

背景音が仮想音響空間に与える聴感上の影響  
The effects of ambient sounds on a 3D virtual sound space quality

大谷 真<sup>†</sup> 岩谷 幸雄<sup>†</sup> 千葉 武尊<sup>†</sup> 鈴木 陽一<sup>†</sup>  
Makoto OTANI<sup>†</sup> Yukio IWAYA<sup>†</sup> Takeru CHIBA<sup>†</sup> Yôiti SUZUKI<sup>†</sup>

### 1. はじめに

日常生活において、我々は様々な音に囲まれて暮らしている。これらの音は、音源から放射されたのちに様々な物理現象、例えば、残響・回折・ドップラ効果など、を通して環境の影響を受けて我々の両耳に届く。我々は、両耳に入力される音響信号に含まれる両耳間時間差・レベル差などの両耳手がかりや周波数特性上の手がかりを利用して、これらの複雑な音の中から直接音の到来方向を知覚することができる。これらの手がかりは、音源から受聴者の両耳までの音響伝達関数を表す頭部伝達関数 (head-related transfer functions: HRTF) [1]と室伝達関数 (room transfer function: RTF)に含まれる。近年、高度なテレコミュニケーションシステムの実現を目的とした音空間の3次元レンダリング手法の開発が行われている。これを実現するためのバーチャルリアリティ技術として仮想聴覚ディスプレイ (virtual auditory display: VAD) がある。VAD システムを実現するための1つのアプローチとして、音源信号にHRTFを合成することで位置情報を付与する手法がよく用いられる。この場合、音声や楽器音などといった特定のオブジェクトに関連付けられる(ターゲット)音を3次元的にレンダリングすることが可能であるが、多くの場合、反射や回折といった直接音以外の音響物理現象はレンダリングされない。しかし、対照的に、日常的な音空間において我々はターゲットからの直接音だけではなく背景音も同時に聴取している。結果として、VADによって呈示される音空間はしばしば不自然なものとなってしまふ。したがって、VADシステムによって受聴者に自然かつ高い臨場感を与える音空間を呈示するためには、ターゲット音だけでなく背景音をレンダリングする必要があると考えられる。

従来の研究により、背景音を再現するための手法がいくつか提案されている。Lokkiらはダミーヘッドによって実測された背景音を用いた[2]。また、Sekiらは視覚障害者のための聴覚訓練システムを開発しており、ターゲット音源と比較して遠方に配置した音源により背景音を模擬している[3]。これらの手法では背景音を実測する必要があり、また、VADによって高精度なレンダリングを実現するためには、その音響出力を頭部運動に対応して変化させる必要があるが、この場合(水平角や仰角における)すべての方向に対して背景音を実測しておく必要がある、現実的ではない。

そこで、本稿では、より容易かつ効果的な背景音のレンダリング手法を提案するとともに、提案法によりレンダリングされた背景音の効果に関する主観評価実験の結果について報告する。

<sup>†</sup> 東北大学電気通信研究所 Research Institute of Electrical Communication, Tohoku University

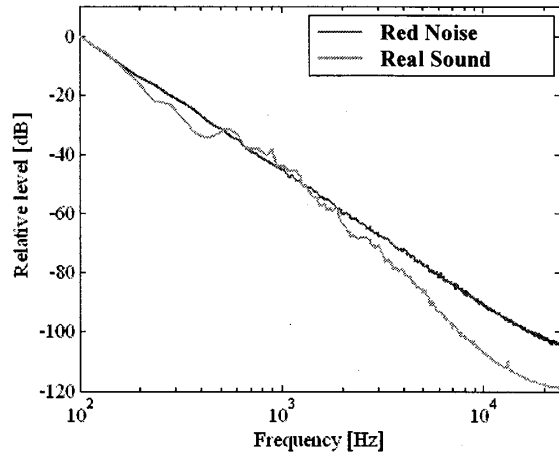


図1 レッドノイズと実際の背景音の周波数特性の比較

## 2. 背景音のレンダリング手法

### 2.1 音源信号の選択

現状のバーチャルリアリティ技術を用いて音空間における全ての物理現象を正確に再現することは難しい。そこで、本研究ではまず模擬的に背景音をレンダリングする手法に焦点を当てる。

まず、背景音として使用する音源信号の選択を行う。SMILE ライブラリ[4]に収録されている背景雑音の周波数特性を、ホワイト/ピンク/レッド(ブラウン)ノイズなどの統計的に定義される典型的なノイズのそれと比較した。その結果、これらのうち、レッドノイズの周波数特性が実際の背景音のものに最も近いことが分かった。そこで、背景音の音源信号としてレッドノイズを用いることとした。図1に、レッドノイズとSMILEライブラリに収録されている屋外の騒音(example09 city sounds/26401.wav)の周波数特性の比較を示す。

### 2.2 レンダリング法

ターゲット音にモノラル音源信号を加えれば、空間情報を含まない背景音を付与することは可能である。この場合、計算コストは小さいものの、再現された背景音は頭内に定位し、不自然な音空間が呈示されることになる。この問題を解決するために、背景音の音源信号(レッドノイズ)に様々な音源位置に対する頭部インパルス応答(HRTFの時間領域表現: head-related impulse response: HRIR)を畳み込み、受聴者の頭外に音像が定位されるように処理を施した。図2にこれらの仮想音源の位置を示す。水平角で0度から350度まで10度ごと、仰角で-30度か

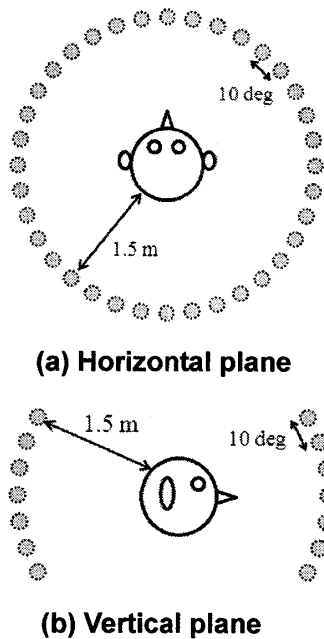


図2 提案法における仮想音源の位置。

ら30度まで10度ごと、とし、合計で252の音源位置に対するHRIRを用いた。HRIRは東北大学電気通信研究所の無響室に設置された球状スピーカアレイにより測定したダミーヘッド(SAMRAI, 高研)のものである[5]。スピーカアレイの半径すなわち音源距離は1.5 mである。仰角-30度以下では胴体や椅子などの影響が大きいため、-30度以下の仰角に対するHRIRは用いなかった。また、対称性を保つために、30度以上のHRIRも用いないものとした。

前述のように音源信号としてレッドノイズを用いるが、呈示される各仮想音源の音源信号の間の相関が高い場合、背景音が頭内に定位してしまうため、252の仮想音源の音源信号の相関が低いことが望ましい。そこで、非常に長い周期と均等性を持つ疑似乱数生成が可能なMersenne twister アルゴリズム[6]を用いて位相がランダムな252の音源信号を生成した。このようにして合成された背景音を、ここでは「周囲配置型背景音」と呼ぶ。

### 3. 実験 1

聴取実験による主観評価により、背景音の付加、及びレンダリング手法の違いが、再現された音空間の「リアリティ」に対してどのような知覚上の影響を与えるのかについて検討した。本研究では、「リアリティ」を「(音空間の)もってもらしさ」と定義する。3つのレンダリング条件に関する一対比較法により心理距離尺度値を測定した。

実験には、著者らが開発したVADシステム[7]を用いた。本システムでは、磁気センサ(FASTRAK, Polhemus)により120 Hzの更新頻度で受聴者の頭部向きを取得し、頭部運動に応じて聴覚への入力信号を変化させることができる。センサの受信器はヘッドホンのヘッドバンドの上に取り付けた。HRIRの信号長は48 kHzのサンプリング周波数で256点である。

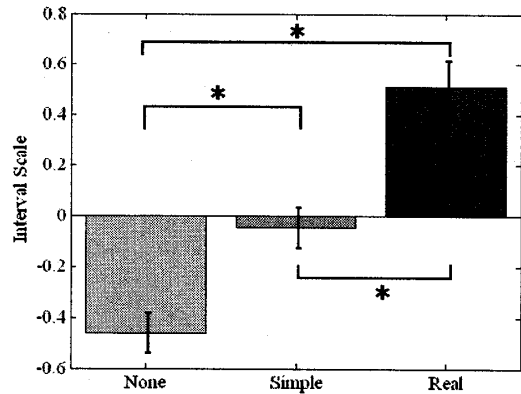


図3 実験1の結果。全被験者の心理距離尺度値の平均値と標準誤差

#### 3.1 方法

##### 3.1.1 被験者

聴力健全な8名の成人(21-24歳)が実験に参加した。

##### 3.1.2 装置及び刺激

実験は防音室で行った。聴覚刺激はUSBオーディオインターフェイス(UA-25, Edirol)、増幅器(AX-9, YAMAHA)、ヘッドホン(HDA200, Sennheiser)を用いて両耳に呈示した。

聴覚刺激はターゲット音と背景音の2つで構成される。ターゲット音としてSMILEライブラリより4つの楽器音(チェロ・フルート・オーボエ・バイオリン)を使用した。ターゲット音はダミーヘッド(SAMRAI, 高研)のHRIRを上記の音源信号に畳み込むことによって作成した。音空間におけるターゲット音の位置を固定するために、被験者の頭部運動に応じてHRIRを切り替えた。背景音はレッドノイズとバイノーラル録音した実環境音の2種類を使用した。実環境音は、集中空調装置から放射される背景音のみが存在するセミナールーム内に設置したダミーヘッドを通じて録音した音を用いた。ターゲット音と背景音の間のSN比は-5, 0, 5, 10 dBの4条件とした。刺激の持続時間は約10 sとした。

##### 3.1.3 実験手順

実験は一対比較法を用いて行なった。被験者には、継続的に呈示された2つの刺激対のうち、「どちらがより普段聞いている音に近いか」を選択させた。刺激呈示中は被験者の頭部運動に制限は加えなかった。実験条件として以下の3条件を設定した。None条件ではターゲット音のみを呈示した。Simple条件ではターゲット音とモノラルのレッドノイズを呈示した。Real条件では、ターゲット音と実環境音を呈示した。これら3レンダリング条件間で比較を行った。その際SN比と楽器音の種類ごとに比較対を作成したため、刺激対は合計48対となった。

#### 3.2 結果及び考察

呈示された音空間のリアリティを定量化するために、一対比較により得られた判断比率をもとにThurstonの比較判断の法則(ケース5)により心理距離尺度値を算出した[8]。心理距離尺度値の値が大きいかほど音空間がよりもってもらしさと知覚されたことを示す。

図3に全被験者間の心理距離尺度値の平均値を示す。この値は *Real* 条件で最も大きく、*None* 条件で最も小さかった。繰り返しのある一要因の分散分析を行った結果、背景音レンダリング条件間に有意差が認められた ( $F(2,20) = 31.26, p < .05$ )。また、Bonferroni の不等式に基づく多重比較の結果、すべての対の間に有意差が見られた (*Simple-Real* 条件,  $p < .05$ ; *Simple-None* 条件,  $p < .05$ )。

これらの結果は、背景音の提示が音空間のリアリティを向上させることを示唆している。しかし、モノラルの背景音は実環境音ほどのリアリティを与えなかった。この原因の1つとして、*Simple* 条件では、ターゲット音が頭外に定位するのに対して背景音は頭内に定位してしまったことが考えられる。実際に数人の被験者は聴覚刺激に違和感を覚えたことを報告している。そこで、実験2では頭外定位するように提示した背景音の効果について検討した。

#### 4. 実験2

実験2では、2.2節で述べた周囲配置型背景音の提示が音空間のリアリティの知覚に及ぼす影響を検討した。

##### 4.1 方法

被験者、装置、刺激、手順は、実験1と同様であった。*None* 条件の代わりに、我々の提案法を用いて作成した周囲配置型背景音を使用する *Surround* 条件を設定した。*Surround* 条件では、ターゲット音は実験1と同様に提示するが、2.2節で述べたようにダミーヘッドの HRIR を畳み込んだレッドノイズ (周囲配置型背景音) を背景音として提示した。

##### 4.2 結果及び考察

図4に全被験者の心理距離尺度値の平均値を示す。平均心理距離尺度値は *Real* 条件で最も大きく、*Simple* 条件で最も小さかった。分散分析の結果、背景音のレンダリング条件の主効果が有意であった ( $F(2,23) = 12.06, p < .01$ )。多重比較を行った結果、*Simple-Real* 条件間 ( $p < .05$ ) 及び *Simple-Surround* 条件間 ( $p < .05$ ) に有意差が見られた。

これらの結果は、被験者は *Simple* 条件と比較して *Real* 条件により提示される音空間の方がよりリアリティが高いことを示しており、実験1の結果が確認された。一方、*Real-Surround* 条件間の間に有意な差は認められなかった。このことは、音空間のリアリティの知覚に対して、周囲配置型背景音が実際の環境でバイノーラル録音された背景音と同程度の効果を持つことを示唆している。

#### 5. 総合考察

上記の実験の結果から、背景音の付与が音空間のリアリティの知覚に影響を与えると結論付けることができる。さらに、実験2の結果は、*Real-Surround* 条件間に有意差がないことを示しており、レンダリングにおいて周囲配置型背景音を付与することで音空間のリアリティを向上させることが可能であると言える。

一方、数人の被験者は、*Real* 条件と *Surround* 条件においてターゲット音と背景音が分離して知覚され不自然な感じがしたと報告している。実際の音空間と今回の実験

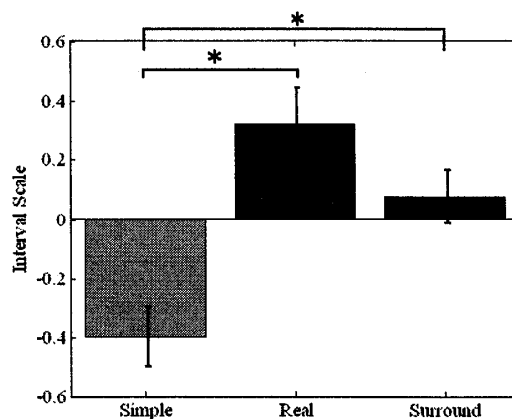


図4 実験2の結果。全被験者の心理距離尺度値の平均値と標準誤差

条件の大きな違いは、背景音提示が頭部運動に対応しているかどうかにある。このことが不自然な印象を与える原因の一つであるかもしれない。背景音の知覚に関する頭部運動の影響は今後の課題である。

#### 6. まとめ

本研究では、背景音のレンダリング手法及びその知覚上の影響について検討を行った。レッドノイズの周波数特性が実際の背景音のそれに類似していたため、レッドノイズを音源信号として用いた。

HRIR との畳み込みを用いて、無相関な位相特性を持つレッドノイズを放射する複数の音源を空間的に配置した。実験の結果、空間化された背景音はリアリティ知覚の改善に効果的であり、人工的な雑音を加えることで仮想音空間をより現実的なものとするのが可能であることが分かった。しかし、数人の被験者は、提示された仮想音空間が不自然であったと報告している。今後は、この不自然さを解決するために頭部運動と背景音知覚の関係について検討を行う予定である。

#### 謝辞

本研究は、科研費 (基盤C) No.20500110, 科研費 (特別推進) No.19001004, 東北大学電気通信研究所共同プロジェクト研究, 東北大学グローバル COE プログラム CERIES により行った。また、有益なご助言ご助力をいただいた東北大学 小林まおり博士に感謝する。

#### 参考文献

- [1] J. Blauert, *Spatial Hearing*. The MIT Press, Cambridge, Massachusetts, 1997.
- [2] T. Lokki, and H. Järveläinen, "Subjective evaluation of auralization of physics-based room acoustics modeling," *Proc. of the 2001 International Conference on Auditory Display*, Espoo, July 29-August 1, 2001.
- [3] Y. Seki and T. Sato, "Development of auditory obstacle perception training system for the blind by using 3-D sound," *Proc. of Conference and Workshop on Assistive Technologies for Vision and Hearing Impairment 2006*, Kufstein, 2006
- [4] K. Kawai, K. Fujimoto, T. Iwase, T. Sakuma, Y. Hidaka and H. Yasuoka, "Introduction of sound material in living environment 2004 (SMILE 2004): A sound source database for educational and

- practical purposes," *4th Joint Meet. ASA/ASJ* (Honolulu), *J. Acoust. Soc. Am.* **120**, pp.3070-3071, 2006.
- [5] S. Yairi, Y. Iwaya and Y. Suzuki, "Individualization feature of head-related transfer functions based on subjective evaluation," *Proc. of International Conference on Auditory Display (ICAD2008)*, Paris, Jun.24-27, 2008.
- [6] M. Matsumoto and T. Nishimura, "Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator," *ACM Trans. on Modeling and Computer Simulations*, 1998.
- [7] S. Yairi, Y. Iwaya and Y. Suzuki, "Development of virtual auditory display software responsive to head movement," *Trans. Virtual Reality Soc. Jpn* (in Japanese), **11**(3), pp.437-446, 2006.
- [8] LL. Thurstone, "A law of comparative judgement," *Psychol Rev*, **35**, 273, 1927.