

Sound Scope Headphone: 音楽用ミキサーを コントロールするヘッドフォン型デバイス

浜中雅俊^{†1} 李昇姫^{†2}

^{†1}科学技術振興機構 さきがけ研究員

^{†2}筑波大学大学院 人間総合科学研究科

m.hamanaka@aist.go.jp

本研究報告では、初心者が音楽を能動的に楽しめるようにすることを旨とし、直感的な操作で複数パートのミキシングが変更できるヘッドフォンを提案する。従来の音楽用ミキサーは煩雑で、音楽初心者が複数パートのボリュームや定位を適切に変更することが困難であった。そこで本研究では、ヘッドフォンに搭載した3種類のセンサで、首を上下左右に振ったり耳を澄ませるポーズをするなど、人間が音を聴くときにする自然な動作を検出し、それをミキサーのコントロールに反映させることで直感的な操作によるミキシングを実現する。

Sound Scope Headphone: Headphone Device for Controlling Audio Mixer

Masatoshi Hamanaka^{†1} Seunghee Lee^{†2}

^{†1}PRESTO, Japan Science and Technology Agency,

^{†2}University of Tsukuba Graduate School of Comprehensive Human Science

A.I.S.T. Mbox 0604. 1-1-1Umezono, Tsukuba, Ibaraki 305-8568 Japan

In this report, we propose a headphone device which can control an audio mixer by natural behavioral manipulation, in order to enjoy music for musical novices. A commercial audio mixer is too complicated for musical novices to control multi-channel volumes and panpots properly. Our headphone device can control an audio mixer using three kinds of sensors mounting on the headphone which detect natural movement when he/she listening to a music.

1. はじめに

我々はジャムセッションシステムの研究[1-4]において、実在する演奏者の音楽的な個性を模倣できる仮想演奏者を生成することを目指し **Guitarist Simulator** を開発してきた。**Guitarist Simulator** は、人間のギタリストと仮想ギタリストあわせて3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していくシステムである。その特長は、人間のギタリストの5分間のMIDI演奏記録を統計的に学習することで、演奏者の振舞いモデル、フレーズモデル、発音時刻ゆらぎモデルという3つのモデルを獲得していた点で

ある。そして、それらのモデルを用いて仮想ギタリストを動作させることによって、演奏者の音楽的な個性の模倣を実現していた。

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、現在拡張を進めている[5]。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。たとえ、複数の仮想演奏者がそれぞれ人間の演奏者の音楽的な個性を獲得していたとしても、ユーザが、それを聴きわけることができなければ個性の違いを感じることは困難である*。

*実際、**Guitarists Simulator**では、3人のギタリストの聴き分けが難しいとの意見があった。

複数パートを聴き分けるためには、パートごとと複数トラックに録音された音源を用意して、音楽用ミキサーで各パートのボリュームや定位を調節すればよい。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操作することは困難であった。たとえば、あるギタリストがソロを始めた瞬間に、そのギターのリズムを上げ、定位を中央に寄せて、その他のボリュームを少しず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ことは初心者には困難であった。また、聴取者のアバターや各パートの位置をマウスで移動してミキシングを変更するシステムでも、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操作が必要で、初心者には困難であった[6-7]。

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を変更できるヘッドフォン **Sound Scope Headphone** を提案する。**Sound Scope Headphone** 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電子コンパス、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した点である。

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、仮想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、本研究のように各音源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[8-11]。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聴けるようにすることは困難であった。

本研究で提案する **Sound Scope Headphone** は、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に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節では **Sound Scope Headphone** を

実現するための課題と解決法について述べ、3節ではシステムの処理方法について説明する。そして4節で実装および展示について述べ、5節でまとめと今後の課題について述べる。

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、次のような状況を考えて頂きたい。

- ・幼なじみから、ライブ演奏の録音をもらった。
- ・幼なじみは、サクソフオンを弾いているらしい。

このようなとき、我々は録音された演奏を聞きながら、自然とサクソフオンの音を探すであろう。そして、サクソフオン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、各パートの定位とボリューム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した。具体的には、探したパートは定位を中央に位置させボリュームを大きくする。一方、その他のパートは定位を左右に振り、ボリュームを小さくする。このようにすれば、たとえ楽器の聴き取りが苦手な音楽初心者でも、容易に自分の聴きたい楽器を探し出すことが可能となる。

このように各パートの音量や定位を調節するためには、パートごとに複数トラックに分けて録音された音源を用意する必要がある。本研究では、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 No. 38)を使用した [12]。

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にするためには、まず人間が音を聴くときに自然に行う動作を、なんらかの方法で検出する必要がある。本研究では、ヘッドフォンに、電子コンパス、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。電子コンパスと傾斜センサはヘッドフォンのアーク部に取り付け，距離センサはスピーカ部の外側に取り付けた(図 1)。距離センサは，当初プラスチックのレバーに取り付けた曲げセンサを用いていたが，現在では赤外線距離センサを用いている。



図 1: ヘッドフォンにとりつけた 3 種類のセンサ

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，ユーザビリティは大きく変化する。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

電子コンパス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する。同様に，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。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。

・顔の仰角との対応づけ

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，まだ正面に複数のパートがあり，目的とするパートが聴きとりにくい場合には，今度は頭を上下に動かすことで，ミキシングを変更することができる。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。このとき，各パートの配置は図 2 のような 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。

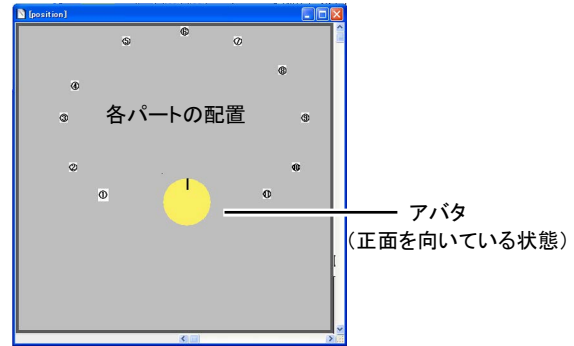


図 2: パートを配置する GUI

・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や声を注聴するような動作を，距離センサで検出する。そして，耳と手との距離によって，聞こえるパートの範囲を設定している。たとえば，手を耳に近づけると，真正面から聞こえるパートの音のみが聞こえるようになり，離すと真後ろ以外すべて聞こえるようになる。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なる。耳と手との距離を適切に設定することによって，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。

3. 処理方法

本節では，システムの処理の流れについて説明する。以下の説明では，電子コンパスから得られる方位角を θ ($-\pi \leq \theta < \pi$)，傾斜センサから得られる仰角を ϕ ($-\pi \leq \phi < \pi$)，距離センサから得られる距離を δ ($0 \leq \delta \leq 1$) 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で向いている方向を 0 とする。距離センサは 0cm から 3cm の距離を検出するが δ はそれを 0 から 1 までの値に正規化したもので，距離が 0cm のときに 1，距離が 3cm またはそれより長いときに 0，0cm と 3cm の間のときは 0 と 1 の間の値を出力する。

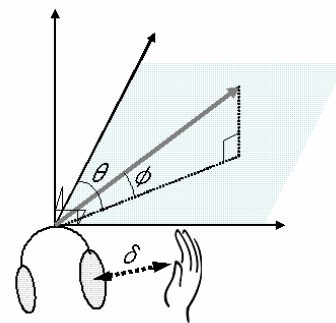


図 3: 方位角 θ ，仰角 ϕ および手と耳の距離 δ

前処理： パートごと複数トラックに分けて録音した音源 S_n を用意し、それぞれのパートの位置を計算機上の2次元空間上に配置する(図2)。同時に聴者のアバタも同じ空間上に配置する。このとき、聴者のアバタから各パートまでの距離を l_n 、方位を θ_n とする。ただし、 θ_n の単位はラジアン、 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： 仰角 ϕ に応じて変化する各パート n の増幅比(減衰比) h_n^ϕ ($0 \leq h_n^\phi \leq 1$)を算出する。ここでは、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次式を用いることにした。

$$h_n^\phi = 1 + l_n \sin \phi - \frac{\sum l_m \sin \phi}{m} \quad (1)$$

h_n^ϕ は仰角が正の場合、遠くに配置したパートでは大きな値を示し、近くに配置したパートでは小さな値を示す関数である(図4)。

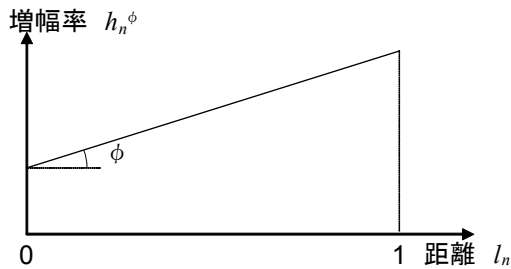


図4: 仰角 ϕ と h_n^ϕ との関係($\phi > 0$ の場合)

ステップ2： 距離 δ に応じて変化する各パートの増幅率 h_n^δ を算出する。 h_n^δ は、0または1のいずれかの値をとる関数で、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、強く押すと、聴者が向いている方向に配置したパートでのみ1を出力するようになる。 $|a|$ は a の絶対値、 θ_n' ($-\pi \leq \theta_n' < \pi$)は、 θ_n と θ の作る角である。

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

たとえば、 $\theta = 30^\circ$ 、 $\delta = 0.5$ の場合では、図5のように、そのとき聴者が向いている方向の後ろ半分の領域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。これにより不要な音を削除することができる。

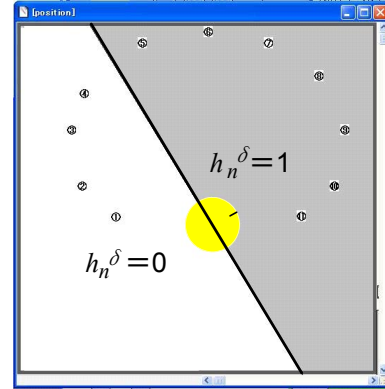


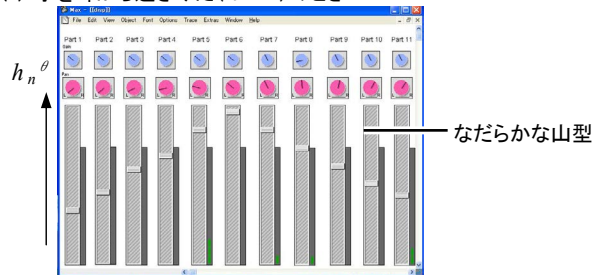
図5: 距離 δ と h_n^δ との関係

ステップ3： 電子コンパスから得られる方位角 θ に応じて変化する各音源の増幅率 h_n^θ を算出する。 h_n^θ 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である。

$$h_n^\theta = 1 - \frac{|\theta_n'|}{\pi(1-\delta \cdot \alpha)} \quad (3)$$

α ($0 \leq \alpha \leq 1$)は調節可能なパラメータで、耳を澄ますポーズをして耳に手を近づけて $\delta > 0$ となった場合の増幅率の変化を設定する。 $\alpha = 0$ の場合には、耳を澄ますポーズをしても各パートの増幅率に変化はないが、 $\alpha > 0$ の場合には、耳に手が近くなるに応じて増幅率が減少する。このとき聴者が向いている方向のパートの増幅率の減少よりも、向いていない方向のパートの増幅率の減少のほうが大きいため、正面の音が相対的に大きな音で聞こえるようになる(図6)。

(a) 手を耳から遠ざけた($\delta = 0$)のとき



(b) 耳に手を近づけたとき($\delta > 0$)のとき

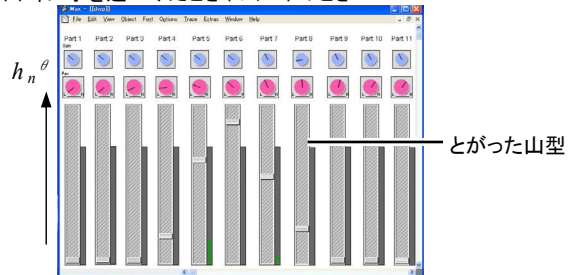


図6: $\alpha > 0$ のときの増幅率の変化

ステップ4: 電子コンパスから得られる方位 θ に応じて、各パートの定位 p_n ($0 \leq p_n \leq 1$) を算出する。 p_n が 0 のとき、そのパートの右の音と左の音の比率は 0 : 1, p_n が 0.5 のとき、1 : 1 とする。 β は調節可能なパラメータで、 $\beta = 0$ の場合には、耳を澄ますポーズをして耳に手を近づけても定位に変化はないが、 $\beta > 0$ の場合には耳に手を近づけるにしたがって、そのとき正面にあるパート以外のパートの定位が後ろに移動していく。たとえば、右前方に定位しているパートは右後方に定位が移動する。このようにすることで、あたかも正面のパートにフォーカスしたように聴くことができるようになる。

$$p_n = \frac{1}{2} + \frac{\theta'_n}{\pi(1 - \delta \cdot \beta)} \quad (4)$$

通常のヘッドフォンでは、たとえば、左右の音量比が 7 対 3 の場合、右方向から音がすることはわかるが、それが右前方なのか右後方なのかはわからない。本研究の場合は聴者が首を左右に動かすことによってそれがわかるようにしている。たとえば、聴者が右へ向いたとき、右前方に定位する音は、左方向に移動するのに対し、右後方に定位する音は右方向に移動する。ただしここでの定位はあくまで頭内定位である。奥行き感等を感じさせる頭外定位を実現するためには HRTF (頭部伝達関数) [13] 等を考慮する必要がある。本研究の目的は音楽ミキサーを直感的に操作することであるからここでは HRTF については扱わない。

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。

ヘッドフォンの右側の出力

$$S_{Right} = \sum_n S_n \cdot h_n^\phi \cdot h_n^\delta \cdot h_n^\theta \cdot p_n \quad (5)$$

ヘッドフォンの左側の出力

$$S_{Left} = \sum_n S_n \cdot h_n^\phi \cdot h_n^\delta \cdot h_n^\theta \cdot (1 - p_n) \quad (6)$$

4. 実装および展示

Sound Scope Headphone は、これまで数回の展示を行い、そのたびに改良を重ねてきた。本節では、それぞれの展示の内容について紹介する。プログラミング環境としては Max/MSP [14] を使用した。

4.1 音楽情報科学研究会デモセッション

2005 年 8 月 4 日に NTT 西京阪名ビルで開催された第 61 回情報処理学会音楽情報科学研究会のデモセッション[15]に 2 種類のヘッドフォンを展示した。

・センサにボリュームを用いた試作機

試作機として、センサにボリュームを使ったものを作成した。360 度回転式ボリュームを埋め込んだ回転台を作成し、聴者がその上に乗り回転することで電子コンパスの代用とした。また耳元のセンサもボリュームで代用した。ボリュームの値は Infusion 社の I-CubeX[16] を用いて AD 変換し MIDI インタフェイスを通じて Max 上に取り込んでいる。展示では、回転台に椅子を取り付け座れるようにした。見学者の意見では、耳元のボリュームをどちらに回転すれば良いのかわからなくなることがあり、操作しづらいというものが多かった (図 7)。



図 7: センサにボリュームを用いた試作機

・電子コンパスを搭載した試作機

センサにフィリップス社の電子コンパス KMZ51 および曲げセンサを用いたものを試作した。電子コンパスの出力は自作回路で 8 ビットに変換し I-CubeX で Max 上に取り込んだ。曲げセンサは塩ビ板を U 字型に曲げたレバーの湾曲部に取り付けた。見学者からはレバーが細く押しにくいとの意見があった (図 8)。



図 8: センサに電子コンパスを用いた試作機



図 9: インタラクティブ東京での展示風景

4.2 インタラクティブ東京 2005

2005年8月25・26日に日本科学未来館で開催されたインタラクティブ東京 2005[17]に出展した。インタラクティブ東京は、一般への公開を目的としているため、一般の方に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、パートの個数分の譜面台に楽器の絵を載せ、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(図9)。

センサは、電子コンパス(3軸MIセンサ)と2軸の傾斜センサが一つの基板上に載ったアイチ・マイクロ・インテリジェント製のAMI302-ATDを用いた。これを採用した理由は、傾斜センサがついていることによって、仰角が読み取れることと、電子コンパスの傾斜補正ができる点、USBで直接パソコンに接続できる基盤に実装されている点である。

操作性を上げるため、曲げセンサを取り付けるレバーを大型化していたが、多くの来場者が訪れたため、最終日展示終了直前にレバーが疲労破損するというアクシデントがあった。

4.3 さきがけライブ 2005

2005年12月22日に東京国際フォーラムで開催されたさきがけライブ[18]に出展した。さきがけライブでは、インタラクティブ東京で展示したセットに加え、ワイヤレス化したヘッドフォンの展示も行った(図10)。ワイヤレス版では、電子コンパスにInfusion社のOrient v1.1、傾斜センサにはオムロン社製のD5R-L02を使用した。いずれもI-CubeXに容易に接続できるという特長がある。I-CubeXはBluetoothでパソコンに接続できるタイプのものを用いた。バッテリーは、ワイヤレスヘッドフォン用の単三アルカリ電池2本と、I-CubeX用の9Vアルカリ電池(006P)を使用していたが、006Pの消耗が早く、駆動可能な時間は約2時間であった。



図 10: ワイヤレス版 Sound Scope Headphone

4.3 文化庁メディア芸術祭

2006年2月24日から3月5日まで開催される文化庁メディア芸術祭[19]の「先端技術ショーケース 未来のアート表現のために」に展示を予定している。期間中多くの見学者が来場することが予想されるため、これまで使用してきた塩ビ板に曲げセンサを取り付けた距離センサでは破損の可能性が高く、シャープ製の赤外線距離センサ（GP2S40J）を採用することにした。このセンサは、赤外線の発光部と受光部からなり、物体に反射した赤外線を受光することで距離が測定できる。感度調節のための可変抵抗を取り付けることで、0cm から 3cm までの距離が測定できるようにした（図 11）。

インタラクティブ東京やさきがけライブで展示したヘッドフォンでは、ヘッドフォンから12芯のシールド線が出ており、ケーブルの重量も重く、取り回しが困難であった。そこで、ヘッドフォンから出ているケーブルの本数を減らすため、電子コンパスと赤外線センサの出力をシリアル信号に変換し、USB ケーブルを通してパソコンに接続することにした。シリアルから USB へ変換する基盤は、ヘッドフォンのスピーカ部に内蔵した。これにより、必要なケーブルはステレオオーディオケーブルと USB ケーブルのみになり、6 芯シールドケーブルで足りるようになった。

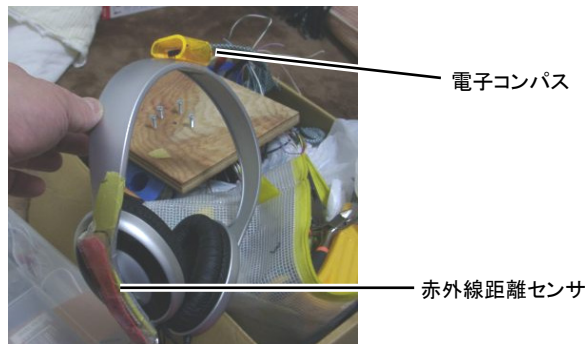


図 11: メディア芸術祭で展示予定のヘッドフォン(製作風景)

5. おわりに

本研究では、直感的な操作で複数パートのミキシングを可能とするヘッドフォンについて述べた。具体的には、ヘッドフォンに複数のセンサを搭載し、人間が音を聴くときの自然な動作を検出し、それを反映したミキシングの変更を行った。今後、センサや、増幅比を決める関数の改良を進めるとともに、ユーザビリティ等の評価を行っていききたい。

謝辞

試作機を製作した池月雄哉氏に感謝します。

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, August 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, March 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, October 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, October 2003.
- [5] <http://www.media.jst.go.jp/scholar/p16/06hamanaka.html>
- [6] Pachet, F. and Delerue, O. A Mixed 2D/3D Interface for Music Spatialization. First International Conference on Virtual Worlds, Lecture Notes in Computer Science (no. 1434), pp. 298-307, 1998.
- [7] Pachet, F. and Delerue, O. On-The-Fly Multi-Track Mixing. Proceedings of AES 109th Convention, Los Angeles, 2000. AES.
- [8] Olivier Warusfel, Gerhard Eckel "LISTEN - Augmenting everyday environments through interactive soundscapes", Proceedings of IEEE Workshop on VR for public consumption, pp. 268-275, 2004.
- [9] Jiann-Rong Wu, Cha-Dong Duh, Ming Ouhyoung, Jei-Tun Wu "Head motion and latency compensation on localization of 3D sound in virtual reality" Proceedings of the ACM symposium on Virtual reality software and technology 1997, pp. 15-20, 1997.
- [10] Camille Goudeseune, Hank Kaczmarek "Composing outdoor augmented-reality sound environments". Proceedings of the 2001 International Computer Music conference(ICMC2001), pp. 83-86, September 2001.
- [11] 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, 2004.
- [12] 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, March 2004.
- [13] イェンス ブラウエルト, 森本政之, 後藤敏幸, ". 空間音響", 鹿島出版会, 1986.
- [14] <http://www.cycling74.com/products/maxmsp>
- [15] 浜中 他: "デモンストレーション: 若手による研究紹介 II", 情報処理学会 音楽情報科学研究会 研究報告 2005-MUS-61-5, Vol. 2005, No. 82, pp. 27-33, August 2005.
- [16] <http://infusionsystems.com/catalog/index.php>
- [17] <http://interactivetokyo.jp/2005/>
- [18] <http://www.the-convention.co.jp/sakigake-live/>
- [19] <http://plaza.bunka.go.jp/festival/about/index.html>