

# 視覚と聴覚のクロスモーダル知覚を用いた 音像定位システムに関する基礎検討

王 夢<sup>1,a)</sup> 小川 剛史<sup>1,2,b)</sup>

**概要:** 拡張現実感における現実世界に重畳表示した仮想オブジェクトのリアリティを向上させるためには、視覚的な提示だけでなく、触覚や聴覚など、より多くの感覚刺激を提示することが重要である。本研究では、聴覚刺激の提示のみで任意の場所に音像を定位することは困難なため、視覚と聴覚のクロスモーダル知覚を用いたシステムを提案する。本稿では、提案システムの実現に向け視覚刺激が音像知覚に与える影響を調査した初期実験について報告する。実験により、視覚と聴覚のクロスモーダル知覚を用いることで、聴覚刺激のみを与えたときと比較して、被験者が音像の位置をより強く認識できることが分かった。

**キーワード:** 拡張現実感, 音像定位, クロスモーダル知覚

## A Basic Study on Auditory Lateralization System Using Auditory-Visual Cross-Modal Perception

OU YUME<sup>1,a)</sup> OGAWA TAKEFUMI<sup>1,2,b)</sup>

**Abstract:** In augmented reality (AR) environment, not only visual stimuli but also tactile or auditory stimuli are important to realize virtual objects with high reality. In this study, we propose an AR system for localizing a sound image in arbitrary position, using auditory-visual cross-modal perception. In this paper, we conducted two initial experiments to investigate the influence of the presentation of visual and auditory stimuli on the auditory lateralization. Experimental results suggest that subjects can recognize more strongly the position of the sound image by using auditory-visual cross-modal perception than only auditory perception.

**Keywords:** Augmented Reality, Auditory Lateralization, Cross-modal Perception

### 1. はじめに

セカイカメラ [1] のような現実世界に情報を付加する拡張現実感 (Augmented Reality, AR) 技術が身近なものになってきている。より現実世界に近く、自然に付加情報を提示するために、AR の整合性問題として幾何学的整合性、時間的整合性、光学的整合性の 3 つの解決すべき問題が存在

し、多くの研究 [2] [3] がなされてきた。しかし、上記の整合性は主に視覚刺激に関するもので、触覚や聴覚などの刺激に関しては対象となっていない。例えば仮想オブジェクトが、「そこに見えて」、「触ることができ」、「衝突すれば音がなる」といったように、現実世界と仮想オブジェクトの真の融合を実現するためには、さらに多くの問題を解決しなければならない。そこで、近年では AR の表現力向上のため視覚的な提示だけでなく、聴覚、触覚刺激等を提示する研究 [6][7] が盛んに行われている。頭部伝達関数 (Head Related Transfer Function, HRTF) とヘッドホンもしくはスピーカを用いた仮想音像提示は空間に音像を正確に定位することができるが、耳や頭部の形状に個人差が存在するため、他人の HRTF を利用すると音像位置の前後誤判定の現

<sup>1</sup> 東京大学大学院学際情報学府  
Graduate School of Interdisciplinary Information Studies,  
The University of Tokyo.

<sup>2</sup> 東京大学情報基盤センター  
Information Technology Center, The University of Tokyo.

<sup>a)</sup> y.ou@ogawa-lab.org

<sup>b)</sup> ogawa@ogawa-lab.org

象が多く生じる [5]。このようなシステムを不特定多数の受聴者で利用可能にするためには頭部伝達関数の個人差を解決する必要がある。また、スピーカアレイなどの多数のスピーカを利用した音像定位に関する研究 [8] がなされているが、装置が大掛かりであり汎用性に欠ける。本研究では、実空間内に設置した少数のスピーカからの聴覚刺激と、HMD に表示する視覚刺激を用いて、空間の任意の位置に音像を定位するシステムを提案する。その際、これまでの研究とは異なり、スピーカアレイのような大規模な装置は利用しないことを前提としている。本稿では、視聴覚のクロスモーダル知覚が人の音像知覚に与える影響を調べた初期実験について報告する。

## 2. 関連研究

本システムの目的は、よりリアルな AR 体験のために、視聴覚のクロスモーダル知覚を利用することで任意の位置に仮想音像を提示することである。本システムに要求される機能は以下の 2 点である。

- 大規模な装置を用いることなく、3次元空間に音像を定位する。
- 視覚刺激を付加することで、スピーカでは提示できない位置へ音像移動を実現する。

ここで、空間内で音源が移動しているように知覚することを音像移動と呼ぶ。

拡張現実感において視覚情報により、任意の位置に音像を定位する研究として、吉川らのラジオマーカ [9] が挙げられる。仮想オブジェクトが音を発する場合、オブジェクトが「そこ」にあるならば、「そこ」から音が聞こえる方が自然という考えをもとに仮想オブジェクトの存在感を強調することを目的としている。このシステムはスピーカ付きのマーカを複数配置し、仮想オブジェクトとスピーカの位置関係から、仮想オブジェクトの移動と共に各スピーカが出力する音量を調節することで、いかにも仮想オブジェクトのいる位置から音が鳴っているかのような音場を提示している。しかし、このシステムで定位できる音像の範囲は複数のスピーカ付きマーカが成す平面上に限定されており、3次元空間内の任意の位置に音像を定位することはできない。

伊藤らは 3D 音場を提示する「音像プラネタリウム方式」 [10] を導入し、複数の超音波スピーカを集約したユニットから、壁面や天井、空間内に用意した反射板に音波を反射させることで任意の場所に音像を定位させるシステムを構築している。さらに、このシステムで音像移動を実現するための研究として、スピーカの方向は変えずに複数の反射面に放射する音の音圧レベルを変更することで、音像が反射面間を移動するかのよう受聴者に知覚させる方法 [11] を提案した。これは、ユーザの座っている前に 2 枚の反射板を固定し、その反射版の間を仮想オブジェクトが

移動するのに応じてスピーカの音量を調整するシステムとなっている。この方法も反射板が作り出す平面内での移動は可能となるが、3次元空間内での移動は考慮されていない。また、このシステムは整えられた特定の環境内でしか利用できない。

どちらの先行研究も制限された空間内の任意の位置に音像を定位できるが、完全な 3次元空間内の任意の位置に音像を提示することはできない。本研究では、汎用的なスピーカを用い、体験者を中心とした完全な 3次元空間内の任意の位置に移動する音像定位システムを目指している。

## 3. 仮想音像定位の実現方法

### 3.1 概要

我々は後ろから車が向かってくるとき、見えていなくても車の位置を知覚できる。これは、ヒトの両耳が頭部の両側についているため、車が発する音が両耳に届くまでに生じる音の変化や差異を知覚できるからと考えられている。このように音の来る方向や距離を知覚できる現象を音像定位と呼ぶ。本研究では、2台のスピーカを受聴者の前方等間隔の位置に配置することで音場を構築し、受聴者の頭部を中心として、左右、遠近、上下の 3 方向に分解して音像定位の実現方法を検討する。

### 3.2 左右方向の音像定位

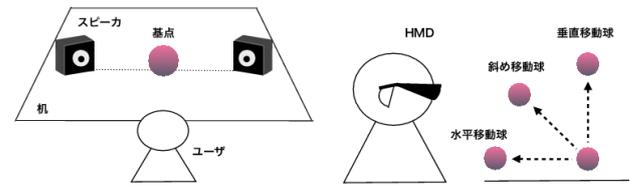
音の両耳への到達時間や音圧の差異による両耳間時間差、両耳間レベル差を利用することで左右の方向感を提示することができる。ステレオシステムは両耳間レベル差を工学的に応用したものである。2つのスピーカから同一の信号を同レベルで再生した場合、受聴者には一つの音像が知覚され、スピーカの間から音が再生させているように感じる。この現象を「ファントム音像」 [4] (図 1(a)) と呼ぶ。このとき、片方のスピーカの音圧レベルをあげることで、音像はそのスピーカの方に移動する。このように、スピーカから再生する音の大きさを制御することで、音像の左右の移動を実現することができる。本システムでは、左右方向の音像定位に「ファントム音像」を利用する。

### 3.3 上下方向の音像定位

ヒトの上下方向の定位精度は左右方向よりも低く、我々が日常生活において上下方向を知覚できるのは、両耳に到達する音の変化を、過去の記憶や経験から判断しているためである。一般的に、音源の位置によらずにヒトは高い周波数の音は高い位置に、低い周波数の音は低い位置に知覚する傾向がある。この現象は提示する信号の周波数によって特定方向に音像が知覚されるもので、方向決定帯域 [5] と呼ばれている。本システムでは提示する音の周波数を上下することにより、音像の上下方向の定位を検証する (図 1(b))。



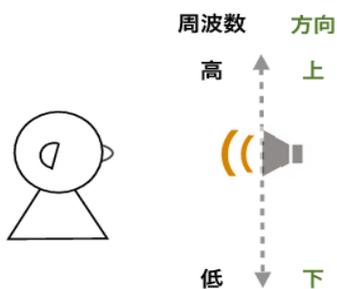
(a) 左右方向の音像定位  
上から見た図



(a) ユーザの後ろから見た図

(b) 横から見た図

図2 提示する視覚情報



(b) 上下方向の音像定位  
横から見た図



(c) 遠近方向の音像定位  
横から見た図

図1 音像定位

### 3.4 遠近方向の音像定位

遠近感は音像の距離の知覚に関連しており、音の大きさが遠近感の知覚に影響を与えていることは多くの研究で示されている。すなわち、音源の位置に関係なく、我々は

大きい音ほど遠くに、小さい音ほど近くに聞こえる [5] (図1(c))。さらに、直接音と反射音のエネルギー比や到達時間の差異、また音色の変化も距離知覚の手がかりになっていることがわかってきているが、本稿では遠近感付与の要素として音の大きさを変更するだけで受聴者の遠い地点から近い地点に移動する音像を定位することを実験で検証する。

### 3.5 クロスモーダル知覚を用いた補完

上記の方法で左右、上下、遠近の3方向の音像定位を行うが、2台のスピーカーのみを使用しているためスピーカー間の平面内に音場が生成されるだけで、3次元空間内、すなわち上下、遠近方向での音像定位は困難であると考えられる。そこで、仮想音像の生成位置に視覚情報を重畳提示することで、音像定位の3次元空間内での音像定位を実現する。

## 4. 実験

### 4.1 実験の目的と準備

本稿では、正確に知覚することが困難である、上下、遠近の両方向での移動音源の実現可能性を検証するために、被験者にHMDを装着させ、視覚刺激と聴覚刺激の両感覚からの音像位置の方向知覚について調査する。ここで、被験者は視覚刺激に強く影響を受け、提示する視覚情報と同じ方向に音像を知覚するという仮説を立てた。

提示する視覚刺激は図2に示すように、球の仮想オブジェクトを両スピーカーを結ぶ線分の midpoint を基点 (図2(a)) として鉛直方向、体験者側斜め上45度、体験者側水平方向への移動の3パターン (図2(b)) とした。本稿では、これらの視覚刺激を、それぞれ鉛直移動球、斜め移動球、水平移動球と呼ぶ。

提示する聴覚刺激は周波数の変化によるエネルギー増減の影響が少ないピンクノイズを利用し、表1に示すように、基準とする音源、周波数を変化させた音源、音圧レベルを変化させた音源、周波数と音圧レベルの両方を変化させた

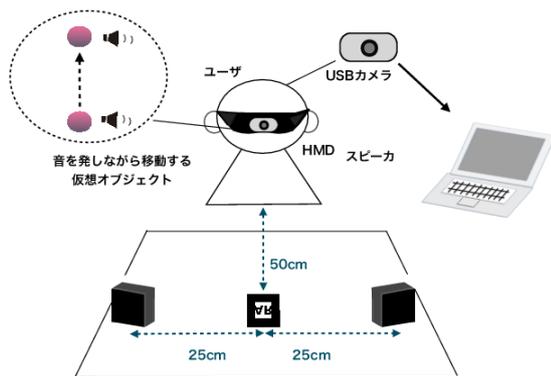


図3 システム概要

音源を用意した。本稿では、これらの4パターンの音源をそれぞれ、基準音、周波数変化音、音圧変化音、両要素変化音と呼ぶ。

ここで、基準音は中心周波数を 5500[Hz]、被験者の耳の位置で 62[dB] となる音源である。周波数は、バンドパスフィルタを利用し、仮想オブジェクトが 3[cm] 移動する毎に、ノイズの中心周波数を 5500[Hz] から、500[Hz] 刻みに高くする。音圧の変化は、仮想オブジェクトが 3[cm] 移動する毎に音圧レベルを約 1.1 倍大きくなるよう設定した。

実験ではシステムのサンプリング周波数を 16[kHz] とし、実験環境は騒音レベル 53[dB] の事務室環境で、被験者は正常な聴力を持つ成人 9 名である。実験環境は机上に 2 台のスピーカを配置することで構築する。両スピーカから出力される音源の音圧レベルや周波数を変化させて、4 パターンの仮想音像を実装した。被験者を椅子に座らせ、HMD を装着させた。スピーカは被験者の左右 25 [cm] の距離に配置した。被験者はスピーカの間に配置される基点を見る体勢となる (図 3)。

## 4.2 実験 1

### 4.2.1 聴覚刺激のみ提示した場合の音像知覚

3 名の被験者に対し、目隠した状態で上記 4 パターンの聴覚刺激をランダムに提示する。それぞれの聴覚刺激から得られる音像の位置や変化を回答するよう指示した。

### 4.2.2 実験結果と考察

提示した聴覚刺激に対して、各被験者が知覚した仮想音

表 1 音源の種類

音源名	周波数の変化	音圧の変化
基準音	なし	なし
周波数変化音	あり	なし
音圧変化音	なし	あり
両要素変化音	あり	あり

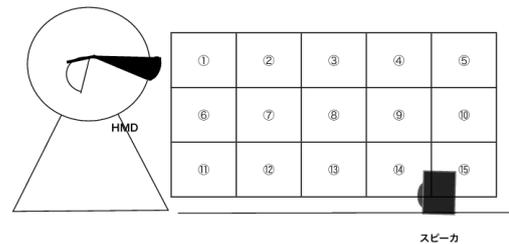


図 4 知覚する音像の位置

像の位置とその位置の変化を表 2 に示す。知覚した音像の位置は図 4 の番号で示し、位置の変化については被験者のコメントを記している。表 2 より、3 名の被験者が知覚した仮想音像の位置は大きくばらつき、こちらが想定した位置に音像を知覚できていないことが分かる。また、提示する聴覚刺激の変化には気付くものの、移動や方向などの動作の知覚に一貫性はなかった。頭部伝達関数を用いた音像定位を利用すれば、聴覚刺激のみの提示でも、より正確な音像位置の知覚が可能であると思われるが、本稿のような単純なシステムでは想定通りの位置に音像を知覚させることはできなかった。

## 4.3 実験 2

### 4.3.1 音源と同方向に移動する視覚刺激の知覚

4 種類の音源に対し、それぞれ 3 パターンの視覚刺激をランダムに提示した。被験者には「音を発しながら移動する仮想オブジェクトを数パターン見せる」ことを伝え、どの方向に移動する球と仮想音像が一番自然に感じたか答えるよう指示した。被験者は、実験 1 を行った 3 名に新たに

表 2 聴覚刺激のみ提示した場合の音像知覚

基準音	位置	変化
被験者 1	12 ↔ 13	少し前後移動
被験者 2	1	変化なし
被験者 3	7	変化なし
周波数変化音	位置	変化
被験者 1	13 ↔ 14	少し前後移動
被験者 2	7	少し奥に移動する感じ
被験者 3	8	少し手前に来る
両要素変化音	位置	変化
被験者 1	12 ↔ 15	前後移動
被験者 2	7	ふらふらと移動している
被験者 3	8	よくわからない
音圧変化音	位置	変化
被験者 1	6 ↔ 15	前後移動
被験者 2	12	よくわからない
被験者 3	6	よくわからない

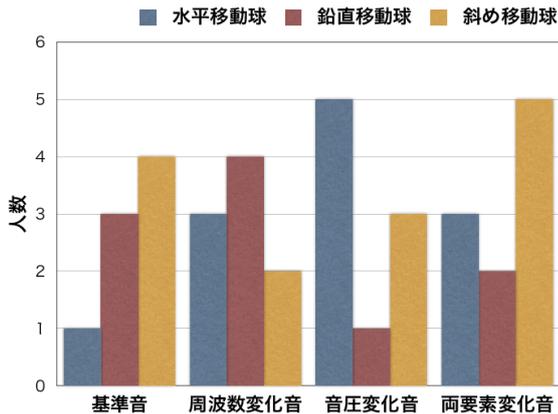


図5 各音源に対して自然と感じるオブジェクトの動き

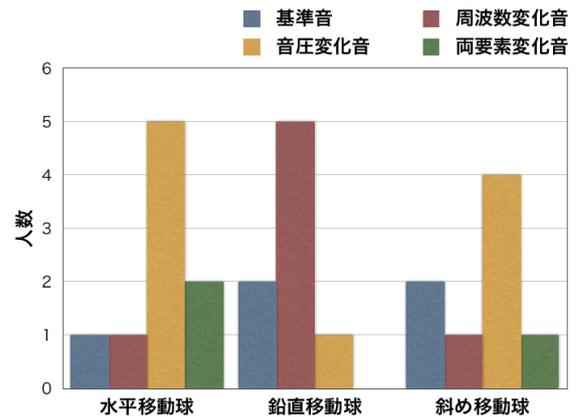


図6 各オブジェクトの移動に対して自然と感じる音源

6名を加え計9名の成人とした。

#### 4.3.2 実験結果と考察

被験者が最も自然に感じた視覚パターンと音源パターンを図5に示す。図5より、周波数変化音を提示した場合、9名中4名の被験者が鉛直移動球の動きが一番自然であったと回答した。また、音圧変化音を提示した場合、水平移動球の動きが一番自然と答える被験者が9名中5名いた。これらの結果より、上下方向は周波数の変化が影響し、遠近方向は音の大きさが影響していると考えられる。さらに、両要素変化音を再生した場合には、斜め移動球の動きが最も自然と答える被験者が9名中5名おり、これは上下、遠近の両要素を組み合わせることで、上下方向、遠近方向が作り出す平面上に音像を定位できる可能性を示している。一方、基準音を提示した場合、斜め移動球が自然であると回答した人数が多かった。これは被験者が音像に特に動きを感じていないため、人とスピーカと球の方向に関する位置関係が変わらない斜め移動球が最も自然であると感じたのだと考えられる。以上から、本システムで音像定位を知覚できる被験者もいるが、より正確に3次元空間内の任意の位置に音像定位するには、3方向の要素の組み合わせ方を検討課題として考えていく必要がある。

#### 4.4 実験3

##### 4.4.1 視覚刺激と同方向に移動する音像の知覚

3パターンの視覚刺激に対し4種類の聴覚刺激をランダムに提示した。被験者には「音を発しながら移動する仮想オブジェクトを数パターン見せる」ことを伝え、視覚パターンに対しどの音像提示が最も自然に聞こえるかを答えるよう指示した。この実験の被験者は、実験2と同じ9名の成人である。

##### 4.4.2 実験結果と考察

被験者が最も自然に感じた視覚パターンと音源パターンを図6に示す。図6より、9名中5名が水平移動球では音

圧変化音が自然であると回答した。同様に、9名中5名が鉛直移動球では周波数変化音が自然であると回答した。これにより、音圧レベルの変化は水平方向、周波数の変化は鉛直方向に影響を与えていると考えられる。斜め移動球では実験2の結果に反して、両要素変化音よりも、音圧変化音と回答した被験者が3名多く、音圧を重視した上で周波数の変化とのバランスを考慮する必要があると考えられる。

実験1、実験2、実験3の結果より、聴覚刺激のみ提示した場合、被験者は音像の位置を判断することが困難であったが、視覚刺激を付加することによって視覚情報とともに移動する音像を知覚することができることがわかった。また、視覚刺激を提示した場合、上下方向の音像定位は周波数の変化が影響し、遠近方向では音の大きさが影響していることが示唆された。斜め方向に音像を提示する場合には、提示する周波数と音圧レベルのバランスを検討する必要があることが分かった。

#### 5. おわりに

本稿では、仮想オブジェクトから音が発しているように知覚させるために、仮想音像の生成位置に視覚情報を重畳提示することで、3次元空間内の任意の位置における音像定位を実現することを目的とし、初期検討として、鉛直方向、垂直方向、斜め方向に移動する仮想音像を、周波数、音圧の変化、視覚刺激の提示より知覚できる可能性を被験者実験により示した。また、斜め方向の音像定位を実現するには、周波数と音圧の提示するバランスを考慮する必要があると分かり、今後検討する予定である。さらに、本稿では考慮しなかった左右方向の音像定位についても調査し、上下、左右、遠近の3方向の組合せによる任意の位置への音像定位を実現する方式を検討する予定である。

**謝辞** 本研究の一部は日本学術振興会科学研究費補助金基盤研究(C)(25330227)の研究助成によるものである。ここに記して謝意を表す。

## 参考文献

- [1] 頓知ドット: “セカイカメラ,”  
<http://sekaicamera.com/> (2013 年 6 月現在).
- [2] M. kanbara and N. Yokoya: “Geometric and Photometric Registration for Real-time Augmented Reality,” Proc. of International Symposium on Mixed and Augmented Reality (ISMAR02), pp. 279–180, 2002.
- [3] 池内克史, 佐藤洋一, 西野恒, 佐藤いまり: “複合現実感における光学的整合性の実現,” 日本バーチャルリアリティ学会論文誌, Vol. 4, No. 4, pp. 623–630, 1999.
- [4] 沢口真生, 中原雅考, 亀川徹: “サラウンド入門,” 東京藝術大学出版, 2010.
- [5] 飯田一博, 森本政之: “空間音響学,” コロナ社出版, 2010.
- [6] J. Sodinik, S. Tomazic, R. Grasset, A. Duenser, and M. Billinghamhurst: “Spatial Sound Localization in an Augmented Reality Environment,” Proc. of OzCHI2006, pp. 111–117, 2006.
- [7] 新島有信, 小川剛史: “拡張現実感における視覚刺激位置が触知覚位置に与える影響の分析,” 日本バーチャルリアリティ学会論文誌, Vol. 17, No. 2, pp. 73–78, 2012.
- [8] “スピーカアレイ装置、信号処理方法およびプログラム,” ヤマハ株式会社, 特開 2006-25153 号公報.
- [9] 吉川祐輔, 宮下芳明: “ラジオマーカ: 仮想オブジェクト位置に音像を再現する AR マーカ,” インタラクシオン 2010 講演論文集, Fresh From the Oven Session, 2010.
- [10] 杉林裕太郎, 栗本総太, 森勢将雅, 西浦敬信, 柴田史久: “壁面反射型オーディオスポットを用いた高臨場感 3D 音場提示手法の検討,” 日本バーチャルリアリティ学会第 15 回大会論文集, 1C4–3, pp. 204–207, 2010.
- [11] 伊藤仁一, 森勢将雅, 西浦敬信, 木村朝子, 柴田史久: “超音波スピーカによる移動音源の実現と視覚補助による定位性能の向上 (2) —音像の移動範囲の拡張—,” 日本バーチャルリアリティ学会第 16 回大会論文集, 12D–5, pp. 151–154, 2011.