

聴覚ディスプレイと音空間知覚

岩谷 幸雄^{†1}

概要: 聴覚に対して音空間を提示（ディスプレイ）するシステムを聴覚ディスプレイという。聴覚ディスプレイは、2チャンネルの信号を合成して両耳に与える方法と、音空間から切り出した波面をスピーカアレイ等によって合成する方法がある。一方、聴覚ディスプレイシステムを高度化することと、人間の音空間の知覚の理解は表裏一体である。本発表では、音空間のレンダリング法とシステム構成法を議論し、システムを高精度化するための音空間知覚の理解のための聴取実験について紹介する。

キーワード: 聴覚ディスプレイ、音空間レンダリング、音空間知覚

Virtual Auditory Display and Spatial Hearing

Yukio IWAYA^{†1}

Abstract: Virtual auditory display (VAD) is a system which display three-dimensional acoustic space to human auditory system. VAD can be realized mainly by two methods. One is given by acoustic two signals to both ears, and another can be a realized loudspeaker array, which is used to synthesize virtual wave front. Furthermore, it is required that we understand auditory process in three-dimensional space well in order to construct an advanced VAD system. In this presentation, rendering method of acoustic space and construction method of VAD, and consideration of spatial hearing are discussed.

Keywords: Virtual auditory display, rendering method of acoustic space, spatial hearing

1. はじめに

音源が空間にあるとき、音源が持つ性質（スペクトル、時間情報、指向性など）と、反射・残響・回折などの伝搬特性が、畳み込まれ、我々の両耳に届く。我々は、知覚された音の信号から、音源の姿（音像）を知覚する。つまり両耳信号の差を使うことによって、視覚と同様に、3次元空間における音像の姿を意識することが可能となる。音像は、位置、音色、動き、指向性など様々な属性を持つ。特に、位置に特化した場合を音像定位と呼ぶ[1]。

我々が、知覚において使う信号数は、たかだか2チャンネルであるから、鼓膜に元音場の信号を与えることができれば、元音場と全く同じ音空間を体験することができる。このような目的の音空間の提示システムを一般に聴覚ディスプレイ(Virtual Auditory Display: VAD)という。VADは、ヘッドホン等を用いて、直接鼓膜に対して所望の信号を提示する形式の場合と、元音場から空間を切り出し、その内部の波面をスピーカアレイ等で合成する方法とに大別できる。本講演では、まず、VADの構成法について議論する。特に、ヘッドホンなどを用いた2チャンネル信号の構成に重点を置く。提示する音信号は、通常原音信号をレンダリングすることにより得られる。

一方、VADを高度化するには、音空間を人間がどのように知覚するかを理解する必要がある。特に、スピーカアレ

イを用いたVADでは、空間的折り返し歪みから逃れることができない。また、聴取者の身体動作に対応して、自然な提示を行うためには、頭部運動を伴う音空間の知覚への理解を深める必要がある。後半では、これら音空間知覚理解のために最近行った聴取実験の結果等について説明を加える。

2. 聴覚ディスプレイの構成法

聴覚ディスプレイを構成するには、2チャンネルの所望信号を両耳にどのように与えるかを考える必要がある。一番直裁な方法は、所望の鼓膜信号を用意して、ヘッドホン等で直接両耳に与える方法である(以後、両耳提示法と呼ぶ)。または、元音場の空間の一部の波面を、聴取空間にスピーカアレイを用いて合成する方法も考えられる(以後、空間合成法と呼ぶ)。本講演では、これら2つの方法のうち、特に前者にスポットを当てて研究事例を紹介する。

2.1 ヘッドホンを用いた両耳提示法

ヘッドホンを用いて音場を提示するには、両耳与えるべき所望の信号を得る必要がある。これを音源信号に対してレンダリングするためには、音源位置から両耳までの空間的な伝搬特性を考え、それをインパルス応答としてとらえ、音源に対して畳み込む必要がある。我々が普段生活する空間は、自由空間ではなく、反射・残響・回折などの物理現象が起こりうる。したがって、臨場感のある音場を提示す

^{†1} 東北学院大学工学部
Tohoku Gakuin University

るには、これらの影響も含めてインパルス応答をモデル化する必要がある。一般的にこのようなインパルス応答は、両耳室インパルス応答 (binaural room impulse response: BRIR) と呼ばれる。一方、これまでの VAD におけるレンダリングでは、自由空間における伝搬特性を考慮することが多かった。これらは、いわゆる頭部伝達関数 (head-related transfer function: HRTF) である。この時間的な表現としてのインパルス応答 (head-related Impulse Response: HRIR) のみを畳み込むということは、音源に対して、位置情報のみをレンダリングすることに等しい。HRTF は、個人性が強いので、理想的には個人化した HRTF を用いることが望まれ、この解決のための研究が進んでいる。もちろん BRIR には、HRTF の影響も含まれるため個人性がある。

ヘッドホンをを用いた音場の提示では、頭部運動に対応させて、畳み込むインパルス応答を切り替えることによって、物理世界と同様に一定の空間位置に音像を止めることが可能となる。これによって、提示された音場の自然性を高めることができる。これを頭部運動感応型 VAD という。

ここで、VAD の進歩について概観する。1980 年に Morimoto らが頭部伝達関数をオフラインで畳み込みした音源を用いた聴取実験を行った [2] のが空間レンダリングの始まりと言える。1990 年代になると、AD/DA を備えた DSP ボードが手に入るようになり、リアルタイムで音源に HRIR をたたみ込めるようになり、VAD も試作された。その後、2000 年代になると、PC の性能が向上し、native CPU で HRIR 程度の畳み込みが行えるようになり、畳み込みエンジンを windows 上でソフトウェアにより構成した VAD が実装され、視覚障害者用の音空間認識訓練システム等が実装された [3]。さらに、スマートフォンやタブレット端末などの携帯端末上でのシステムも可能となっており、可搬性の高いシステムに移行しつつある。これらのシステムでは、主に HRIR によるレンダリングが行われていたが、さらなる臨場感の提示には、BRIR のレンダリングが必要となる。つまり、長い (例えば、3 秒程度) インパルス応答の畳み込みがリアルタイムで実現されなければならない。

そこで、我々はクライアント-サーバ型の VAD システムの開発を行っている。初期のシステム [4] では、クライアントを携帯端末、高速 PC をサーバとして構成し、HTTP 通信により、1.5 秒程度の BRIR を 100 ms 以下の低遅延で畳み込み、頭部方向に応じて切り替えることに成功した (Fig. 1)。最近の研究では、切り替えの遅延をさらに短くするために、IoT の分野で近年注目されているエッジコンピューティングの概念を導入して開発を行っているすなわち、クライアント側では、初期のインパルス応答を畳み込み、サーバ側では、後部のインパルス応答を畳み込んで送信し、加算処理をクライアント側ですることにより、後部の計算結果に遅延があっても、切り替え自体は低遅延で行える。現状のシステムでは、3 秒程度のインパルス応答の畳み込み演算

と、インパルス応答を 30 ms 程度で切り替えられるシステムに成功している [5]。今後、HRTF の個人化とレンダリングの高精細化が進むことにより、VAD システムが手軽に利用できることが望まれる。

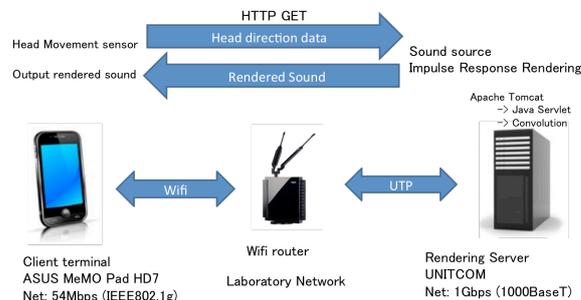


Fig. 1 ネットワーク型聴覚ディスプレイのアーキテクチャ

3. VAD と音空間知覚

3.1 空間合成法と空間的折り返し歪み

スピーカアレイによる空間合成法による聴覚ディスプレイでは、原則半波長以下にスピーカ間隔を狭める必要がある。可聴帯域全体をカバーするシステムを構築するには、8.5 mm 程度の間隔でスピーカを敷き詰める必要があり、したがって空間合成法において、現状では空間的折り返し歪みから逃れることができない。そこで、我々は、レンダリング空間で系統的にスピーカ間隔を操作し、聴取した音空間の印象が弁別できるかについて聴取実験を行った [6]。その結果、4 kHz 程度の周波数までに折り返し歪みがないようなシステムを構築すれば、聴取された音場の印象が、厳密再現された音場と弁別できないことが示唆された (Fig. 2)。

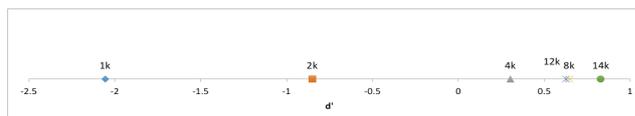


Fig. 2 Psychological distance among upper-limit conditions

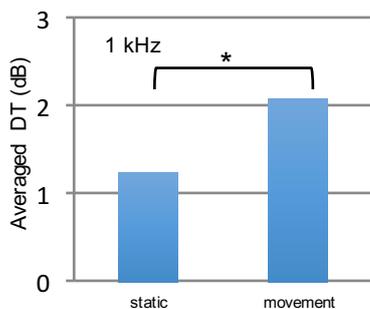
3.2 頭部運動中の音空間知覚

頭部運動を伴う音像定位では、定位精度が向上することが知られている [7]。一方、視覚では空間位置を注目する際に眼球が、小刻みにジャンプして視線を変化させる。これをサッカド運動と呼ぶが、サッカド運動中に入力される視覚刺激は抑制されることが知られている [8]。サッカド抑制が存在することで、外界の知覚を安定させていると言われている。視覚における抑制から類推すれば、聴覚においても、頭部運動時になんらかの抑制が行われていても不思議は無い。そこで、頭部運動を伴う音像定位を行わせ、その際の純音の検知限が変化するかどうかについて聴取実験を行った。その結果、頭頂方向から提示した 1 kHz 純音

の閾値が上昇することが示唆された[9]. VAD において頭部運動を伴う提示の場合、他の提示音の処理を抑えられる可能性がある。

4. おわりに

本講演では、聴覚ディスプレイシステムの構成を概観し、音空間知覚とシステム構成についての聴取実験について事例を紹介した。



閾値の変化(平均値) 1 kHz

Fig. 2 聴取者の閾値の変化の平均値比較

謝辞

本講演で紹介した研究結果の一部は、科学研究費補助金(基盤研究(B)26280078),および総務省 SCOPE (0159-0003)として推進された。

参考文献

- [1] J. Blauert, 森本政之, 後藤敏幸, “空間音響,” 鹿島出版会, 東京, 1986.
- [2] Morimoto Masayuki, Ando Yoichi, “On the simulation of sound localization,” J. of Acoust. Soc. of Jpn. (E), Vol. 1 (3), pp. 167-174 (1980).
- [3] 大内誠, 岩谷幸雄, 鈴木陽一, “視覚障害者のための3次元聴覚情報の提示,” バイオメカニズム学会誌, Vol. 31, No.2, pp. 95-100 (2007).
- [4] Yukio Iwaya, Makoto Otani, Takao Tsuchiya, and Junfeng Li, “Virtual Auditory Display by Remote Rendering via Computer Network,” Proc. of IHH-MSP 2014.
- [5] 伊藤修平, 岩谷幸雄, 大谷真, 土屋隆生, “エッジコンピューティングを用いたネットワーク型聴覚ディスプレイの検討,” 信学技報, EA2016-121 (2017).
- [6] Yukio Iwaya, Makoto Otani, Takao Tsuchiya, and Yasushi Inoguchi, “Perception of acoustical spatial attributes and impression in virtually rendered sound field,” Proc. of ICSV22 (2015).
- [7] S. Perrett and W. Noble, “The contribution of head motion cues to localization of low-pass noise,” Percept. Psychophys., 59 (7), 1018-1026, 1997.
- [8] 光道宏行, “サッカーボールの眼球運動と知覚的安定,” Kyushu University Psychological Research, Vol. 12, 61-68, 2011.