

# 音像空間配置のインタラクティブな制御手法

大谷 健登<sup>1,a)</sup> 鈴木 友美<sup>1</sup> 丹羽 健太<sup>2</sup> 西野 隆典<sup>3</sup> 石黒 祥生<sup>1</sup> 武田 一哉<sup>1</sup>

**概要:** 音像空間配置の操作は聴覚印象を操作する一つの手法として利用でき、受聴者と音像の位置関係や距離が変化することにより音の聴覚印象が変化する。しかし、多数の音像が存在する場合、各音像について個別に操作を行う必要があり、最適な音像空間配置を選択することが困難である。そこで、本研究では音像空間を一次元で表現する手法について提案する。提案手法では聴覚印象特徴量として頭部伝達関数を利用し、事前に聴覚印象特徴量が異なる配置を複数選択する。その後、それらの配置を連続的に結ぶことで、音像空間配置を一次元の軸で操作する事が可能とする。また、楽曲信号等に対する提案手法の応用方法についても議論する。

## 1. はじめに

音楽プレーヤやデジタル音源の普及により、音信号を手軽に加工することが出来るようになり、適切に信号を加工することで、ユーザは各自の好みの聴覚印象を得ることが出来る。例えば、周波数イコライザ [1], [2], [3], [4], [5] は多くの音楽プレーヤに標準で搭載されている機能であり、楽曲全体の周波数特性を操作することで聴覚印象の操作を行うことができる。また、音源信号ごとに個別の信号が存在する場合には、MPEG-SAOC [6] に標準化されているように、各音源信号ごとに音像空間配置を制御することができ、それにより聴覚印象を操作することが可能である。本研究では、音像空間配置を制御することによる聴覚印象操作に着目し、ユーザが容易に聴覚印象を操作することのできる手法を提案する。

音像空間配置を制御するための研究は、受聴位置選択型 (Selective Listening Point: SLP) オーディオ [7]、インタラクティブ音像定位コントローラ [8] など従来から行われている。また、楽器音ごとに音量の制御を行う楽器音イコライザ [9] も、音量が距離のみに依存すると考えることで、音像空間配置制御手法の一つと言える。しかしながら、従来手法では音像空間配置や音量の制御のために多数の操作

軸が必要となってしまう、ユーザが好みの聴覚印象を得ることが困難である。

例として、SLP オーディオ [7] では楽曲中に含まれる各楽器音に対して個別に音像位置を制御することができる。実際には、個々の音源信号  $S_n(\omega)$  及びその音源から受聴位置までの立体音響特性  $H_n(\omega)$  が事前に手元にあることを仮定して、次の式のように出力信号  $y(\omega)$  を得る。

$$y(\omega) = \sum_{n=1}^N H_n(\omega) S_n(\omega) \quad (1)$$

ここで  $N$  は音源数を表す。バイノーラル再生を利用する場合には、 $H_n(\omega)$  に頭部伝達関数 [10] (Head Related Transfer Function: HRTF) を利用することとなり、音像空間配置が変化することで  $H_n(\omega)$  が変化するため、それに伴い出力信号も変化する。SLP オーディオでは、出力信号の聴覚印象も音像空間配置とともに変化するため、ユーザは好みの聴覚印象を持つ音像空間配置を見つけることが重要となる。しかしながら、楽曲には多数の音源が含まれている場合が多く、その場合ユーザが制御しなければならない操作軸数の増加が課題となる。楽曲中に含まれる音源数を  $N$ 、各楽器音に対して 2 次元空間上の  $M$  か所から一か所を選ぶとすると、最適な音像空間配置を決定するため、ユーザは  $M^N$  通りの候補の中からある 1 つの音像空間配置を選択することとなり、各音源を 2 次元平面上で操作するために  $2 \times N$  個の操作軸を同時に操作する必要がある。例えば、楽曲中に 3 音源しか存在せず、音像を配置する空間を  $10 \times 10$  の格子状に区切った場合でも、ユーザは 6 本の操作軸を同時に操作しながら、100 万通りの音源配置の中から最適なものを選択する必要がある。これではユーザは最適な音像空間配置を容易に決定することができず、シ

<sup>1</sup> 名古屋大学  
Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, Aichi 464-8603, Japan

<sup>2</sup> NTT メディアインテリジェンス研究所  
NTT Media Intelligence Laboratories, 3-9-11 Midori-cho, Musashino-shi, Tokyo 180-0012, Japan

<sup>3</sup> 三重大学  
Mie University, 1577 Kurimamachiya-cho, Tsu-shi, Mie 514-8507 Japan

a) ohtani.kento@g.sp.m.is.nagoya-u.ac.jp

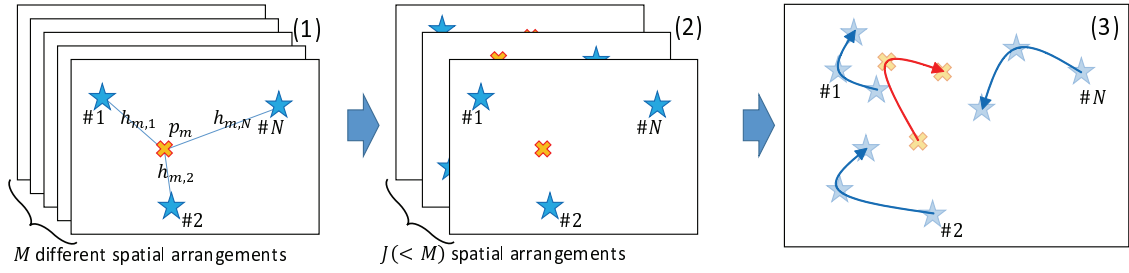


図1 システムの概要. (1) 取り得るすべての音像空間配置について HRTF を利用して聴覚印象特徴量を計算する. (2) 聴覚印象特徴量が異なる  $J$  種類の音像空間配置を取得する. (3) 選択された音像空間配置を滑らかに結ぶ.

システムを利用して最適な聴覚印象を持つ空間音像配置を探すことは難しい.

そこで簡便な音像空間配置の制御のため, 本研究では一次元の軸に沿って音像空間配置を大きく変化させる手法を提案する. 音源ごとの信号は事前に取得されており, 音像の空間印象はバイノーラル再生によりヘッドホンを利用して呈示されると仮定する. バイノーラル再生による音像呈示は HRTF を通して行われるため, 頭部伝達関数を利用することで特定の音像空間配置における聴覚印象特徴量を計算できる. 聴覚印象特徴量は, その音像空間配置でのユーザが受ける音信号の特性を表しており, 類似した聴覚印象特徴量を持つ場合にはユーザは類似した空間印象であると感じることとなる. それらの特徴量が大きく変化するように, 一次元の軸を構成することで, 単一の操作軸のみで音像空間配置を大きく変化させることができるようになる. その結果, ユーザは音を聴きながらインタラクティブに音像空間配置を制御することが出来るようになり, 容易に好みの聴覚印象を持つ音像空間配置を得ることが出来る.

## 2. 提案法

### 2.1 提案法の概要

本節では一次元の軸に沿って音像空間配置を制御する手法について述べる. 前節で述べたように, 選択される音像空間配置は非常に数多く存在しているが, HRTF は音源位置と受聴位置の相対関係のみによって決定されるため, 類似した聴覚印象を持つ音像空間配置も数多く存在しているはずである. そこで, 選択されるすべての聴覚印象の異なる音像空間配置を選択することで, 一次元の軸に沿って音像空間配置の制御が可能となる.

図1に提案法の流れを示す. 本手法は3ステップで構成される.

- (1) 各音像空間配置の聴覚印象特徴量を計算する.
- (2) 聴覚印象特徴量が異なるいくつかの音像空間配置を選択する.
- (3) 選択された空間音像配置を並び替え, 滑らかに結ぶ.

### 2.2 HRTF による聴覚印象特徴量

本研究では本研究では, 聴覚印象特徴量を計算するために HRTF を利用する. 空間中に音源が  $N$  個存在し, それらのパワーがほぼ等しいと仮定すると,  $j$  番目の音像空間配置の聴覚印象特徴量  $\mathbf{f}_j$  は以下の式で表される.

$$\mathbf{f}_j = [\mathbf{H}_{j,1}^{(L)\top}, \mathbf{H}_{j,1}^{(R)\top}, \dots, \mathbf{H}_{j,N}^{(L)\top}, \mathbf{H}_{j,N}^{(R)\top}]^\top, \quad (2)$$

ただし,  $j$  は音像空間配置番号であり,  $\mathbf{H}_{j,n}^{(L)}, \mathbf{H}_{j,n}^{(R)}$  は以下で表される.

$$\begin{aligned} \mathbf{H}_{j,n}^{(L)} &= [H_{j,n}^{(L)2}(1), \dots, H_{j,n}^{(L)2}(\Omega)]^\top, \\ \mathbf{H}_{j,n}^{(R)} &= [H_{j,n}^{(R)2}(1), \dots, H_{j,n}^{(R)2}(\Omega)]^\top, \end{aligned}$$

ここで,  $\Omega$  は周波数ビン数,  $H_{j,n}^{(L)2}(\omega), H_{j,n}^{(R)2}(\omega)$  は  $j$  番目の音像空間配置における  $n$  番目の音源から受聴者までの左・右耳の HRTF である. また,  $i$  番目の音像空間配置と  $j$  番目の音像空間配置の聴覚印象特徴量の類似度を次のように計算できる.

$$k(i; j) = \mathbf{f}_i \cdot \mathbf{f}_j, \quad (3)$$

ただし,  $\cdot$  は内積演算を表す.

### 2.3 聴覚印象特徴の異なる音像空間配置の選択

$J$  種類の異なる聴覚印象を持つ音像空間配置を取得する手法について提案する.  $J$  種類の音像空間配置間の聴覚印象特徴の類似度は次のような行列で計算することが出来る.

$$\mathbf{K} = \mathbf{F}^\top \mathbf{F} = \begin{bmatrix} k(1;1) & \dots & k(1;J) \\ \vdots & \ddots & \vdots \\ k(J;1) & \dots & k(J;J) \end{bmatrix} \quad (4)$$

$$\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_J]. \quad (5)$$

選択された  $J$  種類の音像空間配置の類似度が低い場合には,  $\mathbf{K}$  の非対角成分の値が小さくなる. そこで行列  $\mathbf{K}$  の対角度  $c$  を以下のように計算する.

$$c = \frac{\sum_{i=j} \mathbf{K}_{i,j}}{\sum_{i=j} \mathbf{K}_{i,j} + \sum_{i \neq j} \mathbf{K}_{i,j}}. \quad (6)$$

$c = 1$  の時には行列  $\mathbf{K}$  は対角行列であり、 $c = 0$  の時には行列  $\mathbf{K}$  の対角成分はすべて 0 である。

実際には以下のようなアルゴリズムにより  $J$  種類の聴覚印象の異なる空間音像配置を取得する。

- (1) 存在する  $M$  種類の音像空間配置のうちの 2 種類を選択する全組み合わせに対し、 $2 \times 2$  行列  $\mathbf{K}$  の対角度  $c$  を計算する。
- (2)  $M C_2$  組の中から  $c$  を最大化する 2 種類の 2 音像空間配置の組み合わせを取得する。
- (3) 全ステップの 2 種類の音像空間配置に別の音像空間配置を加え、 $3 \times 3$  行列  $\mathbf{K}$  に対して対角度  $c$  を計算する。 $M - 2$  種類の音像空間配置について対角度の計算を繰り返し、対角度を最大化する音像空間配置を新たに一つ加える。
- (4)  $J$  種類の音像空間配置を得るまでステップ (3) を繰り返す。対角度  $c$  は選択された音像空間配置数に対して単調減少関数であるため、対角度  $c$  の減少幅が小さくなる音像空間配置数を  $J$  とした。つまり対角度を選択された音像空間配置数の関数  $c(j)$  とした時とした時、 $J$  は 2 階微分  $\frac{d^2}{dj^2} c(j)$  を最大化する。

なお、本研究では受聴者の位置は固定とし、音源はすべて受聴者の前方に存在すると仮定する。

## 2.4 音像空間配置の一次元制御のための結合手法

$J$  種類の音像空間配置を結ぶ前に、以下の距離尺度を利用して選択された音像空間配置を並べ替える。

$$d(i; j) = \sum_{n=1}^N \|\mathbf{p}_{i,n} - \mathbf{p}_{j,n}\| \quad (7)$$

ただし、 $\|\cdot\|$  は L2-ノルムを表し、 $\mathbf{p}_{j,n}$  は  $j$  番目の音像空間配置における音源  $n$  の座標を表す。

音像空間配置の並べ替えには以下のアルゴリズムを利用した。

- (1) ランダムに一つ音像空間配置を選択する。
- (2) 選択された音像空間配置と残りの音像空間配置間の距離を式 (7) により計算し、距離が最も小さくなる配置を選択する。
- (3)  $J$  種類の音像空間配置に対しステップ (2) を繰り返す。上のアルゴリズムに沿って並び替えた後、3 次スプライン補間によって滑らかに結ぶ。

## 3. 実験

### 3.1 実験条件

提案法を利用して、音像空間配置の一次元制御経路を取得した。音源数は 3 とし、音源は  $9 \times 9$  の格子状に配置する。 $x$  軸、 $y$  軸の範囲はそれぞれ  $-3 \sim 3$  m、 $0 \sim 6$  m とした。聴覚印象特徴量は名古屋大学 武田研究室にて公開している B&K 4128 で収録されたデータ [11] を利用した。その他の

表 1 実験条件

|             |                                   |
|-------------|-----------------------------------|
| 標準化周波数      | 48 kHz                            |
| HRTF データベース | 名古屋大学武田研究室 [11]<br>B&K 4128 収録データ |
| HRTF 間隔     | $5^\circ$                         |
| HRTF 長      | 10.7 ms (512 pt)                  |
| 音源数         | 3                                 |

実験条件は表 1 に示す。

### 3.2 選択された音像空間配置

選択された音像空間配置と各音源の移動経路を図 2 に示す。今回は提案法により 10 種類の音像空間配置が取得された。黒丸は受聴者位置を示しており、 $(0,0)$  に固定されている。それ以外の色付きの印は各音源の位置を示し、実線に沿って白抜き印へと移動する。

この図から、提案法により様々な音像空間配置が取得されていることがわかる。例えば、音像空間配置 (1) では音源 2 が受聴者に近く、(2) では音源 3、(3) では音源 1 が受聴者に近づいている。また、各音源が受聴空間全体を動き回っていることも見て取ることができ、一次元の軸に沿って様々な音像空間配置を得ることが出来ていることが確認できる。

## 4. 提案法の応用例

### 4.1 音楽プレーヤへの応用

本研究の応用先として、音楽プレーヤが挙げられる。現在でも、ほとんどの音楽プレーヤには周波数イコライザなどの聴覚印象を操作するための機能が付随しており、ユーザは各自の好みの聴覚印象を得るために調整を行うことが出来る。しかしながら、現在利用されている多くの機能は楽曲全体に作用してしまうため、個別の楽器の音量の制御や楽器の音像空間配置などを制御することはできない。楽器の音像空間配置を制御可能となることで、従来よりも柔軟な聴覚印象の制御を行うことが出来るようになると考えられるが、単純に各楽器音の位置を制御することが出来るだけでは、非常に多くの操作軸を利用した複雑な操作が必要となってしまう。そこで提案法を利用することで、図 3 に示すように単一の操作軸を利用するだけで複数の楽器音の音像空間配置を制御可能となり、利用者の望む音像空間配置を容易に実現できると考えられる。

### 4.2 拡張聴覚システム

提案法は、音楽プレーヤのみでなく現実の音空間へも応用可能である。例えば、周囲の音をマイクロホンアレイで取得し、方向ごとに音信号を分離した後、分離された信号に対して提案法を適用するシステムを考える。このシステムを利用することで、ユーザは周囲が騒がしい中で特定の方向の音を強調・減衰させたり、音の到来方向を自由に変

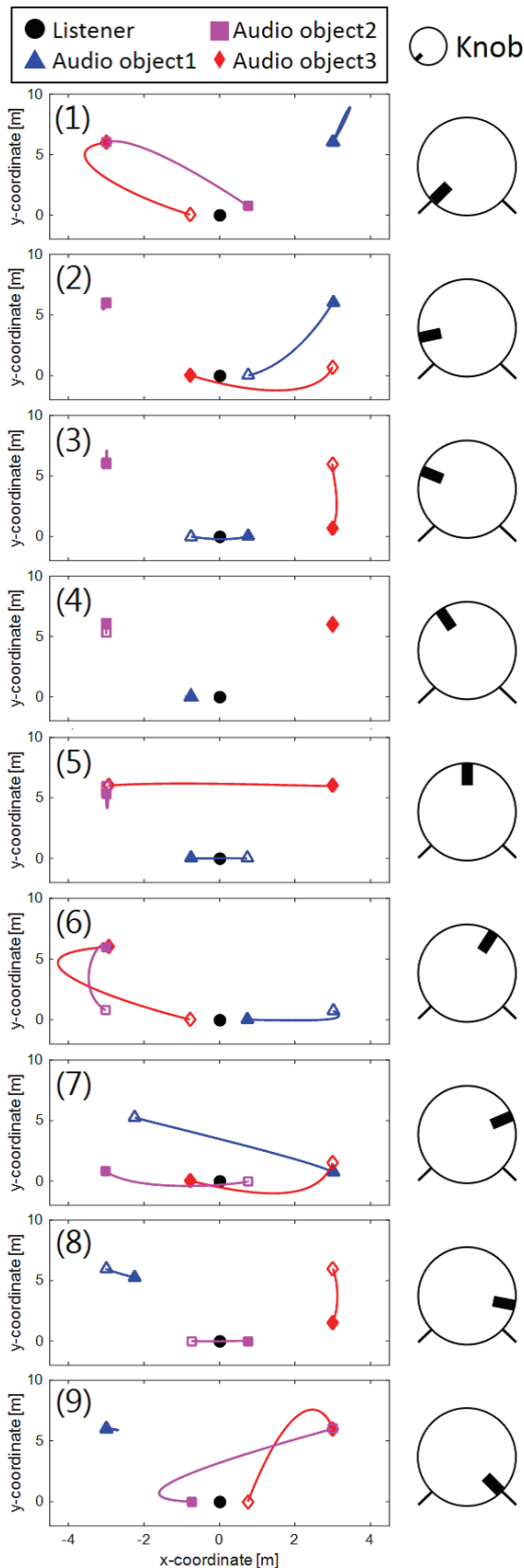


図 2 選択された音像空間配置. 黒丸が受聴者位置を表す. それ以外の印は音源を表し, 実線が音源の移動経路を表す. 各音源は白抜き印に向かって移動する. ユーザが単一の軸を操作することで, 全音源が同時に図の順に移動する.

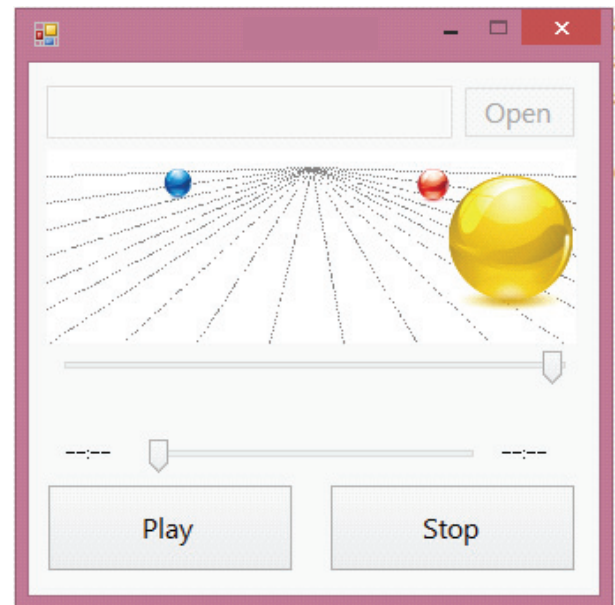


図 3 提案法の音楽プレーヤーへの応用例. 描画領域内の球が楽器音位置を表し, 球の大きさが音量を表す. ユーザは描画領域下部にある一次元のスライダーを操作するだけで楽器の音像空間配置を制御することができる.

更させたりといったようなことが出来る. このシステムは聴覚に対する拡張現実であるため拡張聴覚システムと呼ぶことが出来る.

## 5. まとめ

本研究では, 音像空間配置を一次元の操作軸によって制御する手法を提案した. 本手法を利用することで, ユーザは複数の音源について個別に音像空間配置を制御する必要がなくなり, 一次元の操作軸によって容易に音像空間配置を制御することが可能となる. 提案法は, 3ステップで構成されている. 初めに, 考え得るすべての音像空間配置に対し, HRTF を利用して聴覚印象特徴量を計算する. 次に, 聴覚印象特徴量が異なる音像空間配置を複数選択し, その後選択された音像空間配置を滑らかに結ぶ. 結果として, 様々な音像空間配置を滑らかに制御することのできる一次元の操作軸を得ることが出来る.

今後の課題としては, 音源分離手法の導入が挙げられる. 実際には音源ごとに個別の信号を得ることが出来る場合は稀であり, CD などでミックスダウンされた音源に対しても提案手法を利用できるように音源分離手法の導入が必要となる. また, 立体音場再生にはバイノーラル再生を仮定しており, HRTF の個人性も問題となるため, そちらについても検討していく必要がある.

謝辞 本研究の一部は科研費 No. 25280060 及び No. 15K00235 によるものである.

## 参考文献

- [1] Swamy, M. and Thyagarajan, K.: Digital bandpass and bandstop filters with variable center frequency and bandwidth, *Proc. of the IEEE*, Vol. 64, No. 11, pp. 1632–1634 (1976).
- [2] Erfani, S. and Peikari, B.: Variable cut-off digital ladder filters, *International Journal of Electronics*, Vol. 45, No. 5, pp. 535–549 (1978).
- [3] Tan, E.: Variable lowpass wave-digital filters, *Electronics Letters*, Vol. 18, No. 8, pp. 324–328 (1982).
- [4] Regalia, P. and mitra, S.: Tunable digital frequency response equalization filters, *IEEE Trans. on Audio, Speech and Language Processing*, Vol. 35, No. 1, pp. 118–120 (1987).
- [5] Orfanidis, S.: Digital parametric equalizer design with prescribed nyquist-frequency gain, *Journal of the Audio Engineering Society*, Vol. 45, No. 6, pp. 444–455 (1997).
- [6] Engdegard, J., Resch, B., Falch, C., Hellmuth, O., Hilpert, J., Hoelzer, A., Terentiev, L., Breebaart, J., Koppens, J., Schuijers, E. and Oomen, W.: Spatial Audio Object Coding (SAOC) - The Upcoming MPEG Standard on Parametric Object Based Audio Coding, *124th Audio Engineering Society Convention* (2008).
- [7] Niwa, K., Nishino, T. and Takeda, K.: Development of selectable viewpoint and listening point system for musical performance, *Proc. of the 19th International Congress on Acoustics (ICA2007)*, PPA-06-011 (2007).
- [8] Kamado, N., Nawata, H., Saruwatari, H., Shikano, K. and Nomura, T.: Interactive controller for audio object localization based on spatial representative vector operation, *Proc. of International Workshop on Acoustic Echo and Noise Control (IWAENC2010)* (2010).
- [9] Itoyama, K., Goto, M., Komatani, K., Ogata, T. and Okuno, H.: Instrument equalizer for query-by-example retrieval: improving sound source separation based on integrated harmonic and inharmonic models, *Proc. of the 9th International Society for Music Information Retrieval (ISMIR 2008)*, pp. 133–138 (2008).
- [10] Blauert, J.: *Spatial hearing: the psychophysics of human sound localization*, MIT press (1996).
- [11] Takeda Lab., Nagoya Univ.: Head Related Transfer Functions Database, <http://www.sp.m.is.nagoya-u.ac.jp/HRTF/> [Online accessed; 21 Apr. 2015].