

# サウンドスコープヘッドフォン

浜中 雅俊\*<sup>1</sup> 李 昇姫\*<sup>2</sup>

Sound Scope Headphones

Masatoshi Hamanaka\*<sup>1</sup> Seunghee Lee\*<sup>2</sup>

**Abstract** – This paper describes Sound Scope Headphones, which enable the user to control an audio mixer through natural head and hand movements. When listening to music, a listener might want to hear particular parts more clearly or reduce the level of some parts. Such a capability is possible with commercial audio mixers that perform multi-channel recording and mix-downs. However, commercial audio mixers are too complicated for musical novices to be able to control numerous channel volumes and panpots. Our headphone device controls an audio mixer by using three sensors: a digital compass, a tilt sensor, and a distance sensor. These sensors are mounted on the headphones and they detect natural movements, such as of the head or a hand placed behind an ear when the user is listening to music. By putting their hand behind their ear, the listener adjusts the distance sensor on the headphones to focus on a particular part that they want to hear. We made three distance sensor prototypes and experimental results show that musical novices are more able to find a target instrument in a multi-channel recording with an infrared distance sensor.

**Keywords** : audio mixer, digital compass, tilt sensor, infrared distance sensor

## 1. はじめに

我々は人間の演奏者と計算機上の仮想演奏者がインタラクティブに即興演奏するジャムセッションシステムの研究 [1] ~ [5] において、人間のギタリスト 1 人と人間の個性を模倣した仮想ギタリスト 2 人がコード進行のみを決めた曲で自由にソロや伴奏を交代しながら即興演奏するシステム Guitarist Simulator を開発してきた。Guitarist Simulator は、元々ギタリスト向けに作成したシステムであるが、演奏初心者でも楽しめるよう、拡張を進めている。そこで問題となるのは、初心者が複数のパートの演奏を聴き分ける能力である。仮想演奏者が人間の音楽的な個性を獲得できていても、ユーザが、それを聴きわけることができなければ、仮想演奏者ごとの個性の違いを楽しむことは困難である。複数のパートを聴きわけるためには、パートごと別々のトラックに録音した音源を用意し、音楽用ミキサーで各パートの音量や定位（左右の音量比）を調節すればよい。しかし、市販の音楽用ミキサーは、操作が煩雑で初心者が直感的に扱うことは困難であった<sup>1</sup>。

そこで本研究では、音楽初心者でも複数パートの聴き分けが容易となるよう、直感的な操作で各パートのミキシングが変更できるヘッドフォン、サウンドスコープヘッドフォンを提案する。その特長は、頭を上下左右に振ったり、手を耳に近づけて耳を澄ませるようなポーズをするなど人間が音を聴くときに自然に行う動作を、ヘッドフォンに搭載した地磁気センサ、傾斜センサ、距離センサの 3 種類のセンサで検出することで音楽ミキサーのコントロールを可能にする点である。そして、頭を上下左右に振る動作や、手を耳に近づける動作を組み合わせることで、自分が聴きたいパートの音を探しながら演奏を聴くという音楽の新しい楽しみ方を可能とする。具体的には、左を向けば左から聴こえていたパートが正面で聴こえるように、右を向けば右から聴こえていたパートが正面で聴こえるようにする。また、上を向けば遠くに配置した、下を向けば近くに配置したパートが大きく聴こえるようにする。さらに、耳に手を近づけると、そのとき正面で聴いているパートがさらに強調され、あたかもフォーカスしたように聴くことができるようにする。

以下、2. 章ではサウンドスコープヘッドフォンの実現上の課題とその解決法を検討し、3. 章ではシステムの処理方法を説明する。そして 4. 章で実装について述べ、5. 章では関連研究について議論する。さらに、6. 章では評価実験について述べ、7. 章でまとめを述べる。

\*1: 筑波大学大学院システム情報工学研究科, 科学技術振興機構さきかけ (兼任)

\*2: 筑波大学大学院人間総合科学研究科

\*1: University of Tsukuba, Graduate School of SIE, PRESTO: Japan Science and Technology Agency

\*2: University of Tsukuba, Graduate School of CHS

1: たとえば、あるギタリストがソロを始めた瞬間に、そのギターを音量を上げ、定位を中央に寄せて、その他のパートの音量を少しづつ下げてその定位を調節するなど、一連の動作を瞬時に行う必要がある。

## 2. 音を探せるヘッドフォン

本研究では、直感的な操作で複数パートのミキシングを自由に変更できるヘッドフォンを作成することで、自分が聴きたいパートを探しながら演奏を聴くという音楽の新しい楽しみ方を提案する。たとえば、次のような状況を思い浮かべて頂きたい。

- 幼なじみから、ライブ演奏の録音をもらった。
- 幼なじみは、ギターを弾いているらしい。

このとき、我々は録音を聞きながら、自然とギターの音を探すであろう。そして、ギターの音をもっとよく聴きたいと思うであろう。本節では、そのようなことを実現するための課題とその解決法について述べる。

### 2.1 どのように聴きたい音を探すのか

「聴きたい音を探す」ことを可能にするためには、探した音とその他の音が区別できなければならない。本研究では、各パートの定位と音量に違いをもたせることで、探したパートとその他のパートが区別できるようにする。具体的には、探したパートは定位を中央に位置させ音量を大きくする。一方、その他のパートは定位を左右に振り、音量を小さくする。このようにすれば、たとえ楽器の聴き分けが苦手な音楽初心者でも、自分の聴きたい楽器を探し出すことが可能となる。

各パートの音量や定位を独立に調節するためには、パートごと別のトラックに録音された音源を用意する必要がある。そのような音源として RWC 研究用音楽データベース (RWC-MDB-J-2001) がある [6]。

### 2.2 どのように動作を検出するのか

直感的な操作によるミキサーのコントロールを可能とするためには、まず人間が音を聴くときに自然に行う動作を、検出する必要がある。本研究では、ヘッドフォンに、地磁気センサ、傾斜センサ、距離センサという3つのセンサを搭載することで、頭を上下左右に振る動作、耳を澄ませるポーズをして手を耳に近づける動作を検出する。具体的には、頭部の方位角、仰角および耳と手の距離を測定する。地磁気センサと傾斜センサはヘッドフォンのアーク部に取り付け、距離センサは右スピーカの外側に取り付けた (図1)。

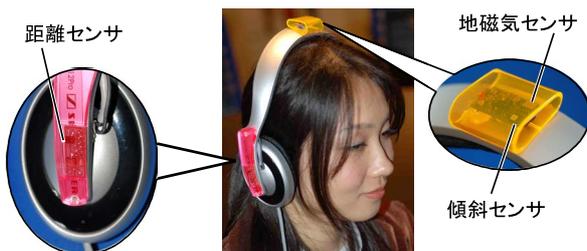


図1 ヘッドフォンにとりつけたセンサ

Fig.1 Sensors mounted on the headphones.

2.3 どのようにミキサーの操作と対応づけるか  
音を聴く動作とミキサーの操作との対応づけの良し悪しによって、操作性は変化する。本研究では、以下のような対応づけを試みた。

- 顔の方向との対応づけ

地磁気センサを用いて頭部の方向を検出し、右を向けば右から聞こえていた音が正面から聞こえるように、左を向けば左から聞こえていた音が正面から聞こえるようにする。すると、聴者は首を左右に動かすという自然な動作で、聴きたいパートを探し、そのパートが正面から聞こえるようにすることができる。このとき、各パートの配置は図2のようなグラフィカルユーザインタフェース (GUI) を用いて手動で変更することができる。中心にある円が聴者のアバタの位置と頭部の方向を表し、その周囲にある丸数字が各パートの配置をあらわしている。

- 顔の仰角との対応づけ

正面に複数のパートがある場合、頭を左右に動かし聴きたいパートが正面から聞こえるように操作しても、まだ聴きとりにくい場合が生じる。このとき、頭を上下に動かすことで、ミキシングを変更し目的とするパートを強調できるようにする。ここでは、ものを観るときの身体動作のメタファーを参考に対応づけを行う。具体的には、傾斜センサで頭部の上下動を検出し、上を向けば遠くに配置したパートの音量を大きくし、下を向けば近くに配置したパートの音量を大きくする。

- 耳に手をあてる動作との対応づけ

耳に手をかざして正面から聞こえる音を注聴する動作を、距離センサで検出する。そして、耳と手の距離に応じて聞こえるパートの範囲を設定する。具体的には、手を耳に近づけると、真正面のパートの音のみが聞こえるようにし、離すと真後ろ以外すべて聞こえるようにする。また、その中間ぐらいの距離にした場合には、前半分の音のみが聞こえるようにする。すると、耳と手の距離を適切に調節することで、あたかもフォーカスしたように自分の聴きたいパートを聴くことができるようになる。

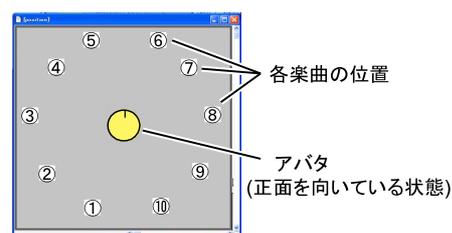


図2 パートを配置する GUI

Fig.2 GUI for locating the position of parts.

### 3. 処理方法

本章では，システムの処理の流れについて説明する．説明では，地磁気センサから得られる方位角を  $\theta$  ( $-\pi \leq \theta < \pi$ )，傾斜センサから得られる仰角を  $\phi$  ( $-\pi \leq \phi < \pi$ )，距離センサから得られる距離を  $\delta$  ( $0 \leq \delta \leq 1$ ) とする (図 3)．角度の単位はラジアンを用い，仰角，方位角は聴者が初期状態に向いている方向を 0 とする．距離センサは 0 から 3cm の距離を検出するが  $\delta$  はそれを 0 から 1 に正規化した値である．

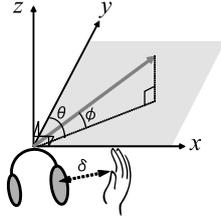


図 3 方位角  $\theta$ ，仰角  $\phi$  および手と耳の距離  $\delta$ .  
Fig. 3 Direction  $\theta$ , angle of elevation  $\phi$ , and distance  $\delta$ .

前処理  $n$  ( $0 \leq n \leq m$ ) 個のパートを複数トラックに分けて録音した音源  $S_n$  を用意し，各パートを計算機上の 2 次元空間上に配置する．同時に聴者のアバタも同じ空間上に配置する (図 2)．このとき，聴者のアバタから各パートまでの距離を  $l_n$ ，方位角を  $\theta_n$  とする． $l_n$  ( $0 \leq l_n \leq 1$ ) は一番遠くに配置したパートまでの距離を 1 として正規化した値である．  
ステップ 1 顔の仰角  $\phi$  に応じて変化する各パートの増幅率  $h_n^\phi$  ( $0 \leq h_n^\phi \leq 1$ ) を算出する．本研究では，上を向けばより遠くに配置したパートの音量が大きくなり，下を向けば近くに配置したパートの音量が大きくなるような関数として次式を用いる．

$$h_n^\phi = \begin{cases} 0 & \bar{h}_n^\phi < 0 \\ \bar{h}_n^\phi & 0 \leq \bar{h}_n^\phi < 1 \\ 1 & 1 \leq \bar{h}_n^\phi \end{cases} \quad (1)$$

ただし，

$$\bar{h}_n^\phi = 1 + l_n \sin \phi - \frac{1}{m} \sum_m l_m \sin \phi$$

各パートを図 4(a) のように配置した場合， $\phi = 0$  では，方位角  $\theta$  が同じであれば近く配置したパートも遠くに配置したパートも同じ増幅率となる (図 4(b))．一方  $\phi < 0$  では，遠くに配置したパートの増幅率が減少する (図 4(c))．図 4(b),(c) は増幅率をミキサー風に表示したもので，スライダーの左から順に 1 番から 10 番のパートに対応している．スライダーが上にあるほど，増幅率が高い．

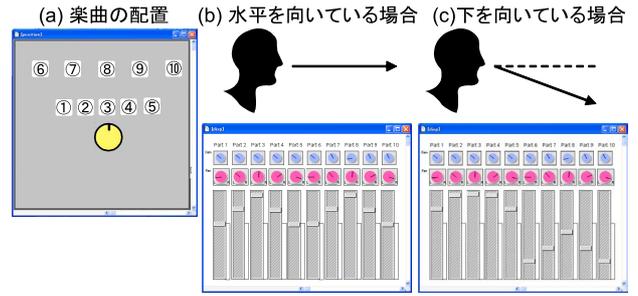


図 4 仰角  $\phi$  と各パートの増幅率.

Fig. 4 Angle of elevation  $\phi$  and the mixing console.

ステップ 2 手と耳の距離  $\delta$  に応じて変化する各パートの増幅率  $h_n^\delta$  ( $0 \leq h_n^\delta \leq 1$ ) を算出する．本研究では， $h_n^\delta$  として，0 または 1 のいずれかの値を取り，耳と手の距離が離れている状態では，すべてのパートで 1 を出力するが，近づくと，1 を出力する部分が次第に少なくなっていき，さらに近づくと，最終的には聴者の正面に配置したパートでのみ 1 を出力するような関数を定義する．ここでは，そのような関数として次式を用いる．

$$h_n^\delta = \begin{cases} 1 & \pi \cdot \delta \geq |\theta'_n| \\ 0 & \pi \cdot \delta < |\theta'_n| \end{cases} \quad (2)$$

$|a|$  は  $a$  の絶対値， $\theta'_n$  ( $-\pi \leq \theta'_n < \pi$ ) は， $\theta_n$  と  $\theta$  の作る角である．図 5 は， $\theta = \pi/3$ ， $\delta = 0.5$  の場合で，聴者が向いている方向の後ろ半分に配置したパートで  $h_n^\delta = 0$ ，前半分に配置したパートで  $h_n^\delta = 1$  となる．これにより不要なパートを除外し，自分の聴きたいパートを強調することができる．

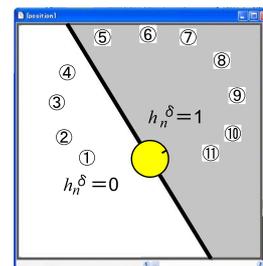


図 5 距離  $\delta$  と各パートの増幅率  $h_n^\delta$ .

Fig. 5 Distance  $\delta$  and amplification ratio  $h_n^\delta$ .

ステップ 3 顔の方位角  $\theta$  に応じて変化する各パートの増幅率  $h_n^\theta$  ( $0 \leq h_n^\theta \leq 1$ ) を算出する．本研究では，聴者が向いている方向に配置されているパートでは大きな値を，そうでないパートでは小さな値を示す関数として次式を用いる．

$$h_n^\theta = \begin{cases} 0 & \tilde{h}_n^\theta < 0 \\ \tilde{h}_n^\theta & 0 \leq \tilde{h}_n^\theta \end{cases} \quad (3)$$

ただし,

$$h_n^\theta = \begin{cases} 0 & \delta = 0 \\ 1 - (\alpha \cdot |\theta'_n|) / (\pi \cdot \delta) & \delta > 0 \end{cases}$$

各パートを図2のように配置しているとする、ユーザが左を向いているときの増幅率は図6(a)のようになり、正面を向いているときは図6(b)のようになり、右を向いているときは図6(c)のようになる。



図6 距離  $\theta$  と各パートの増幅率  $h_n^\theta$ .  
Fig. 6 Direction  $\theta$  and amplification ratio  $h_n^\theta$ .

$\alpha$  ( $0 \leq \alpha < 1$ ) は調節可能なパラメータで、 $\delta < 1$  の場合の増幅率の変化を設定する。 $\alpha = 0$  の場合には、耳に手を近づけても各パートの増幅率に変化はないが、 $\alpha > 0$  の場合には、耳に手が近づくにつれて増幅率が減少する。このとき聴者が向いている方向よりも、向いていない方向の増幅率の減少のほうが大きいので、正面の音が相対的に大きな音で聞こえるようになる。各パートを図2のように配置しているとする、 $\delta = 1$  のときの増幅率は図7(a)のようになる。耳に手を近づけて  $\delta < 1$  となった場合の増幅率は図7(b)のようになり、正面のパートがより明瞭に聞こえるようになっている。

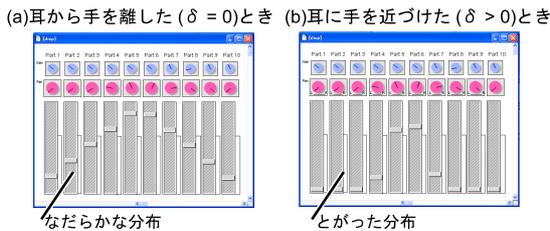


図7  $\alpha > 0$  のときの増幅率の変化.  
Fig. 7 Amplification ratio while  $\alpha > 0$ .

ステップ4 顔の方位角  $\theta$  に応じて変化する、各パートの左右の音の増幅率を算出する。本研究では、右を向くと右から聴こえていた音が正面から聴こえ、左を向くと左から聴こえていた音が正面から聴こえるような関数として、次式を用いる。

$$\text{右の増幅率 } h_n^{right} = \cos\theta \quad (4)$$

$$\text{左の増幅率 } h_n^{left} = \sin\theta \quad (5)$$

式(4), (5)は、等距離パンニングと呼ばれ、定位が移動しても音からの距離が常に一定に感じられる手

法である。この他に左右の増幅率の和を常に一定とするリニアパンニングという手法がある。6.章の評価実験で両者の比較を行う。

ステップ5 すべてのパートの信号を加算しヘッドフォンから出力する。その際、ステップ1から4までで求めた増幅率を掛け合わせる。 $S_{right}$  は右スピーカの出力、 $S_{left}$  は、左スピーカの出力である。右スピーカの出力:

$$S_{right} = \sum_n S_n \cdot h_n^\phi \cdot h_n^\delta \cdot h_n^\theta \cdot h_n^{right} \quad (6)$$

左スピーカの出力:

$$S_{left} = \sum_n S_n \cdot h_n^\phi \cdot h_n^\delta \cdot h_n^\theta \cdot h_n^{left} \quad (7)$$

#### 4. 実装

本章ではハードウェアの実装および、具体的な使用例について述べる。

##### 4.1 ヘッドフォンの製作

我々は、以下の2つの方針に基づき、ハードウェアの実装を行った。

- 軽量でかつ丈夫であること
- 計算機に容易に接続可能であること

ハードウェアの重量は177gで、USBケーブルとオーディオケーブルをパソコンに接続することで動作する。ヘッドフォン ヘッドフォンには、ヘッドバンドの内側にレールがありスピーカの位置が調節可能なゼンハイザー製 HD212Pro を採用した。ヘッドバンドの外側に距離センサを搭載できるので、頭の大きさに応じてスピーカの位置を調節した場合でも安定した動作が期待できるという特徴がある。

地磁気センサ、傾斜センサ ヘッドバンドの頭頂部にはアイチマイクロインテリジェント製の姿勢検知モジュール (AMI302-ATD) を搭載した。地磁気センサは地球の磁力線の方向を測定するため、一般に、傾斜角が大きくなるほど誤差が増大するが、このセンサは、地磁気センサと2軸の傾斜センサを搭載しており、傾斜誤差を補正できるという特徴がある。検出は120ミリ秒毎、分解能は2度である。

距離センサ ヘッドバンドの右外側には、手と耳の距離を測定するための距離センサを搭載した。距離センサとしては、曲げセンサを利用したものと、赤外線距離センサ (シャープ製 GP2S40J) を利用したものの2種類を試作した。前者は、プラスチックのレバーに取り付けた曲げセンサの抵抗値の変化を測定し距離を求めるもので、後者は、センサから出力された赤外線が物体に反射し戻ってくる際の赤外線の反射角度から距離を求めるものである。

プロテクター 各センサの破損や故障等を防ぐため、アクリルでプロテクターを製作した。赤外線距離セ

ンサのプロテクターの色は、様々なものを試した結果、センサが太陽光の影響を受けづらかったピンク色を使用した。

マイコンヘッドフォンの右ハウジング内部に、ルネサス製のマイコン (R8C/15) を搭載し、各センサからの情報を統合し、シリアル信号として出力するようにした。これにより、ヘッドフォンから出るワイヤーの本数や太さを減少できるようになった。マイコンは 120 ミリ秒毎に情報を出力している。各センサの近くに LED を搭載し、センサが正常に動いていない場合に点滅の間隔を変化させることで、動作確認ができるようにした。

USB 変換基板 マイコンから出力されるシリアル信号を USB 変換基板 (FTDI 製 FT232RL) を用いて、USB でパソコンと接続可能にした。これにより、マイコンおよびセンサを USB のバスパワーで駆動することが可能となり、電源が不要となった。

4.2 使用例 1: Guitarist Simulator での使用  
人間と計算機上の仮想演奏者が即興演奏する Guitarist Simulator において、各演奏者の聴きわけに使用した。Guitarist Simulator では 3 人のギタリストが同時に演奏する形式であるため、2 人が同時にソロを弾いている場合など、音楽初心者にとって聴きわけが特に困難となる場合が生じていた。

3 人のギタリストの演奏音を図 8 のように配置し、3 人の音楽初心者がそれを聴いた結果、全員が各ギタリストの演奏を聴きわけることができた。具体的には、人間の演奏者 B とその演奏者の個性を模倣した仮想演奏者 C を同時に演奏させた場合、その両者が似通った演奏をしていることに気が付くことができた。

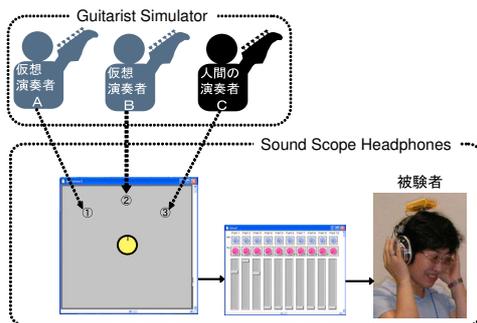


図 8 Guitarist Simulator での使用。  
Fig. 8 Using with Guitarist Simulator.

4.3 使用例 2: 複数の楽器の聴きわけでの使用  
ヘッドフォンを 10 種類の楽器の聴きわけに使用した。使用した曲はジャズのスタンダードナンバー (The Entertainer, RWC-MDB-J-2001, No. 38 [6]) で、複数の楽器が同時に演奏されているという特徴がある。各楽器を図 2 のように配置し、3 人の音楽初心者が使

用した結果、全員が各楽器の音色を聴きわけながら演奏を楽しむことができた。図 9 は、システムを展覧会で展示した様子である。展示の際には内容が伝わりやすくなるための工夫を施した。具体的には、楽器をステージの周囲に配置し、各パートの音量に応じてライトの明るさを ELATION 社製の MIDI 調光器 CYBERPAK で連続的に変化させた。そうすると、聴者が耳を澄ませるポーズをして、ある特定のパートを聴いている場合には、その楽器がライトで明るく照らされるようになるため、視覚的にも音量の変化を確認できるようになった<sup>2</sup>。



図 9 展覧会展示の様子。  
Fig. 9 Exhibition of the headphones.

4.4 使用例 3: 複数の楽曲の聴きわけでの使用  
Sound Scope Headphones の機能を利用して、多くの未知の曲の中から、自分の聴きたい曲の選択を可能とするシステムを構築した (図 10)。このシステムの特長は、頭を上下左右に振ったり手を耳にかざす直感的な操作により、多くの楽曲を音でプレビュー (プリリスニング) しながら、目的の楽曲を円滑に探し出すことを可能としている点である。システムには、選曲モードと再生モードという 2 つのモードがある。まず、選曲モードでは、10 曲が 2 次元空間上に配置され、同時に再生される。このとき、あたかも周囲を見渡すように、頭を上下左右に振ることで自分の聴きたい曲を探ることができる (楽曲プレビュー機能)。さらに、手を耳にかざし耳を澄ますポーズをすると、空間上で楽曲が聴こえる範囲が連続的に変化し、ユーザが聴きたいと思う楽曲を強調することができる (フォーカス機能)。ユーザが聴きたい曲を見つけその曲を 5 秒以上聴いていると、その曲が選択され、再生モードに移行する (スイッチ機能)。再生モードでは、通常 CD を聞くときと同じように 1 曲のみが再生される。再生モードで、左右に首を振ると選曲モードに移行する。

<sup>2</sup> 展示の様子を収めたビデオは、以下の URL からダウンロードできる。  
<http://music.iit.tsukuba.ac.jp/hamanaka/video/SSH.rv>

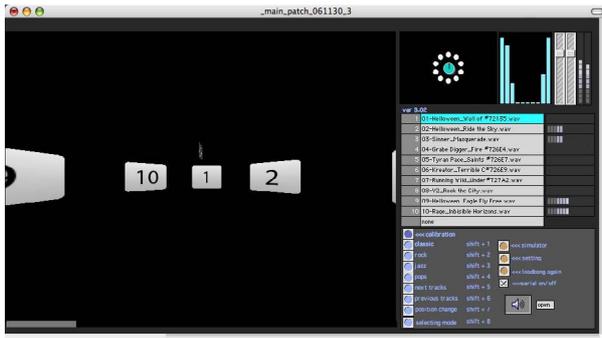


図 10 音楽検索への応用.  
Fig.10 Application for Music Retrieval.

## 5. 関連研究

本研究で提案したサウンドスコープヘッドフォンは、複数のパートや楽曲の聴き分けを容易にする新たな音楽鑑賞インタフェースである。本章では、サウンドスコープヘッドフォンの関連研究について検討し、本研究の位置づけを明確にする。

頭部の方向や位置を検出するセンサを搭載したヘッドフォンは従来から存在していたが、その目的は、仮定の音源位置を固定することで臨場感を高めることであったため、本研究のように各パートのボリュームや定位を積極的にコントロールする目的では使われてこなかった [7] ~ [12]。したがって、たとえば聴きたい楽器のすぐ近くに聴きたくない楽器が定位していた場合、聴きたい楽器のみを選択的に聴けるようにすることは困難であった。一方、空間音響システム [13], [14] では、聴取者のアバタや各パートの位置をマウスで移動することでミキシングを変更することを可能にしていた。しかし、各パートのミキシングを適切に変更するためには、ソロが始まったパートを近くに配置したり、ソロが終わり伴奏に戻ったパートを遠くに配置しなおすなどの煩雑な操作が必要なため、初心者には困難であった。本研究では、音楽初心者でも複数パートの聴き分けが容易となるよう、頭を上下左右に振ったり、手を耳に近づけて耳を澄ませるようなポーズなど、人間が音を聴くときに自然に行う動作をヘッドフォンに搭載した地磁気センサ、傾斜センサ、距離センサの 3 種のセンサで検出することで音楽用ミキサーのコントロールを可能とした。

サウンドスコープヘッドフォンは、4.4 節で紹介したように新たな楽曲選択インタフェースとしても使用することが可能である。従来の楽曲検索手法や楽曲推薦システムには、類似度に基づく手法 [15] ~ [18] や協調フィルタリングに基づく手法 [19], [20] などがあるが、これらは、多くの楽曲から限られた数の楽曲に絞り込むことには有効であるが、絞りこまれた後の楽曲

をどのように試聴するかについては議論されておらず、ユーザは 1 曲ずつ楽曲を試聴していかなくてはならなかった。これに対し、Musicream [21] では、流しそうめんのように楽曲が流れるインタフェースを提案した。また、Papipuum [22] や SmartMusicKIOSK [23] は、音楽要約、すなわち傷の付いたレコード盤上での針飛びで生じる速聴きのような方法を可能とした。これらのシステム [21] ~ [23] は、楽曲の選択や曲の試聴の時間を節約できるという点が優れていた。しかし、ボタンインタフェースや GUI を操作しながら曲を 1 曲ずつ試聴しなくてはならないという点では、従来の楽曲検索手法や楽曲推薦システムと同様の問題、すなわち、限られた時間内で出来る限り多くの楽曲を試聴しようとする、1 曲あたりの試聴時間が短くなっていき、曲の試聴よりも楽曲選択の操作に多くの時間や意識を費やさねばならないという問題があった。これに対して、本研究では多くの楽曲を同時に再生する楽曲レビューという新しい音楽の聴き方を提案した。

複数の音源を同時に再生し、各曲の音量を変化させることにより選曲を可能にしていた点では、Melting-Sound [24] が本研究と関連がある。Melting Sound ではディスプレイ上に配置した音源とマウスポイントとの距離によって各曲の音量のコントロールを行っていた。したがって、ユーザは画面を見ながら操作することを前提していた。これに対して、サウンドスコープヘッドフォンでは音を聴くときの直感的な動作を検出するため、ディスプレイを使用するのが困難な場合でも選曲が可能という特長がある。梅本らによって提案された一覧再生手法 [25] では、ユーザの正面に位置する仮想的な円周上に一定間隔で配置した 6 つの曲を一定速度で回転させることによって、複数の曲の聴き分けを実現していた。そして、送りボタン、戻しボタン、選択ボタンという 3 種類のボタンを操作することによって、ディスプレイが使用できない状態でも楽曲の選択を可能としていた。しかし、目標の曲を選択するためには、その曲が正面に来るまで待ってから選択ボタンを押す必要があるため受動的であるといえる。これに対し、サウンドスコープヘッドフォンは、各楽器の音量を能動的に変化させて望む曲を強調して聴くことが可能である。

## 6. 評価実験

本研究で構築したサウンドスコープヘッドフォンについて以下の 4 点を評価した。

- 適切なセンサが使用されているか
- センサとミキサーとの対応づけが適切であるか
- 楽曲選択インタフェースとして使用できるか
- ディスプレイのない環境でも選曲できるか

実験者は、音楽初心者3名で、3.章で述べた、調節可能なパラメータ  $\alpha$  は、予備実験を行い実験者ごとに任意に設定させた。

### 6.1 センサの比較

サウンドスコープヘッドフォンは、3種類のセンサを搭載しているが、このうち、手を耳に近づける動作を検出するセンサは様々なものが考えられ、そのうちどれを用いるかによって操作性に違いが出ると考えられる。そこで、(a) 手を耳に押し付ける強さを検出する曲げセンサを用いたものと、(b) 赤外線距離センサを用いたものの2種類を製作しどちらが調節しやすいかを評価した。この2種類のセンサはいずれも小型軽量で丈夫であるという特徴がある。実験では、複数パートの中から特定のパートを選択するまでの時間を計測した。参考までに、(c) 距離センサの代わりに可変抵抗を搭載したヘッドフォンについても測定を行った(図11)。使用したのは、10個のパートからなる曲RWC-MDB-J-2001, No. 38である。各パートは、図2のように配置した。実験の流れは以下の通りである。

1. 各パートの位置をランダムに変更し、10個のパートのうち、1つをランダムに指定。
2. 曲の開始する位置をランダムで決定し、その位置から再生を開始。
3. 指定したパートを見つけるまでの時間を測定。実験者は、各パートの名前とその音色をあらかじめ知っており、それぞれのヘッドフォンを30分間使用し、十分に操作に慣れた後、実験を開始した。10試行ごとにヘッドフォンを交換し、それぞれのヘッドフォンについて100回の試行を行った。

表1は100回の結果の平均で、曲げセンサを用いたものに比べて赤外線距離センサを用いたもののほうが、より迅速にパートの選択ができることがわかった。曲げセンサを用いたものは、アクリルのレバーがばねの働きをしており、微妙な操作がしにくいという意見が多かった。パートを見つけるまでの時間が3人の平均で最も短かったのは赤外線距離センサであることから、手を耳に近づける動作を検出するためのセンサとしては、赤外線距離センサが適切であると言える。そこで、以下の実験では赤外線距離センサを用いる。

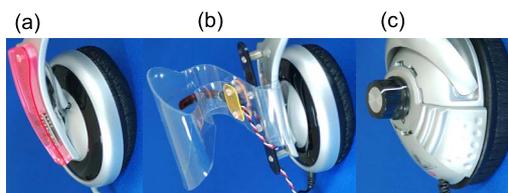


図11 3種類のセンサ。  
Fig. 11 Three types sensor.

表1 3種類のセンサの比較。  
Table 1 Comparison of three focus sensors.

	(a) 曲げセンサ	(b) 赤外線センサ	(c) 可変抵抗
実験者 A	1.28 sec.	1.12 sec.	1.84 sec.
実験者 B	1.04 sec.	0.84 sec.	0.72 sec.
実験者 C	2.01 sec.	0.74 sec.	1.02 sec.
平均	1.44 sec.	0.90 sec.	1.19 sec.

### 6.2 ミキサーとの対応づけの比較

音を聴くときの自然な動作と、音量変化との対応づけの良し悪しによっても、操作性は変化すると考えられる。そこで本節では、対応づけに変更を加えた場合に性能が変化するかを評価する。具体的には、今回は、3.章で述べたように、首を左右に動かしても音源までの距離が同じに感じられる等距離パンニングという手法を用いたが、これを (b) 左右の増幅率の和が常に一定となるリニアパンニングに変更した場合の性能について評価した。また、今回は手を耳に近づけると正面のパートが強調されるような対応づけを行ったが、逆に (c) 手を耳から離すと正面のパートが強調されるように対応づけを行った場合の性能を評価した。

(b) と (c) について、6.1 節と同じように特定のパートを見つけるまでの時間の測定を行った。実験は、(b) と (c) それぞれに設定したヘッドフォンを30分間使用し、十分に操作に慣れた後開始し、10試行ごとにヘッドフォンを交換し、100回の試行を行った。表2は100回の測定値を平均したものである。ただし (a) は、表1の赤外線センサの結果と同じである。

実験の結果 (a) 等距離パンニングに比べて (b) リニアパンニングのほうがパートを見つけるまでの時間が長かったことから、(a) 等距離パンニングのほうが適していると考えられる。一方、(a) 手を耳に近づけると強調される場合と、(c) 手を耳から離すと強調される場合では、ほとんど違いがなかった。ただし実験後、実験者から意見を聞いたところ、いずれの実験者も (a) 手を耳に近づけると強調される場合のほうが自然に操作できたと述べた。

今回の実験では、パートを見つけるまでの時間の測定で性能を評価したが、今後、操作の快適性や疲労度などについて測定する方法について検討していきたい。

表2 ミキサーとの対応づけの比較。  
Table 2 Comparison of links between functions of audio mixer and movements.

	(a) 等距離	(b) リニア	(c) 離すと強調
実験者 A	1.12 sec.	1.34 sec.	1.02 sec.
実験者 B	0.84 sec.	1.45 sec.	0.97 sec.
実験者 C	0.74 sec.	0.99 sec.	0.92 sec.
平均	0.90 sec.	1.26 sec.	0.97 sec.

### 6.3 標準プレーヤとの比較

4.4節で提案した楽曲選択システムを使用して、円滑な選曲が可能であるか評価実験を行った。実験には、RWC 研究用音楽データベース：ジャズ音楽 (RWC MDB-J-2001) を用いた。データベースには、50 のジャズ曲が収められており、そのうちソプラノサクスの含む楽曲は 1 曲である。そして、そのソプラノサクスの含む曲を探し出すまでの時間を測定した。実験者は、データベースにある曲を聴いたことはないが、ソプラノサクスの音色は知っていた。楽曲リストは、データベースにある順で与え、ソプラノサクスの含む楽曲は 50 曲中 50 番目にあった。ただし、実験者には、データベースに何曲の曲があるかは知らせていない。表 3 は、測定結果をまとめたものである。いずれの実験者も短時間で、目的の曲を見つけることができた。提案システムを用いた測定の後に Windows の標準プレーヤである Windows Media Player を用いて同様の測定を行った。用いたのは同じデータベースであるため、実験者が曲の内容を覚えている場合には標準プレーヤが有利になる可能性が高い。しかし、結果は逆で、いずれの実験者も提案するシステムを用いた場合のほうが、短時間でソプラノサクスの含む曲を選択できた。この結果は、楽曲リストから特定の音色を含む曲など、目的の曲を探すのに有効なシステムであることを示している。

表 3 提案システムと標準プレーヤとの比較。  
Table 3 Comparizon of our system and standard music player.

	提案システム	標準プレーヤ
実験者 A	224 sec.	845 sec.
実験者 B	423 sec.	1145 sec.
実験者 C	642 sec.	751 sec.
平均	429 sec.	914 sec.

### 6.4 ボタンインタフェースとの比較

ディスプレイを使用するのが困難な場合に (a) 本システムと (b) 梅本ら [25] によって提案されたようなボタンインタフェースによる選曲方法とのどちらが有効であるか評価した。本研究では (b) として、10 曲がユーザの周囲に配置され、それらが一定の速度 (15 度/秒) で回転し送りボタン、戻しボタン、選択ボタンという 3 種類のボタンで選曲を行うシステムを実験用に構築した。そして、(a),(b) の両方とも、公園を散歩しながら画面を見ない条件で操作した。

実験には、RWC 研究用音楽データベース：ジャズ音楽 (50 曲)、ポピュラー音楽 (100 曲) からランダムで抽出した 10 曲から一番好きな曲を選択するまでの時間を測定した。表 4 は、10 試行ごとに (a) と (b) を切り替え、それぞれ 100 回の試行を行った平均である。

実験の結果、提案システムのほうがボタンインタフェースによる選曲方法より短い時間で選曲できることがわかった。実験後、実験者の 1 人からは (b) を使用すると音が周囲を回転し、乗り物酔いをしたような感じになったとの報告があった。また、残る 2 人の実験者からも、提案システムのほうが快適に選曲できたとのコメントが得られた。

表 4 提案システムとボタンインタフェースによる選曲システムとの比較。

Table 4 Comparizon of our system and music selection by button interface.

	(a) 提案システム	(b) ボタンインタフェース
実験者 A	15.4 sec.	18.7 sec.
実験者 B	13.2 sec.	19.4 sec.
実験者 C	18.2 sec.	18.6 sec.
平均	15.6 sec.	18.9 sec.

## 7. まとめ

本論文では、人間が音を聴くときに自然に行う動作をヘッドフォンに搭載した、地磁気センサ、傾斜センサ、距離センサで検出し、ミキシングを変更するサウンドスコープヘッドフォンについて述べた。

本研究の意義として、まず第一に、複数のパートの中から特定のパートを強調する手法を提案し、自分の聴きたいパートを強調しながら聴く新たな音楽の楽しみ方を実現した点があげられる。第二に、ヘッドフォンにセンサーを搭載し、音を聴くときに自然に行う動作を検出し、さらに検出した動作とミキサーの操作との対応づけを行った点があげられる。これにより、ユーザは煩雑なマウス操作や音楽用ミキサーの操作から開放され、円滑にパートを選択することが可能となった。第三に、構築したシステムを応用して、楽曲選択システムを実現した点があげられる。これにより、ディスプレイが見れない状況でも、楽曲選択が可能となり、ユーザが未知の楽曲に触れる機会を増やすことが可能となった。第四に、実験により提案システムの性能を評価した点があげられる。センサの比較では、耳を澄ませるポーズを検出するセンサとして、赤外線距離センサが適していることがわかった。また、センサから検出した動作とミキサーの対応づけの比較をした結果、左右の増幅率の決定はリニアパンニングより等距離パンニングが適していることがわかった。さらに、提案した楽曲選択システムは、特定の音色を含む曲など、目的の曲を探すのに有効で、ボタンインタフェースと比べても操作しやすいことが確認できた。

今後、さらに詳細な実験を行い、提案システムのどの機能が優れているのかについて検討していく。また、構築したヘッドフォンのその他の利用法についても検討していきたい。

参考文献

- [1] Masatoshi Hamanaka, Masataka Goto, Hideki Asoh, Nobuyuki Otsu: A Learning-Based Jam Session System that Imitates a Player's Personality Model, Proceedings of the 2003 International Joint Conference on Artificial Intelligence(IJCAI2003) , pp. 51-58 (2003).
- [2] 浜中 雅俊, 後藤 真孝, 麻生 英樹, 大津 展之: Guitarist Simulator: 演奏者の振舞いを統計的に学習するジャムセッションシステム, 情報処理学会論文誌, Vol. 45, No. 3 , pp. 698-709 (2004).
- [3] 浜中 雅俊, 後藤 真孝, 麻生 英樹, 大津 展之: 学習するジャムセッションシステム: 演奏者固有のフレーズの獲得, 情報処理学会 音楽情報科学研究会 研究報告 2002-MUS-47-13 , Vol. 2002, No. 100, pp. 71-78 (2002).
- [4] Masatoshi Hamanaka, Masataka Goto, Hideki Asoh, Nobuyuki Otsu: A Learning-Based quantization: Unsupervised Estimation of the Model Parameters, Proceedings of the 2003 International Computer Music conference (ICMC2003), pp. 369-372 (2003).
- [5] 浜中 雅俊: 個性を模倣した仮想演奏者の実現, 情報処理学会会誌, Vol. 47, No. 4, pp. 374-380 (2006).
- [6] 後藤 真孝, 橋口 博樹, 西村 拓一, 岡 隆一: RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース, 情報処理学会論文誌, Vol.45, No.3, pp.728-738 (2004).
- [7] Oliver Warusfel, and Gerhard Eckel: LISTEN - Augmenting Eeveryday Environments Through Interactive Soundscapes, Proceedings of IEEE Workshop on VR for public consumption , IEEE Virtual Reality, pp. 268-275 (2004).
- [8] Jiann-Rong Wu, Cha-Dong Duh, Ming Ouhyoung, Jei-Tun Wu: Head Motion and Latency Compensation on Localization of 3D Sound in Virtual Reality, Proceeding of ACM Symposium on Virtual Reality Software and Technology, Lausanne, Switzerland, ACM Virtual Reality Software and Technology, pp. 15-20 (1997).
- [9] Camille Goudeseune, and Hank Kaczmariski: Composing Outdoor Augmented-reality Sound Environments, Proceedings of International Computer Music Conference (ICMC2001), pp. 83-86 (2001).
- [10] 佐藤 光一, デジタルコードレスサラウンドヘッドフォンの開発, PIONEER 技術情報誌, Vol. 14, No. 2, pp. 66-73 (2004).
- [11] 久木元 伸如, Ewe C. Huat, 竹田 仰, プロジェクション型没入ディスプレイにおける3次元音場生成のための実用的検討, 電子情報通信学会 技術研究報告, EA99-32, pp. 29-36 (1999).
- [12] 比嘉 恭太, 西浦敬信, 木村朝子, 柴田史久, 田村秀行, 視覚・聴覚を併用した複合現実感システムの開発(1) 視覚的MR と聴覚的MR の同時提示の実現 , 日本バーチャルリアリティ学会第11回大会論文集, pp. 285-288 (2006) .
- [13] Francois Pachet, and Oliver Delerue.: A Mixed 2D/3D Interface for Music Spatialization, Proceedings of International Conference on Virtual Worlds, pp. 298-307 (1998).
- [14] Francois Pachet, and Oliver Delerue: On-The-Fly Multi-Track Mixing, Proceedings of AES Convention, Audio Engineering Society (2000).
- [15] Elias Pampalk: A MATLAB toolbox to compute music similarity from audio, Proceedings of International Conference on Music Information Retrieval(ISMIR2004), pp. 254-257, (2004).
- [16] Thomas Soding and Alan. F. Smeaton: Evaluating a music information retrieval system - TREC style, Proceedings of International Conference on Music Information Retrieval(ISMIR2002), pp. 71-78, (2002).
- [17] George Tzanetakis and Perry Cook: Musical genre classification of audio signals, IEEE Transactions on Speech and Audio Processing, Vol. 10, pp. 293-302, (2002).
- [18] Fabio Vignoli and Steffen Pauws: A music retrieval system based on user-driven similarity and its evaluation, Proceedings of International Conference on Music Information Retrieval(ISMIR2005), pp. 272-279, (2005).
- [19] William W. Cohen and Wei Fan: Web-collaborative filtering: Recommending music by crawling the Web, Proceedings of the 9th international World Wide Web conference on Computer networks/the international journal of computer and telecommunications netowrking, Vol. 33, pp. 685-698, (2000).
- [20] Alexandra Uitdenbogerd and Ron van Schyndel: A review of factors affecting music recommender success, Proceedings of International Conference on Music Information Retrieval(ISMIR2002), pp. 204-208, (2002).
- [21] 後藤 孝行, 後藤 真孝: Musicream: 楽曲を流してくっつけて並べることのできる新たな音楽再生インタフェース, WISS2004 論文集, pp. 53-58, (2004).
- [22] Keiji Hirata and Shu Matsuda: Interactive Music Summarization Based on GTTM, Proceedings of International Conference on Music Information Retrieval(ISMIR2002), pp. 86-93, (2002).
- [23] 後藤 真孝: SmartMusicKIOSK: サビ出し機能付き音楽試聴機, 情報処理学会論文誌, Vol. 44, No. 11, pp. 2737-2747, (2003).
- [24] 神原 啓介, 安村 通晃: MeltingSound: なめらかなオーディオブラウジング, ヒューマンインタフェースシンポジウム 2003 論文集, pp. 817-820, (2003).
- [25] 梅本 あずさ, 柴尾 忠秀, 水口 充, 浦野 直樹: 音声提示型インタフェースの実装と評価, 第7回インタラクティブシステムとソフトウェアに関するワークショップ (WISS'99), pp. 83-92, (1999).

(2007年3月26日受付)

[ 著者紹介 ]

浜中 雅俊 (正会員)



2003年筑波大学大学院工学研究科電子・情報工学専攻博士課程修了。2003-2004年日本学術振興会特別研究員PD, 2004年より2006年まで科学技術振興機構さきがけ研究員(専任)として独立行政法人産業技術総合研究所において音楽情報処理の研究に従事。現在,筑波大学大学院システム情報工学研究科講師。2004-2005年オランダ・ナイメヘン情報認知研究所(NICI)客員研究員。博士(工学)。2001年情報処理学会山下記念研究賞,2001年SCI(5th World Multiconference on Systemics Cybernetics and Informatics) in Art 優秀論文賞,2003年筑波大学大学院優秀論文賞(博士課程長賞),2005年ICMC2005 Best Paper Award(Journal of New Music Research Distinguished Paper Award)各賞受賞。

李 昇姫



1999年筑波大学大学院芸術学研究科博士課程修了。博士(デザイン学)。同年,オランダデルフト工科大学デザイン工学部主任研究員。2000年筑波大学芸術学系講師。2001年より筑波大学大学院人間総合科学研究科感性認知脳科学専攻講師として現在至る。2004年オランダアイントホーベン工科大学客員研究員。1997年アジアデザイン学会最優秀論文賞受賞。2003年度21世紀COE拠点形成プログラム「こころを解明する感性科学の推進」の事業推進担当委員。現在は,感性情報評価およびTangible技術を用いた感性インタラクションの研究に取り組む。