

# 包囲形スピーカアレイを用いた リアルタイム室内音場可聴化システムの実装

Implementation of real-time room auralization  
using a surrounding loudspeaker array

岡本拓磨<sup>1,2)</sup>, Katz Brian F.G.<sup>4)</sup>, Noisternig Markus<sup>5)</sup>, 岩谷幸雄<sup>1,3)</sup>, 鈴木陽一<sup>1,3)</sup>

Takuma OKAMOTO, Brian F.G. KATZ, Markus NOISTERNIG, Yukio IWAYA and Yōiti SUZUKI

- 1) 東北大学電気通信研究所, 2) 東北大学大学院工学研究科, 3) 東北大学大学院情報科学研究科  
(〒980-8577 仙台市青葉区片平 2-1-1, {okamoto@ais, iwaya@, yoh@} riec.tohoku.ac.jp)  
4) LIMSI-CNRS (BP 133, 91403 Orsay Codex, France, brian.katz@limsi.fr)  
5) IRCAM (1, place Igor-Stravinsky, 75004 Paris, France, Markus.Noisternig@ircam.fr)

**Abstract :** Real-time auralization systems have been developed for room acoustic simulations. For high quality spatial rendering, many of these systems have used binaural rendering. This paper presents the implementation of a real-time room acoustics auralization system using an array of 157 loudspeakers. The room acoustic model uses a combination of an iterative image-source model and feedback delay networks to create the reflections and reverberation. Higher-order Ambisonics (HOA) is used to generate spatial room impulse responses. The speaker array uses a density grid of 50 cm and covers the walls and ceiling of a treated listening room. A distributed network system has been used to create the auralization. All software components of this system are Open-Source. In this paper, we introduce this system and discuss about the synchronous estimation of 157 audio signals.

**Key Words:** *Surrounding loudspeaker array, Auralization, Higher-order Ambisonics*

## 1. はじめに

コンサートホールなどの音空間モデルを作成し、音場を計算機シミュレーションによって推定した音を人間に提示する技術を可聴化技術 (Auralization)[1] という。この技術は、実際の空間の音場の計算だけでなく、計算機上で仮想的に作り出された音空間を生成できるため、バーチャルリアリティ技術としても非常に注目を集めている。

これまでの可聴化技術の多くは、空間の一点において計算された音を提示する技術に留まっていたため、人の動きには対応することができなかった。それに対して、人の動きにも対応できるとする室内音場シミュレーションのためのリアルタイム可聴化技術の発展が期待されている。

また、これまでの高品質空間レンダリング技術の多くは、再生系が単純で安価であるという理由から、ヘッドホンを用いたバイノーラル再生方式を採用してきたが、M. Noisternig *et al.* は、ヘッドホンを用いたバイノーラル再生によるリアルタイム可聴化技術を提案している [2]。

我々は、この技術を発展させ、多チャンネルスピーカによる再生系に拡張したリアルタイム室内音場可聴化の実装を試

みる。再生系には、室内の床と天井の計 5 面に格子状に計 157 ch のスピーカを配置した包囲型スピーカアレイ [3, 4] を用いる。本システムでは、聴取者の位置情報を位置センサにより取得し、仮想音源と聴取者間の反射音を虚像法により生成し、各スピーカから提示するインパルス応答を高次アンビソニックの原理 (HOA: Higher-order Ambisonics)[5] を用いて生成する。我々は、反射音の伝搬経路をリアルタイムで可視化する処理と各スピーカから提示する 157 ch の音情報を生成する処理とを、オープンソースによる 5 台の PC による計算とそれらの OSC (Open Sound Control)[6] プロトコルおよび UDP (User Datagram Protocol) 通信により実装した。本報告では、構築したシステムの概要およびシステムの基礎的な評価について述べる。

## 2. 包囲型スピーカアレイの構築

多チャンネルスピーカによる再生系として、床以外の壁面 4 面と天井の計 5 面にスピーカを格子状に 50 cm 間隔で配置した包囲型スピーカアレイを構築した。スピーカと各壁面、天井との距離は 30 cm である。スピーカ配置および室内寸

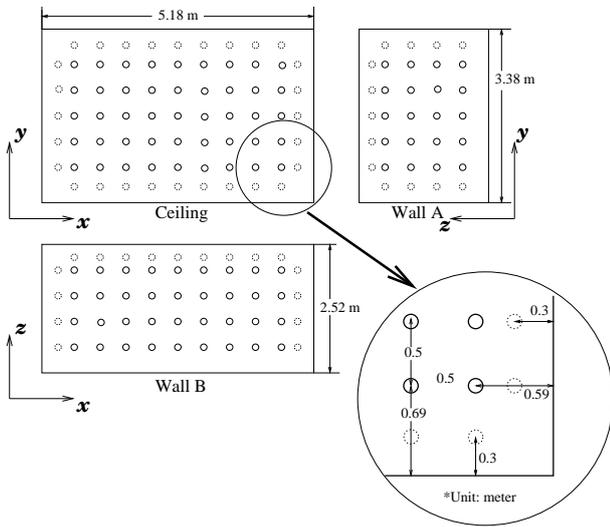


図 1: 包囲型スピーカアレイのスピーカ配置

法を図 1 に示す．部屋の残響時間は 0.20 s であった．157 ch の同時再生を実現するために，我々は合計 14 台の D/A を 4 台の PC に分割して制御することによって実装した．全ての D/A のクロックをそろえるために，クロックジェネレータ (Rosendahl Nanosyncs HD) を用いた．

天井には 6 自由度位置センサを設置しているため，聴取者の位置情報をリアルタイムに取得可能である [4]．

### 3. 構築したシステムの概要

このシステムを実現するためには，(i) コンサートホールなどの音空間をモデル化する技術，(ii) 音源位置と聴取位置から聴取者位置での室内インパルス応答をリアルタイムに計算する技術，(iii) 計算した室内インパルスを多チャンネル再生系に変換する技術，(iv) 多チャンネル音データを同期させて再生する技術が必要となる．

(i) については，室内音場ソフトウェアパッケージ CATT-Acoustic v8.0[7] を用いて建築物のモデルを作成した．文献 [2] 同様，単純な直方体から複雑なホールの形状データを作成できる．以下，(ii) ~ (iv) を実現するために，図 2 に示すシステムを構築した．また，文献 [2] 同様，(v) 計算した室内の反射音の伝搬経路をリアルタイムで可視化する技術についても実装した．

#### 3.1 初期反射経路の算出

反射音のシミュレーションには，音の反射を鏡面反射として計算する虚像法 [8] を用いる．システムへの実装は，虚像法演算ライブラリ EVERT[9] に基づくオープンソースリアルタイムソフトウェアである EVERTims により行った．これにより音源位置や聴取者位置に応じて反射音の伝搬経路をリアルタイムに算出できるようになった．

#### 3.2 初期反射経路の可視化

EVERTims により得られた反射音の伝搬経路を 3 次元可視化オープンソフトウェアである VirChor (VirtualChoreographer)[10] により可視化する．EVERTims と VirChor の処理は 1 台の PC (Linux) により実装した．図 3 に反射音が

存在しない状態，1 次反射までおよび 3 次反射までの可視化を示す．手前の球が音源で，奥の顔の形が聴取者であり，音源から壁面を反射して聴取者まで届く光線が EVERTims によって算出された直接音と反射音の伝搬経路を表す．音源や聴取者の位置を変化させると，EVERTims により反射音の伝搬経路が更新され，それに伴い VirChor 上の伝搬経路も更新される．

#### 3.3 聴取者位置における室内インパルスの生成

レンダリング音の生成と提示は，包囲型スピーカアレイの制御 PC (Mac OS X) 4 台とリアルタイム多チャンネルオーディオソフトウェアである Pd (Pure data)[11] により実装した．Pd は Windows, Mac OS, Linux 等に対応したクロスプラットフォーム仕様のソフトウェアである．また，OSC プロトコルや UDP 通信によりマシン間でデータや命令を相互通信も可能である．本システムにおいて，Pd は EVERTims をインストールしたメイン PC，およびオーディオ制御用 PC 4 台のみにインストールされた．

音源位置情報，聴取者位置情報，EVERTims によって算出された虚像音源情報がメイン PC から Pd の OSC プロトコルによって 4 台の Mac に逐次転送される．ここで，虚像法は反射の次数が上がるほど計算に時間がかかることや，後部残響は初期反射と比べてパワーの減衰が大きいいため，本システムでは 3 次までの反射音を EVERTims により高速算出し転送する．

各マシンは受信した音源と聴取者の位置情報，3 次までの反射音情報から聴取者位置での室内インパルス応答を算出する．3 次以降の後部残響については，3 次反射までのパワー減衰情報をもとに FDN (feedback delay networks)[12] により近似的に算出する．

#### 3.4 高次アンビソニック (HOA) に基づく 3 次元音空間レンダリング

音源と聴取者の位置情報，EVERTims によって算出した初期反射情報，FDN により算出した後部残響情報から，聴取者の位置における 3 次元音空間レンダリングを行う．本システムでは，HOA の原理に基づく方法による実装を行う．HOA では，音空間を再現する空間の中心から距離  $r$ ，水平角  $\theta$ ，仰角  $\phi$  の位置  $\mathbf{r} = (r, \theta, \phi)$  における音圧  $p$  を球面展開係数  $A_{mn}$  にエンコードし，再生系でデコードすることによって音空間を再現するアルゴリズムである．位置  $\mathbf{r}$  での音圧  $p$  を球面調和解析すると，球面展開係数  $A_{mn}$  とは以下の関係が成り立つ．

$$p(\mathbf{r}, \omega) = \sum_{m=0}^{\infty} i^m j_m(kr) \sum_{0 \leq n \leq m, \sigma = \pm 1} A_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta, \phi) \quad (1)$$

ここで， $\omega$  は角周波数， $Y_{mn}^{\sigma}$  はオーダー  $n$ ，次元  $m$  の球面調和関数の第一次と二次の球面ベッセル関数であり，次式で与えられる．

$$Y_{mn}^{\sigma}(\theta, \phi) = \sqrt{(2m+1)(2-\delta_n) \frac{(m-1)!}{(m+n)!}} P_{mn}(\sin\theta) \begin{cases} \cos(n\theta) & \text{if } \sigma = +1 \\ \sin(n\theta) & \text{if } \sigma = -1 \end{cases} \quad (2)$$

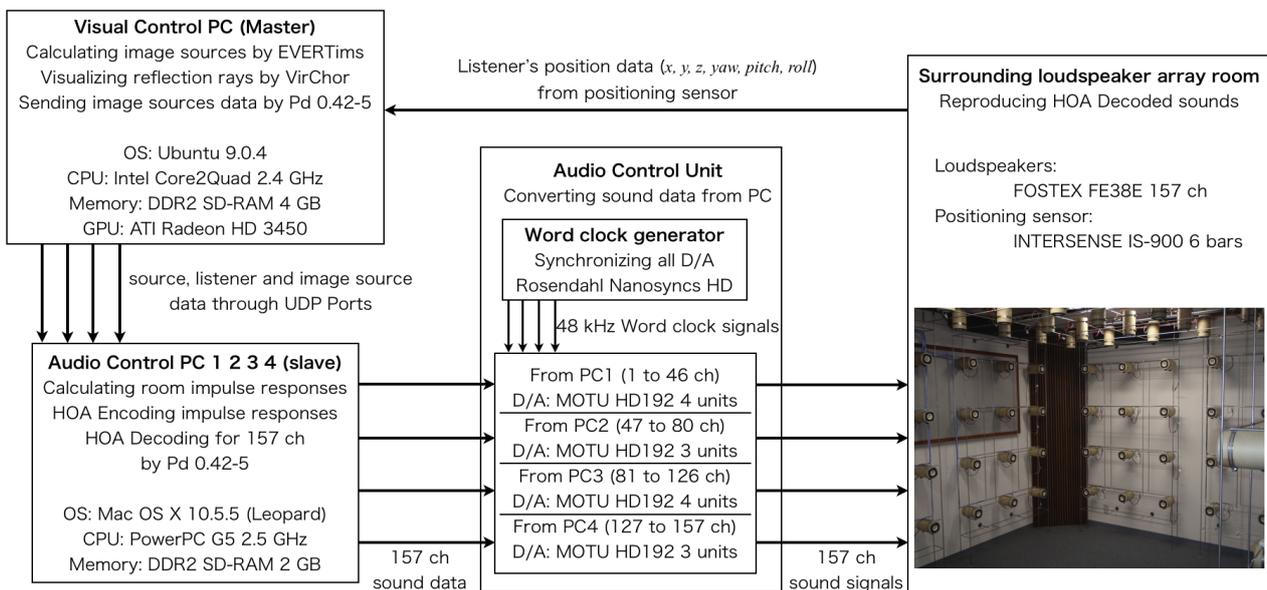


図 2: 157 ch リアルタイム可聴化システム

ここで、 $P_{mn}(\sin\theta)$  はオーダー  $n$ 、次元  $m$  のルジャンドル陪関数、 $\delta_n$  はクロネッカーのデルタである。

HOA は、一度 3 次元空間を  $A_{mn}$  としてエンコードしてしまえば、デコードの際にはどのような再生系にも適用可能であり、再生系に依存しないことが利点である。また、エンコードする次数を大きくすることによってより忠実な音空間を再現できることも挙げられる [2]。

本システムでは、音源と聴取者の位置情報から直接音による  $p(r, \omega)$  を、EVERTims によって算出した初期反射情報から得られる初期反射音による  $p(r, \omega)$  をそれぞれ球面展開係数  $A_{mn}$  に 4 次のアンビソニックエンコードを行う。FDN により算出した後部残響情報に関しては適当な方向情報を与え、1 次のアンビソニックエンコードを行う。現在は、システムの限界では 3 次までの初期反射音を合計 100 個まで HOA エンコードが可能である。

最後に、エンコードされた球面展開係数  $A_{mn}$  を包囲型スピーカの位置情報に合わせてデコードする。デコードするスポットは部屋の中心 ( $x = 2.59, y = 1.69, z = 1.26$ ) である。4 台のマシンで対応するスピーカに応じたインパルス応答を HOA デコードにより算出し、音源に畳み込み、最後に 157 ch 同時再生することによってスポット位置でのシミュレーションで得られた 3 次元音空間のリアルタイム可聴化を実現する。現在、音源には既収録された音データや、4 台目の PC にマイクロホンから入力される音信号を、算出したインパルス応答に畳み込むことも可能となっている。なお、4 台目の PC に入力された音信号は OSC プロトコルによって 1~3 台目の PC に転送される。

#### 4. 性能評価

文献 [2] でのシステムにおいても EVERTims による反射音を計算しているため、計算時間については同等である。しかし、我々のシステムでは 157 ch を 4 台のマシンに配分し、

14 台の D/A (12 ch/台) を用いているため、各 D/A 間やマシン間での同期を確認する必要がある。そこでまずは、室内インパルス応答の計算や HOA エンコード・デコード処理などはまったく行わない状態での Pd 制御による時間同期精度についての基礎的検討を行う。

##### 4.1 TSP 信号を用いた時間同期確認

Pd では、OSC プロトコルや UDP 通信を用いて各マシン間でのデータや命令の転送を行うことができる。今回は、PC1 で音を出す命令を出し、その命令を UDP で PC2~4 に転送し、全マシンの全チャンネルが同時に音を出力するプログラムを作成し、その時間同期についての検討を行った。

出力する音はサンプリング周波数 48 kHz、16 bit 量子化のリニア PCM の TSP (Time Stretched pulse) 信号 [13] を用いる。出力された音をオーディオインターフェースである M-Audio ProFire Lightbridge、A/D である Behringer ADA8000、収録ソフトウェアである Desidesign Pro Tools 8 M-Powered を用いてサンプリング周波数 48 kHz、16 bit リニア PCM の 8 ch 同時収録を行う。出力が 157 ch のため、出力位置を様々なパターンで接続しての収録を行った。

TSP 信号はその性質上、自己相関を計算すると、完全に時間が一致した場合のみ大きな値をとり、1 サンプルでもずれると小さな値となる。その性質を利用し、収録データと音源である TSP 信号の相互相関を計算することにより時間同期の確認を行った。その結果、各 PC に制御された D/A はどの PC 内でも全チャンネル完全に同期して再生されていることがわかった。しかし、異なった PC 間での同期はとれておらず、どのマシン間でも最大約  $\pm 300$  タップ以内 (= 約 6 ms) の時間ずれが発生していることがわかった。これは、PC1 で発令した命令を UDP で PC2~4 で転送する際の転送速度が毎回マシンごとに異なっていることが原因である。157 ch を完全に同期再生するためには、UDP 通信速度の固定化や OSC プロトコルの改良などが必要であると考

えられる。

## 5. おわりに

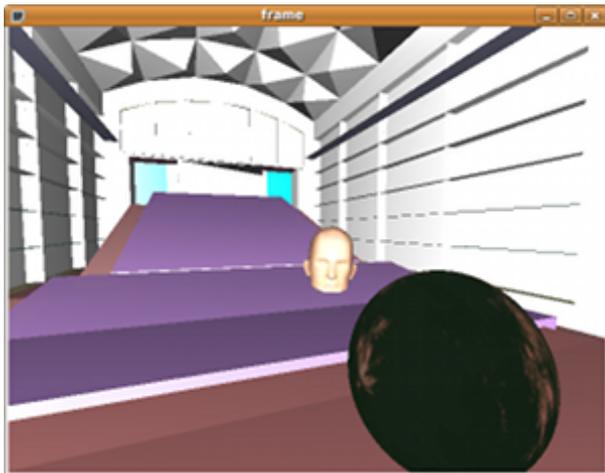
本報告では、157 ch の包囲リアルタイム室内音場可聴化システムの実装についてのシステム紹介、および基礎的な時間同期についての性能評価を行った。

今後は、異なったマシンで制御される信号を同期させて全チャンネルを同期して提示するためのシステムの改良を行う。その上で、システム全体の遅延や再現精度に関する性能評価を行う。

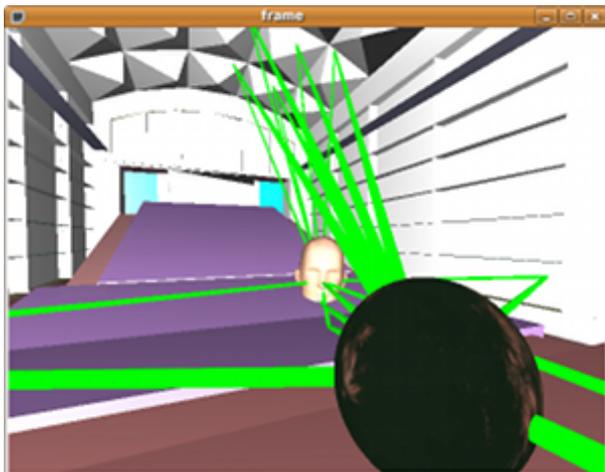
謝辞 本研究は東北大学大学院工学研究科電気情報系 GCOE プログラム (CERIES) の助成による。

## 参考文献

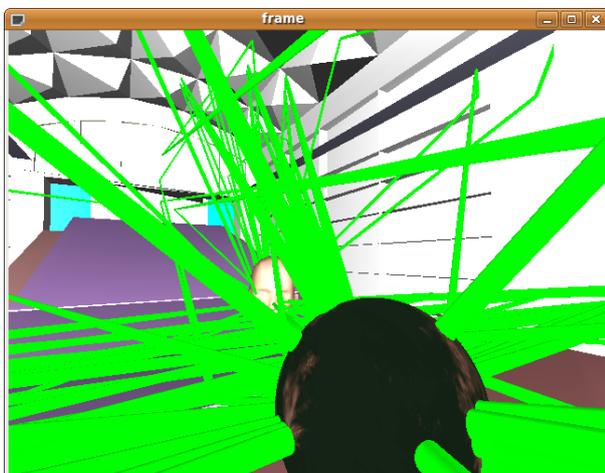
- [1] M. Kleiner, B. I. Dalenback and P. Svensson, "Auralization: an overview," *J. Audio Eng. Soc.*, vol. 41, no. 11, pp. 861-875, Nov. 1993.
- [2] M. Noisterning, B. F. G. Katz, S. Siltanen and L. Savioja, "Framework for real-time auralization in architectural acoustics," *Acta Acustica United with Acustica*, vol. 94, no. 6, pp. 1000-1015, Nov. 2008.
- [3] T. Okamoto, R. Nishimura and Y. Iwaya, "Estimation of sound source positions using a surrounding microphone array," *Acoust. Sci & Tech.*, vol. 28, no. 3, pp. 181-189, May 2007.
- [4] D.-G. Kang, Y. Iwaya, R. Miyauchi and Y. Suzuki, "Comparison of sound localization performance between visual and real three-dimensional immersive sound field," *Acoust. Sci & Tech.*, vol. 30, no. 3, pp. 216-219, May 2009.
- [5] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *J. Audio Eng. Soc.*, vol. 53, no. 11, pp. 1004-1025, Nov. 2005.
- [6] <http://opensoundcontrol.org/>
- [7] <http://www.catt.se/>
- [8] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943-950, Apr. 1979.
- [9] S. Laine, S. Siltanen, T. Lokki, and L. Savioja, "Accelerated beam tracing algorithm," *Applied Acoustics*, vol. 70, no. 1, pp. 172-181, Jan. 2009.
- [10] <http://virchor.wiki.sourceforge.net/>
- [11] <http://puredata.info/>
- [12] J.-M. Jot and A. Chaigne, "Digital delay networks for designing artificial reverberators," *Proc. AES 90th Int. Convention*, 3030, Feb. 1991.
- [13] Y. Suzuki, F. Asano, H. Y. Kim and T. Sone, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," *J. Acoust. Soc. Am.*, vol. 97, no. 2, pp. 1119-1123, Feb. 1995.



(a) Non reflection environment



(b) 1st order reflection in the 3D visualization unit



(c) 3rd order reflection in the 3D visualization unit

図 3: VirChor による反射音経路のリアルタイム可視化