

# 視覚・聴覚を併用した複合現実感システムの開発(6) —実時間反射音生成に基づく音場再生の実現—

## Mixed Reality System using Audio and Visual Senses (6) ---Implementation of Sound Field Reproduction Based on Generation of Reflected Sound in Real Time---

吉持達也, 森勢将雅, 西浦敬信, 木村朝子, 柴田史久, 田村秀行

Tatsuya Yoshimochi, Masanori Morise, Takanobu Nishiura, Asako Kimura, Fumihisa Shibata, and Hideyuki Tamura

立命館大学大学院 理工学研究科  
(〒525-8577 滋賀県草津市野路東1-1-1)

**Abstract:** This paper describes the experimental results of a method of sound field reproduction for an MR system that merged the real and virtual worlds in both the audio and visual senses, in which reflected sounds are generated in real time. We have already developed an MR system using audio and visual senses with a headphone and speakers. However, on the existing MR system, there are individually differences for sound localization because the system adopts a prepared HRTF of unrelated person. Therefore, we proposed a method of sound field reproduction, in which direct sound is presented from a headphone and reflected sounds are presented from speakers. To evaluate the basic performance of sound localization, we conducted the evaluation experiment and compared proposed system with existing system using HRTF. As a result, the performance of sound localization was slightly degraded, but according to the condition, it would be better.

**Key Words:** Mixed Reality, Audio and Visual Senses, Reflected sound

### 1. はじめに

現実世界と仮想世界を融合する複合現実感 (Mixed Reality; MR) 技術は, 人工現実感 (Virtual Reality; VR) の発展系であり, VRにはない大きな潜在能力をもつものとして期待されている. 一般的に, MRでは視覚的な融合に主眼が置かれているが, 我々は視覚的 MRだけでなく聴覚的 MRを実現し, 視覚・聴覚両面で現実世界と仮想世界を矛盾なく融合する視聴覚併用 MRシステムとして, 「2×2方式視聴覚併用 MRシステム」を研究開発した[1]. さらに, 視聴覚併用 MR空間における表現力向上を目指し, 仮想物体による実音の反射・遮断, 実物体による人工音の反射・遮断の現象を実現した[2].

これまでの視聴覚併用 MRシステムでは, 3次元音場を表現する手法として, 事前に用意した他人の頭部伝達関数 (Head-Related Transfer Function; HRTF) を用いたバイノーラル再生方式による音の提示を行ってきた. しかし, HRTFは人によって異なるため, 音像の定位感には個人差があるという問題があった. このことから, 音響分野では HRTFを用いない音像提示手法として, スピーカを複数用いた手法が研究されている[3]. しかし, スピーカによる音提示は, 音が人の耳に届くまでに時間差が生じるため, 実時間性が問われる MRでは不適切と言える. そこで, ヘッドホンとスピーカを併用するという着想[4]を応用し, 直接音と反射音を分割した個人特徴を用いない音像提示によって, 視聴覚併用 MRシステムにおける反射音生成に基づく音場再生システムを検討する.

### 2. 直接音・反射音分割提示による音像提示方法

#### 2.1 直接音・反射音の分割提示

人は, 両耳間での音の到来する時間差や音量差, 頭部の形状による音の変化, 壁による音の反射の影響などを自らの経験をもとに考慮し, 音の到来方向を判別している. そのため, 現実に近い音場を生成することができれば, 音の到来方向を判別することができると言える.

既存の技術では, 複数のスピーカを用いる方法で現実に近い音場を生成できることがわかっている[5]. この方式の場合, スピーカは空間中に音を放射するため, 各人の HRTFで音を聴取できる. しかし, 体験者の行動に対し実時間での応答が重要な MRにおいては, スピーカから発する音が空間中を伝わることによって生じる物理的な遅延に起因して, 不都合が生じてしまう可能性がある. 一方, ヘッドホンによる音提示では音の遅延が発生しないため, MRにおいて適切な音の提示方法であると考えられる. しかしながら, 両耳を繋いだ線を対称に前後の定位感を間違える前後誤判断や頭の中で音が鳴っているように感じる脳内定位を引き起こしてしまうという問題がある.

そこで, 図1のようにヘッドホンとスピーカを併用し, 体験者に最初に届く直接音はヘッドホンから, それ以外の反射音をスピーカから提示し音像定位を補うことで, HRTFを用いない音場再生システムを実現する.

#### 2.2 システム構成

構成は大きく分けて視覚的 MRの処理ならびに MR空間を管理する視覚情報生成兼 MR空間管理処理部と, 聴

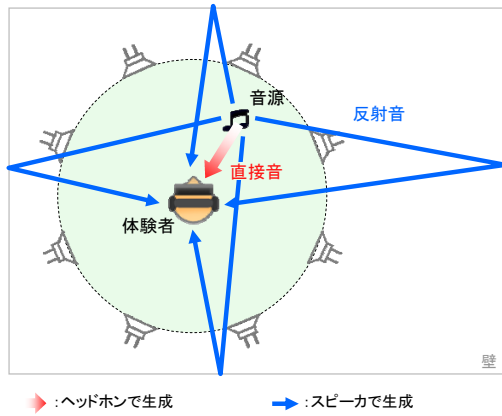


図1 直接音と反射音の分割提示

覚的 MR を処理する聴覚情報生成処理部の 2 つからなっている[4].

本システムでは聴覚的 MR を提示する方法として、開放型ヘッドホンとスピーカ 8 本を併用する. 開放型ヘッドホンを用いることで、実音と人工音の同時聴取を可能にする. また、反射音を体験者の全周から提示するため、スピーカ 8 本を円状に等間隔に設置する.

### 2.3 直接音の提示方法

直接音は、比較的個人差の影響の少ない両耳間時間差と両耳間強度差のみを考慮し、実時間で生成する. このときの両耳間時間差の生成には、Woodworth と Schlosberg による代表的なモデルを使用する. 両耳間強度差は、音源から両耳までの経路差による距離減衰のみを考慮する.

また、ヘッドホンとスピーカの特性による音質差をなくすため、TSP 信号を用いてそれぞれの特性を測定し、ヘッドホンの特性の逆フィルタを生成して特性を合わせる. 音量についても同様にレベルを合わせる必要があるため、ヘッドホンとスピーカそれぞれからホワイトノイズを再生し、マイクロホンを用いて耳元で測定して音量差が 1[dB] 以内に収まるように音量レベルを合わせている.

### 2.4 反射音の提示方法

MR 空間で反射音を実現するには、体験者の位置姿勢や仮想音源の位置に合わせて実時間で反射音を生成する必要がある. そこで、反射音を音源から体験者に届くまでの反射回数の少ない低次反射成分と反射回数の多い高次反射成分に分けて考える.

低次反射成分は直接音に次いで音のエネルギーが大きく、体験者の位置や音源の位置によって音が変化しやすいため、音像の定位感に大きく影響を与える. 一方、高次反射成分は多数の音線があらゆる方向から到来するが、エネルギーが小さい残響であるため、直接音、低次反射成分に比べて体験者の位置や音源の位置による音の変化を感じにくい.

そこで、反射音全体を計算する場合に比べ、大幅に計算量が少ない低次反射成分は実時間で計算を行い、膨大な計算量を必要とする高次反射成分は事前に体験者や音源の位置を固定した環境で生成する. 高次反射成分は低次反射

成分に実時間で付加して提示する.

反射音の生成方法としては、反射経路を算出し、反射音のインパルス応答を求め、音源に畳込むという方法をとる. 反射経路の算出には鏡像法を用い、求めた反射経路からそれぞれ到達時間や距離減衰、壁との反射角度を求めてインパルス応答を生成する. 具体的には、まず、経路長からインパルスの到達時間、距離減衰を求める. そして、音が壁に反射した角度をもとにした実測値の壁での反射特性を反射した回数分量込むことで、インパルス応答を生成する. そうして経路ごとのインパルス応答を生成し、全ての経路のインパルス応答を時間軸上で加算することで、反射音のインパルス応答を生成する.

また、反射音の提示にはスピーカを複数使用するため、反射音を反射経路により振り分ける必要がある. そこで、1 つの反射経路の音を 2 つのスピーカに振り分けて提示するという方法をとる[6].

## 3. 定位性能評価実験

### 3.1 実験の目的と準備

本システムにおける音の定位性能を評価するために、2 つの実験で定位精度と音質の評価を行う. また、バイノーラル再生を用いたシステムとの比較を行うため、比嘉ら[1]のシステムにおける視聴覚併用での開放型方式を既存のシステムとし、本システムとの比較評価も行う.

実験環境を図 2 に示す. 本実験では、実音は 1 つのスピーカから提示し、人工音はスピーカとヘッドホンを併用した提案手法で提示する. なお、反射音は実時間処理を行うため 1 次反射のみを低次反射成分とする. 実音源、仮想音源は共に半径 2.0[m] の円上、被験者の目の高さ (1.15[m]) の位置に配置する. 実験中、被験者はイスに座り、常に正面を向くように指示する.

実験で使用したシステムのサンプリング周波数は 16[kHz]、フレーム周期は 64[ms]、映像の処理速度は約 28.4[fps] である. 実験環境は暗騒音 48[dBA] のオフィス環境で、被験者は正常な聴力を持つ学生 10 名である. 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ため、被験者はヘッドホンと HMD (視野角: 水平方向

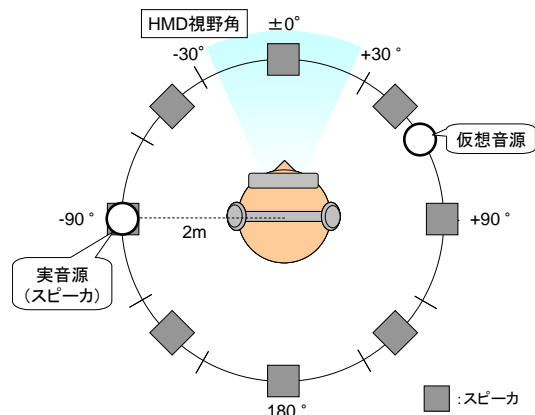


図2 実験環境

51度, 垂直方向37度)を装着し, 仮想音源の位置に追従する形でCGのヘリコプタを重畳描画する。

### 3.2 実験1: 固定音源による混合音評価実験

**【実験内容】** 実験1では固定音源を用いて定位性能を評価する。実音, 人工音ともに固定音源とし, 同時に聴取したときの定位精度や音質を評価する。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。なお, 実験環境のスピーカ配置では12方位全てからの実音提示ができないため, 被験者が座ってい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。

**【実験結果と考察】** 本システムと既存のシステムの実音と人工音の定位精度に関する評価結果を図3, 図4に, 音質に関する評価を図5, 図6に示す。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。音質評価の縦軸はMOS (Mean Opinion Score) 値を示している。

既存システムと本システムを比較すると, 人工音が90度方向に定位しやすい既存のシステムに対して, 本システムでは30度などの斜めの方向に定位している。これは, 反射音によって定位精度の補助効果があることを示していると考えられる。すなわち, 斜めの方向の定位精度に関してはHRTFによる個人差が大きく, 反射音を全周から提示することにより自分のHRTFを通して聞くため, 定位しやすくなっていると考えられるからである。脳内定位をコメントしていたのは10名中1名のみであったため,

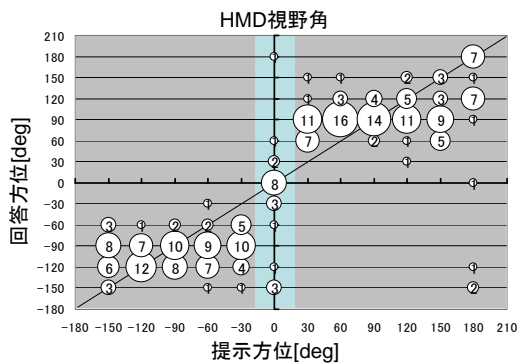
反射音の提示によりその影響は抑えられていると考えられる。しかし, 前後誤判断に関しては多数見られることから, 影響を抑えきれておらず, 音像定位の補助が十分ではなかったと考えられる。これは, 直接音に両耳間時間差と両耳間強度差しか考慮していないことが原因であると考えられる。実音の定位精度が低いのは, 既存のシステム同様, ヘッドホンを装着していることによりHRTFが変化していることや, 実音と人工音を同時に提示するためそれぞれの音を聴取しづらくなっていることが原因として考えられる。

音質に関して人工音の評価が向上しているのは, 反射音を提示するようになったためだと考えられる。また, 実音の評価が低い, これは被験者が既存のシステムでの実験時と異なるため, 人工音の評価が向上したことにより相対的に低くなっていると考えられる。

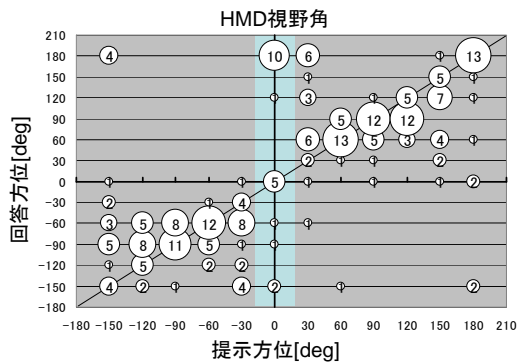
### 3.3 実験2: 移動音源による混合音評価実験

**【実験内容】** 実験2では, 移動音源を用いて定位性能を評価する。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を評価する。定位感の評価は被験者による5段階評価(1: bad ~ 5: excellent)の主観評価とする。なお, 仮想音源は被験者を中心に半径2.0[m]の円周上を旋回し, 速度は0.5, 1.0, 5.0[m/s]の3種類とする。

**【実験結果と考察】** 本システムと既存のシステムでの人工音の定位感に関する評価結果を図7, 図8に示す。横軸は仮想音源の旋回速度, 縦軸はMOS値を示している。

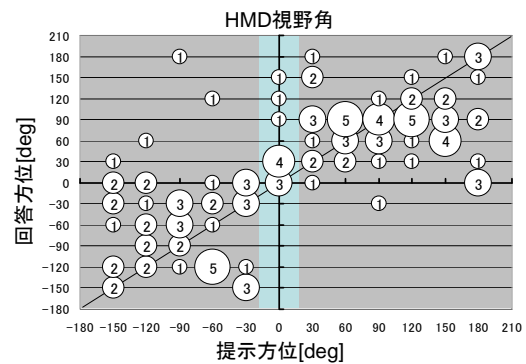


(a) 人工音

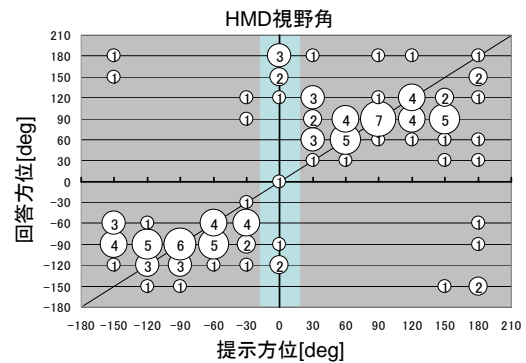


(b) 実音

図3 実験1・定位精度の結果 (既存のシステム)



(a) 人工音



(b) 実音

図4 実験1・定位精度の結果 (本システム)

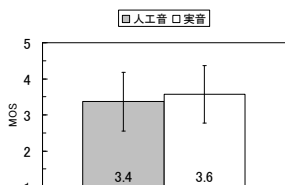


図5 実験1・音質評価 (既存のシステム)

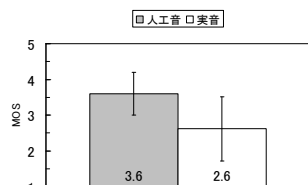


図6 実験1・音質評価 (本システム)

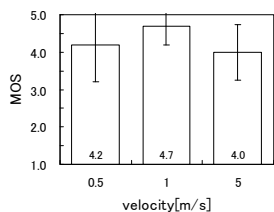


図7 実験2結果 (既存のシステム)

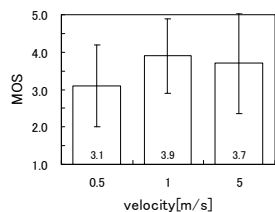


図8 実験2結果 (本システム)

既存のシステムと評価の傾向が似ているため、音質同様、被験者の違いにより相対的に評価が下がっていると考えられる。相対的に見た場合、5[m/s]の評価が上がっているのは、既存のシステムよりもフレームレートが向上し、フレーム間での映像と音の変化が細かくなったためだと考えられる。また、固定音源より移動音源の方が音像の定位感がわかりやすいというコメントから、実際にMRで本システムを利用する場合は自由に頭を動かすことが可能なため、音像の定位方向がわかりやすくなると考えられる。他に、音が途中で飛んだように聞こえるというコメントがあったが、これは旋回中に前後誤判断が起り、定位方向が前後に移動してしまったためだと考えられる。

#### 4. 直接音の改善

本システムにおける音像の定位精度低下の原因の一つとして、前後誤判断が多く起こっていたことがあげられる。これは、ヘッドホンから提示する直接音に両耳間時間差と両耳間強度差しか考慮していないことが最も影響を与えていると考えられる。そこで、直接音を前後方向において差異をつけることでその問題を解決する。

後方から聞こえてくる直接音は耳の後ろに当たり回折して聞こえるため、回折しにくい高音部分は小さくなくとされる。また、実際に0度方向と180度方向からホワイトノイズを再生し、録音したものをスペクトル解析したところ、6.5[kHz]以上の周波数で9[dB]程度の減衰が見られた。そこで、カットオフ周波数6.5[kHz]で3, 6, 9[dB]減衰するフィルタを作成し、左右90度方向から後方にかけて30度毎に高音が3[dB]ずつ下がるようにフィルタを直接音に畳み込むシステムに変更した。

改良を加えたシステムにおいて、実験1, 2と同様の実験を行った。被験者は正常な聴力を持つ学生10名である。人工音の定位精度に関する評価結果を図9に示す。

改良前のシステムに比べ、全体的に正答率の向上が見られ、特に180度方向に関しては大きな向上が見られる。しかし、定位方向のばらつきも増加している。これは今回の改良が被験者自身のHRTFに合わず、さらに定位精度

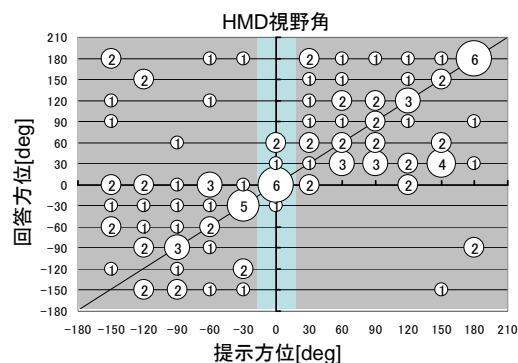


図9 人工音の定位精度の結果 (改良後)

が低下した被験者がいたからだと考えられる。また、依然として前後誤判断が見られるため、複数人のHRTFを平均するなどして、共通の特徴を正確に求め、システムに適用する必要があると考えられる。

#### 5. むすび

本稿では、視聴覚併用複合現実空間における個人特徴を用いない音の提示法として、直接音と反射音を分割提示する方法を提案し、その性能の評価を行った。結果、人工音の定位精度は既存のシステムに比べ30度などの斜めの方角に関して向上しており、反射音による音像定位の補助効果が確認できた。また、音質に関しても人工音の評価は向上しており、反射音を付加することにより臨場感が向上することも確認できた。しかし、全体的な定位精度としては低下していたため、直接音の改善を行うことによって定位性能の向上を図った。

改善案として、後方から音が到来した場合、音が耳の裏にあたり回折して聞こえるため、回折しにくい高音成分を低減させるという方法を実装し、定位精度の評価を行った。結果、多少の向上は見られたものの大きな向上が見られなかったため、この点については更なる検討が必要である。  
謝辞 本研究の一部は、科研費・基盤研究A「視聴覚併用複合現実空間の表現力向上に関する研究」による。

#### 参考文献

- [1] 比嘉嘉太, 他: “視覚・聴覚の現実と仮想を融合する2x2方式複合現実感システムの実現”, 日本VR学会論文誌, Vol. 13, No. 2, pp. 227 - 237, 2008.
- [2] 吉野将治, 他: “視覚・聴覚を併用した複合現実感システムの開発(4)―複合現実空間での音の反射・遮断の実現―”, 同学会第13回大会論文集, pp. 556 - 559, 2008.
- [3] 榎並和雅, 他: “超臨場感 (Ultra - Reality) の現状と将来像”, 日本音響学会誌, Vol. 64, No. 5, pp. 322 - 327, 2008.
- [4] 吉野将治, 他: “視覚・聴覚を併用した複合現実感システムの開発 (5) ―ヘッドホンとスピーカの併用による音提示法の拡張―”, 信学会, 信学技報, Vol. 109, No. 374, PRMU2009-200, pp. 305 - 310, 2010.
- [5] 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.
- [6] 大久保洋幸, 他: “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”, 日本VR学会論文誌, Vol. 5, No. 3, pp. 965 - 974, 2000.