



# VR コンテンツにおける人声の放射点を考慮したレンダリング法の効果

Effectiveness of Rendering Method Considering Radiation Point of Human Voice in VR Content.

杉本岳大<sup>1)</sup>, 木下光太郎<sup>1)</sup>, 中山靖茂<sup>1)</sup>

Takehiro SUGIMOTO, Kotaro KINOSHITA, and Yasushige NAKAYAMA

1) 日本放送協会 (〒 157-8510 東京都世田谷区砧 1-10-11, sugimoto.t-fg@nhk.or.jp)

**概要:** VR コンテンツにおける音のレンダリングでは、発音源とリスナとの間の距離に基づいて音圧に距離減衰を付与するので、発音源の正確な位置や伝搬経路の情報が不可欠である。本稿では、これまでに提案した人声の放射点およびリスナまでの伝搬経路を考慮した人声のレンダリング法の効果を検証する目的で、VR コンテンツによる主観評価実験を実施した。実験結果から、発声する人物の頭部中心に人声の音声オブジェクトを配置して頭部中心を規準とする等方的な距離減衰を付与するシンプルな手法よりも、提案手法の方が有意に好まれることがわかった。

**キーワード:** VR, 開口端補正, 距離減衰, 人声, MPEG-I Immersive audio

## 1. 緒言

イマーシブメディアは、あたかもその世界に飛び込んだかのような体験を提供できる次世代のメディアであり、ゲーム業界をはじめとして様々な分野で研究開発が進んでいる [1-3]。NHK でもイマーシブメディアを今後の放送技術の進展における重要な研究課題の一つと位置づけており、AR (拡張現実) や VR (仮想現実) を活用したコンテンツの制作や提示のための技術開発を進めている [4]。

イマーシブメディアの映像には 360° 映像やボリュメトリック映像が採用されており、コンテンツ空間内の任意の位置でコンテンツを楽しむことが特徴である [5]。そこでイマーシブメディア用の音響技術も、リスナの自由な移動に追従して音をレンダリングできるように、six-degrees of freedom (6DoF, 図 1 を参照) 対応が期待されている [6]。

ところで VR コンテンツでは、発音源とリスナとの間の距離に基づいて音圧に距離減衰を付与するので、発音源の正確な位置や伝搬経路の情報は距離減衰の計算には不可欠である。発音源の音声信号が人声の場合、シンプルな実装では人物の頭部中心を人声の放射点とみなす場合があるが、人声の放射点が頭部中心でないことは自明である。特に人声は最も馴染み深い音声オブジェクトの一つであるため、コンテンツへの没入感を阻害しないように、可能な限り実際の物理現象に則して、振る舞いを模擬することが望ましい [7,8]。

我々はこれまで、人声の放射点およびリスナまでの伝搬経路を考慮した人声のレンダリング法を提案してきた [9-11]。本稿では、6DoF 動作が可能な VR コンテンツを用いて、提案手法の効果を主観評価実験で検証した結果を報告する。なお本稿の検討は、MPEG-I Immersive audio [12,13] の規格化活動の一環として実施した。

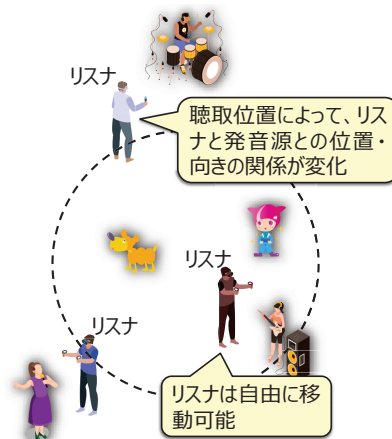


図 1: 6DoF 音響空間の模式図。6DoF では、聴取位置によってリスナと発音源との位置・向きの変化が変化する。

## 2. 人声の放射点を考慮したレンダリング法 [9-11]

本節では、人声のレンダリングにおける提案手法を説明する。図 2 に提案手法の模式図を示す。図 2 では人物の頭部を剛体球と仮定し、声道における開口端補正の影響を考慮して人声の放射点を口唇の数 cm 前方に設定している。次に、コンテンツ空間におけるリスナの聴取位置に応じて、以下の手続きで音圧の距離減衰を算出する。

**条件 1** リスナ位置から人声の放射点が見通せる場合、距離の逆二乗則に基づく距離減衰を付与する。

**条件 2** 人物の頭部の陰になってリスナ位置から人声の放射点が見通せない場合、頭部周囲の回折による伝搬経路の伸長を考慮して、距離の逆二乗則に基づく距離減衰を付与する。

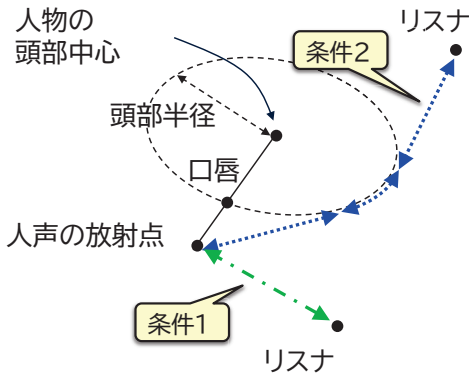


図 2: 提案する人声のレンダリング法の模式図。一点鎖線(緑)が条件 1 に、点線(青)が条件 2 に対応する。

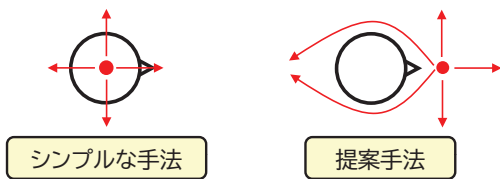


図 3: 左: シンプルな手法の模式図。右: 提案手法の模式図。

本稿では、提案手法の効果を検証するために、人物の頭部中心に人声のオブジェクトを配置し、頭部中心を規準として等方的な距離減衰を付与する手法（以下、シンプルな手法）との比較をおこなう。図 3 に、シンプルな手法と提案手法それぞれの模式図を示す。

### 3. 実験環境

#### 3.1 実験システム

実験システムは、MPEG-I Immersive audio の評価環境の規準を満たす機材で構成した。

- Platform PC
  - Graphics Card: nVIDIA / GeForce RTX 3090
  - Processor: Intel(R) / Core i9-10900K, 10 cores, 20 logical processors
  - Memory: 64GB RAM
  - Storage: 2TB SSD
  - OS: Windows 10 Pro 64-bit
- VR system: HTC / VIVE Pro
- Sound card: Focusrite / Scarlett 2i2 3rd Gen
- Headphone: Beyerdynamic / DT-990 Pro

コンテンツの映像は Unity を用いて再生した。また音声は MPEG-I Immersive audio の規準モデル [13] に対してシンプルな手法および提案手法を実装してレンダリングし、Max/MSP 経由で再生した。



図 4: 主観評価実験を実施した防音室。

#### 3.2 実験室

主観評価実験は、ラジオスタジオのフロアを模した防音室で実施した(図 4)。評価者が実空間を歩き回れる範囲は、縦 3 m × 横 4 m とした。

### 4. 主観評価実験

#### 4.1 実験手法

評価は、7 段階の評価スケールを有する A/B テストでおこなった。A/B の一方にはシンプルな手法に基づいてレンダリングした音、もう一方には提案手法に基づいてレンダリングした音がランダムに割り当てられる。評価者は、映像と同時に提示されるレンダリング音を聴取し、コンテンツを表象する音として尤もらしいか (plausible) の観点に基づいて、定位・音量・音色・距離感などを総合して音質を評価した。7 段階の評価スケールを以下に示す。

- +3: A が非常に良い
- +2: A が良い
- +1: A がやや良い
- 0: A と B は同じ
- 1: B がやや良い
- 2: B が良い
- 3: B が非常に良い

評価者は現役の音響研究者 5 名であり、コンテンツ内の GUI 上に表示された A/B のボタンで音を切り替えながら、自由な位置・方向で評価した。A/B を切り替えると再生音は変化するが、映像は途切れずに再生され続ける条件である。

#### 4.2 評価コンテンツと刺激

評価コンテンツには、MPEG-I Immersive audio の規格化用に Philips が制作した “Hospital” を使用した。このコンテンツには、医師が発声しながら廊下を歩いて往復し続けるシーンが含まれており、医師が発した人声に対していずれかの手法を適用した。今回の評価実験では、発音源が移動しない条件でも検証するために、医師が静止しているシーンも制作した。

図 5 に評価に用いた “Hospital” を示す。医師の人声の放射特性には、カナダ国立研究機構が測定したデータを使用しており [14]、放射角度の分解能は 15° である。

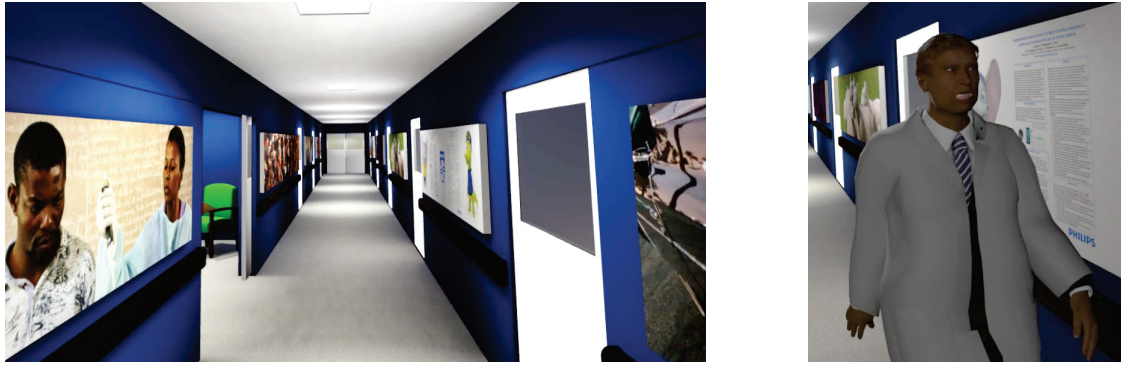


図 5: Philips 制作の評価コンテンツ “Hospital”. 廊下の様子 (左) と発声しながら歩行する医師 (右) .

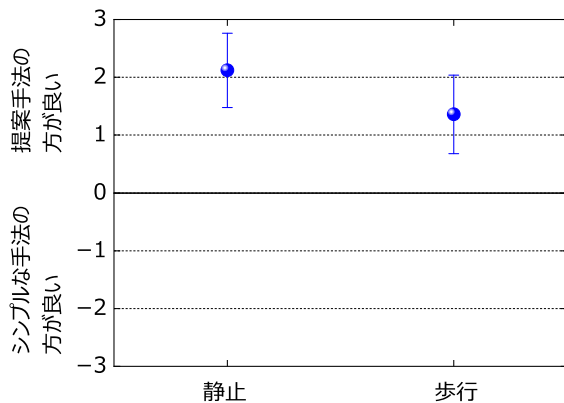


図 6: 主観評価実験の結果. 丸印は評点の平均、エラーバーは 95%信頼区間を示している.

刺激は、評価者の位置・向きに合わせてリアルタイムにレンダリングした。またバイノーラル化には、ベルリン工科大学が中心となって測定したダミーヘッドの頭部伝達関数を用いた [15]。なお、提案手法のレンダリングでは、医師の頭部半径を 10 cm に設定した。

#### 4.3 実験結果と考察

図 6 に実験結果を示す。丸印は評点の平均、エラーバーは 95%信頼区間を示している。図 6 から、静止、歩行の両条件ともに提案手法が好まれたことがわかる。

また、歩行の方が静止よりも評点の方が下がる傾向が見られている。評価者の内観報告から、移動し続ける音声オブジェクトの音響特性の比較は、静止している音声オブジェクトの音響特性の比較よりも、適切な音量で聴取できる時間が限られることから、評価しづらいと感じるタスクであったことがわかっており、その影響と考えられる。さらに、評価コンテンツは音刺激の切り替えに関係なく連続して再生されるため、評価コンテンツの同じ箇所同士を直接比較することができないことも、評価の難易度が上がった一因と考えられる。

## 5. 結言

本稿では、6DoF 動作が可能な VR コンテンツを用いて、提案手法の効果を主観評価実験で検証した。その結果、人

物の頭部中心に人声のオブジェクトを配置して頭部中心を規準とする等方的な距離減衰を付与するシンプルな手法と比較すると、人声の放射点とリスナまでの伝搬経路を考慮した提案手法が有意に好まれることがわかった。

今後は、提案手法を人声以外に拡張した場合の効果についても検証を継続していく。

## 参考文献

- [1] Immersive media technologies: The acceleration of augmented and virtual reality in the wake of COVID-19, White paper, World Economic Forum, 2022.
- [2] M.-A. Moïnnereau, A. de Oliveira Jr., and T. Falk: Immersive media experience: A survey of existing methods and tools for human influential factors assessment, Quality and User Experience, Vol. 7, No. 5, 2022.
- [3] Rec. ITU-R BT.2420-6: Collection of usage scenarios of advanced immersive sensory media systems, 2024.
- [4] NHK 放送技術研究所: Future Vision 2030-2040, [https://www.nhk.or.jp/str/future\\_vision/index.html](https://www.nhk.or.jp/str/future_vision/index.html)
- [5] 青木秀一: より自由な映像体験に向けたイマーシブ映像の標準化動向 —360 度映像からイマーシブメディアへの進化—, 日本音響学会誌, 78 巻 3 号, pp. 150–156, 2022.
- [6] 杉本岳大, 堀内俊治: イマーシブメディア実現に向けた音響技術, 映像情報メディア学誌, Vol. 77, pp. 17–21, 2023.
- [7] T. Sugimoto, T. Hasegawa, T. Iwasaki, H. Kubo, S. Oode, and H. Okubo: Requirements for 6DoF audio service and 6DoF audio metadata, 2023 AES Intl. Conf. Spatial and Immersive Audio, Conference paper 33, 2023.
- [8] 杉本岳大, 長谷川知美: 6DoF 音響コンテンツの制作要件, 映像情報メディア学会年次大会, 32B-4, 2023.
- [9] T. Sugimoto and K. Kinoshita: Modeling of distance attenuation for human voice in 6DoF audio content

considering exact radiation point, AES Europe 2024, Convention express paper 211, 2024.

- [10] 杉本岳大, 木下光太郎: 人声の放射点を考慮した距離減衰特性の付与方法, 日本音響学会講演論文集, pp. 1471–1472, 2022.
- [11] 木下光太郎, 杉本岳大: 人声の距離減衰の聴感印象評価, 日本音響学会講演論文集, pp. 1037–1038, 2022.
- [12] ISO/IEC 23090-4: Immersive audio (規格化中) .
- [13] J. Herre and S. Disch: MPEG-I Immersive audio – Reference model for the virtual/augmented reality audio standard, J. Audio Eng. Soc., Vol. 71, pp. 229–240, 2023.
- [14] W. T. Chu and A. C. C. Warnock: Detailed directivity of sound fields around human talkers, Research Report No. RR-104, National Research Council of Canada, Institute for Research in Construction, 2002.
- [15] F. Brinkmann, A. Lindau, S. Weinzierl, G. Geissler, S. van de Par, M. Müller-Trapet, R. Obdam, and M. Vorländer: High resolution and full-spherical head-related transfer function database for different head-above-torso orientations, J. Audio Eng. Soc. Vol. 65, No. 10, pp. 841–848, 2017.