複合現実感，ハプテックインタフェースの研究に従事．2001 年より 2002 年まで Mayo Clinic にて Special Project Associate．電子情報通信学会，情報処理学会，ヒューマンインタフェース学会，ACM，IEEE 各会員．2006 年本学会学術奨励賞等受賞．

柴田 史久　（正会員）



1996 年大阪大学大学院基礎工学研究科博士前期課程修了．1999 年同 研究科博士後期課程修了．大阪大学産業科学研究所助手を経て，2003 年 4 月より立命館大学理工学部助教授．現在，同 情報理工学部情報コミュニケーション学科准教授．工学博士．モバイルコンピューティング，複合現実感等の研究に従事．IEEE，電子情報通信学会，日本ロボット学会，情報処理学会等の会員．2005 年本学会学術奨励賞受賞．

田村 秀行　（正会員）



1970 年京都大学工学部電気工学科卒．工業技術院電子技術総合研究所，キヤノン（株）等を経て，2003 年 4 月より立命館大学理工学部教授．現在，同 情報理工学部メディア情報学科教授．工学博士．1997 年より 2001 年まで，MR システム研究所にて「複合現実感研究プロジェクト」を率いた．本学会元理事，現在，評議員，複合現実感研究委員会顧問．編著書「Mixed Reality」(Ohmsha & Springer)「コンピュータ画像処理」（オーム社）など．電子情報通信学会フェロー，IEEE，ACM，情報処理学会，人工知能学会，映像情報メディア学会等の会員．情報処理学会論文賞，人工知能学会功労賞等を受賞．