

演奏の感じを伝える音楽演奏システム

中家武志* 平川正人†

*島根大学大学院 総合理工学研究科 数理・情報システム学専攻

†島根大学 総合理工学部 数理・情報システム学科

本論文では、4×4配列のスピーカ群を用いたテーブル型の音楽演奏システムを提案する。利用者が行うテーブル上での操作デバイスの動きに従って、操作位置からの鳴動及び操作デバイスの物理的反動がフィードバックされ、同時に映像も投影される。従来の電子的な楽器システムでは得ることが困難な「演奏している感じ」を利用者に伝達でき、コラボレーション演奏の支援に繋がる。

A Music Performance System Capable of Providing Play and Feel

Takeshi Nakaie* Masahito Hirakawa†

**Interdisciplinary Graduate School of Science and Engineering
Shimane University, Japan*

†*Interdisciplinary Faculty of Science and Engineering
Shimane University, Japan*

In this paper, we present a gesture-based music performance system at which an immersive sound field using a 16-channel sound table is set up. The system captures user's hand motion over the table, and projects computer-generated graphics onto the table. The user can lively interact with the system through both auditory and visual channels.

1. はじめに

音楽の演奏は、音楽を聴く以上に楽しく、そして刺激的な体験である。コンピュータ音楽の分野では、演奏者のためのシステムが様々な研究・開発されており、今日ではコンピュータを用いて手軽に音楽演奏システムを構築することができる。

例えば、オーディオの録音、編集、ミキシングのような音楽制作に必要な一連の作業をデジタル処理により行うことができる一体型システムであるDAW(Digital Audio Workstation)ソフトウェアを利用することで、プロフェッショナルなレコーディング・スタジオと遜色ない品質で音楽の制作や演奏が行える。そのうえ、仮想的な演奏者とジャム・

セッションのように合奏することも可能である[1]。これらシステムのインタフェースの多くは、テープ、磁気ヘッドレコーダ、スイッチ、つまみ、スライダなどを採用しており、古くからそのようなハードウェアに親しんでいるスタジオ・ミュージシャンにとっては十分に実用的である。

また、音響や映像のリアルタイム生成が容易になったことを背景に、聴衆が参加することで動的に状態が変移するような、インスタレーションの形態をとる芸術作品なども多くみられるようになった。例えばinstant city [2] は聴衆らが協力して半透明のブロックを積み上げていくことにより次第に音楽が構築されていく作品である。これは広義に解釈すれば聴衆同士による音楽の演奏とも言える。

このように、コンピュータ・ベースの音楽演奏が広く一般的になった一方で、それらに初めて触れる人々の多くは、その独特の演奏感覚に戸惑いを感じている。

例えば、初心者を対象としたシステムに Drop Spin Fade [3] や Beat Jigsaw [4] などがある。これらは画面上でグラフィカルに、もしくは物理的に、音楽を構成する「要素」を配置して即興演奏を行うことができる。演奏者は要素の色や形、位置などから演奏結果を変化させようと試みるが、視覚的な情報と音を直感的に関連づけるのは困難な場合が多い。

対して、管楽器、弦楽器、打楽器のようなアコースティックな楽器においては、演奏者の息づかいや指をはじく強さのような、物理的な力の加え方が音の大きさや音色に影響を与える。また、音は通常、楽器本体から発生するので演奏者は楽器の振動を直に感じ取っている。この振動という物理的なフィードバックを通して楽器と対話を重ねることで、演奏者は自身の力で音を出している感覚、すなわち「演奏している感じ」を得ることができる。これは楽器を上手にコントロールするために欠かせない要素である。

また、多くの場合、コンピュータ・ベースの音楽は一組のスピーカを通して提示される。このスピーカは演奏者が一人であれば十分に役目を果たすかもしれないが、複数人が協調して演奏する際には、全員の演奏パートが重なって再生される状況が発生する。各演奏者が自分の演奏に対して適切なコントロールを維持するには、演奏者ごとの空間的な音場が判る仕組みが必要であろう。

総じて、コンピュータ・ベースの音楽システムには「演奏している感じ」を得られる機構が欠けている。これらに適度なフィードバックを実現すれば、初心者が演奏する際の大きな助けになると考える。

本論文では、コンピュータ・ベースの音楽演奏システムを提案する。システムは16基のスピーカが4x4レイアウトに置かれたテーブルが中心となる。全てのスピーカは制御され、テーブル上の任意の位置で音を再生することができるようになっている。以降、このテーブルをSoundTableと呼ぶ。

また、システム上部に設置した2基のカメラによって利用者の3次元空間内での操作を認識すると同時に、ビデオプロジェクトを通してSoundTable上面にコンピュータ生成映像を投影する。

これら聴覚と視覚のフィードバックにより、従来のコンピュータ・ベースの演奏システムでは得ることが困難であった「演奏している感じ」を利用者に伝達する。

本論文の構成は以下の通りである。まず初めに、第2章で関連研究を紹介する。続いて、第3章ではシステムの概要を示す。第4章ではソフトウェアの仕組みを説明する。第5章ではシステムに対する評価実験の結果を示す。第6章ではシステムを利用したインタラクションの例を示し、最後に第7章で本論文をまとめる。

2. 関連研究

従来のDAWソフトウェアやヤマハのサイレントギターのような電気楽器とは異なる、新しい種類の音楽演奏システムをいくつか紹介する。

SONASPHERE [5] はコンピュータ画面上で利用できる新しい種類の音楽演奏システムのひとつである。SONASPHEREは、音のサンプルやエフェクト、ミキサーなどを3次元空間上の球体オブジェクトとして表現し、これらを信号の流れに沿って接続することで音楽が構成される。興味深いのは、オブジェクト同士が物理的な属性を持っており、引き合って衝突したり、落下して跳ね返ったりする。オブジェクトの動きと音は密接に関係しており、演奏に無限の変化をもたらすと共に、その様子を可視化し聴衆に対して提示することもできる。

SONASPHEREは通常のコンピュータを利用するが、[6]-[8]にあるようなテーブル型のものや[9]-[11]のような身振りを利用するものなど、身体的なインタラクションを取り入れた音楽演奏システムも提案されている。我々は特にテーブル型が有用であると考えている。

Audiopad [6] やreactable [7] は、テーブル面に置かれた幾つかの樹脂製のオブジェクトを操作する音楽演奏システムである。システムはオブジェクトの位置を認識し、それらの動きから音を生成する。またオブジェクトの操作に応じたグラフィカルな投影がテーブル面へ行われる。

Block Jam [8] は、テーブル面でブロックを配置するという点で、Audiopadやreactableと類似している。ブロックはクリック操作とジェスチャ操作に対応したディスプレイを持ち、互いに繋げることができる。ディスプレイをクリックするとブロック間を飛び移るように音が再生される。

本論文で提案するシステムはAudiopad, reactable, およびBlock Jamのようなテーブル型のインタフェースを採用している。異なるのは、利用者がテーブル上の3次元空間内でオブジェクトの操作が可能である。また、マルチ・チャンネルのスピーカを使用することで空間的な音場を発生させることができ、利用者はあたかも手元に楽器があるように、操作した場所からの音を聴くことができる。これは複数人が協調して演奏する際に、各演奏者が自分の演奏に対して適切なコントロールを維持するのに役立つ。

マルチ・チャンネルのスピーカを使用するという考え方は[12]にも見られるが、このシステムはCAVEのようなマルチ・スクリーン映像表示のために設計されているため、かなり複雑で高価である。

3. システム概要

本論文では、テーブル型音楽演奏システムを提案する。システムの構成を図1に示す。システムはSoundTableを中心として、2基のUSBカメラ、プロジェクタ、およびPCによって構成される。

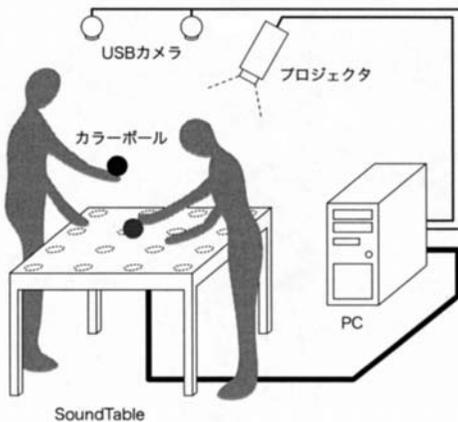


図1. システム構成

SoundTableは、図2のように16基のスピーカ群を4x4配列で内蔵している。

テーブル面の大きさは90cm四方で、高さは73cmである。その上面には図3のように白色の保護カバーが被せてあり、コンピュータ生成映像を投影するためのスクリーンを兼ねている。

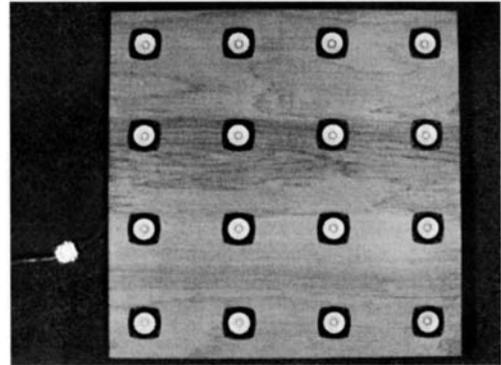


図2. SoundTable

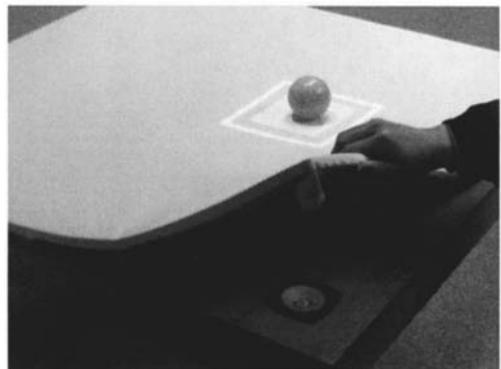


図3. スクリーンを被せたSoundTable

利用者は、テーブル上でカラーボールを操作して音楽を演奏する。カラーボールの動きは、2基のUSBカメラによって3次元的に追跡される。マルチ・チャンネルの出力のために、2基の8チャンネル出力サウンドカードをPCに内蔵しSoundTableと接続した。プロジェクタは操作に応じたコンピュータ生成映像をテーブル面に投影する。

4. ソフトウェア

PC上のソフトウェアは大きく2つのモジュールで構成される。1つは利用者の入力操作を処理するモジュール、もう1つは音や映像の出力を制御するモジュールである。

4.1. 操作の認識

既存のテーブル型音楽演奏システムの多くは、操作オブジェクトがテーブル面上に配置される。これはステージで指示を出す音楽家のように、よりダイ

ナミックな方法で即座にアイデアを形にできるからである。反対に、より複雑な演奏を行うためには、オブジェクトや操作の種類を増やすことになるが、これは同時にユーザビリティを低下させる可能性がある。

本研究では初心者でも利用できる音楽演奏システムの実現を考えるので、ジェスチャの種類を限定した。これについては第6章で詳細に述べる。

操作の追跡についてはいくつかの手法を検討した。そのひとつに、格子状に敷設されたアンテナと静電容量計測技術を応用したSmartSkinという手法がある[13]。これは利用者の負担が少ない直感的な操作方法を実現するが、検出可能な範囲がテーブル表面のみに制限されてしまう。高さ情報を取得するために超音波センサの利用も検討したが、複数人に対する計測が困難である。

結果として、今回はカラーボール（直径7cm）に対してコンピュータビジョンベースのジェスチャ認識を行い、手の動きを追跡することにした。ボールの色を区別することで、利用者ごとの操作を容易に識別できる。

カラーボールはSoundTableの上部に設置された2基のUSBカメラで捕捉する。色域抽出によりボールの領域を検出し、三角測量法によって空間座標を算出している。

4.2. 音の定位

SoundTableには16基のスピーカが内蔵されている。隣接する2基のスピーカの間の距離は24cmである。これは音楽演奏に関わる手やボールの大きさと比較してもかなり粗い。

音の再生される位置をより精細にコントロールするために、音場計算モジュールを実装する。より多くのスピーカを利用する方法も考えられるが、システムが大掛かりになり実用的でなくなることを危惧した。

音場計算には正弦波と余弦波によるクロスフェードを応用する。一組のスピーカ間において音が聴こえる方向を調整するには、各チャンネルの出力レベルに正弦波と余弦波を掛け合わせて行うのが一般的である（図4）。これを2次元に拡張し、テーブル面の任意の位置で音の再生を実現した。

音は、PCに内蔵した2枚の8チャンネルサウンドカードより出力され、16基のスピーカを並列に制御可能である。同時に再生できる音の最大数は16であ

り、これは現時点の実装におけるハードウェア性能の制限によるものである。

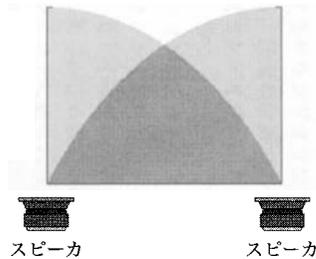


図4. 正弦波と余弦波を利用したクロスフェード

5. 評価

システムの有効性を評価するために、2つの実験を行った。1つは音の定位性能についての実験、もう1つはジェスチャの認識性能についての実験である。

5.1. 音の定位に関する実験

音の定位に関する性能を評価するため、音場計算モジュールで指定した位置と利用者が知覚した音の位置の誤差を分析した。さらに、再生位置が移動する場合の滑らかさについても検証した。実験は10人の学生（男性7人、女性3人）で行った。実施にあたっては、音の再生位置をプログラムできるテストツールを作成し、被験者全員が同一の環境になるよう配慮した。

被験者はシステムの音を聴き、音の数とそれぞれが聴こえた位置を回答する。被験者には、SoundTableの前にまっすぐ立ってもらい、音の位置はスピーカの取り付けられた位置に依らないことを予め伝えてある。再生音には5種類の異なる楽器（ピアノ、ベース、シンバル、ドラム、シンセサイザ）による8秒間のシーケンスを用意した。

システムが再生した音の位置と、被験者が答えた音の位置の誤差を集計した結果を図5及び図6に示す。図5は音の位置が静止している場合の分布であり、合計290回の結果を集計したものである。水平な横軸と縦軸はそれぞれ横方向と縦方向の誤差の値を示している。ここで、誤差の値1.0は隣り合うスピーカ間の距離と等しい。横方向の誤差の平均値は0.19、縦方向の誤差の平均値は0.41であった。

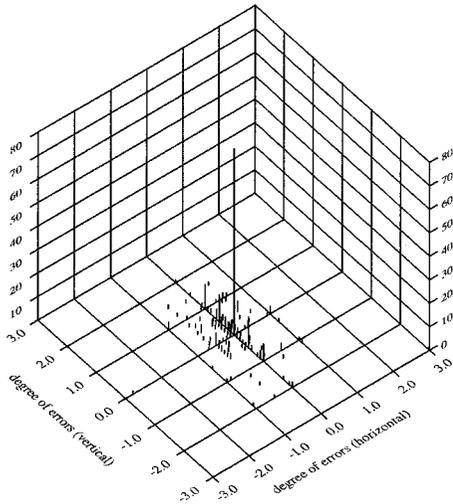


図5. 音の位置が静止している場合の誤差の分布

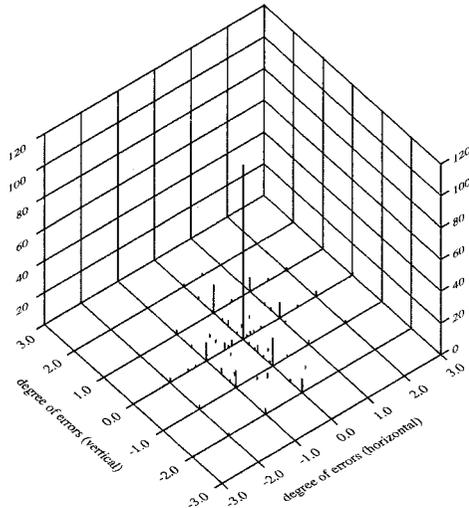
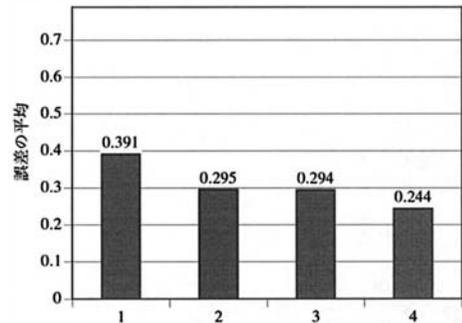


図6 音の位置が移動している場合の誤差の分布

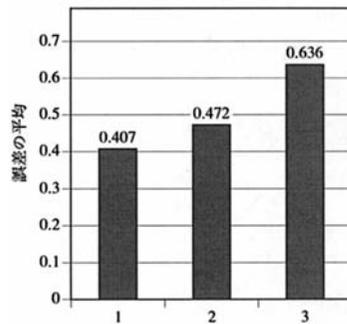
図6は音の位置がテーブル上を移動している場合の分布であり、合計250回の結果を集計したものである。横方向の誤差の平均値は0.41、縦方向の誤差の平均値は0.56であった。

平均して、横方向の誤差よりも縦方向の誤差のほうが大きいことが分かる。この差は人間の頭部の構造に由来するものと考えられる。また、音の位置が移動している場合は、音の位置が静止している場合と比べて誤差が大きいことが分かる。

続いて、同時に再生される音の数と、誤差の平均値の関係を図7に示す。音の位置が静止している場合は、誤差の平均値に大きな変化は見られないが、音の位置が移動している場合は、音の数が増えるに従い、誤差の平均値が著しく増加している。今後、より多くの音を再生した場合についての追加実験が必要になるであろう。



同時に再生される音の数 (音の位置が静止している場合)



同時に再生される音の数 (音の位置が移動している場合)

図7. 再生される音の数と誤差の平均値の関係

5.2. ジェスチャ認識に関する実験

空間的な位置座標の取得に関する性能を評価するため、高さ方向の認識精度についての実験を行った。

90cm四方のテーブル面を10cm間隔の格子状に区切り、その格子点上で5cmずつ50cmまで高さを変更しながらカラーボールの認識を行った。

高さの測定誤差のヒストグラムを図8に示す。横軸は、カラーボールの高さとシステムが認識した高さの差を表す。縦軸は各誤差の度数を表す。この測定実験全体における誤差の平均値は-5.9cmであった。

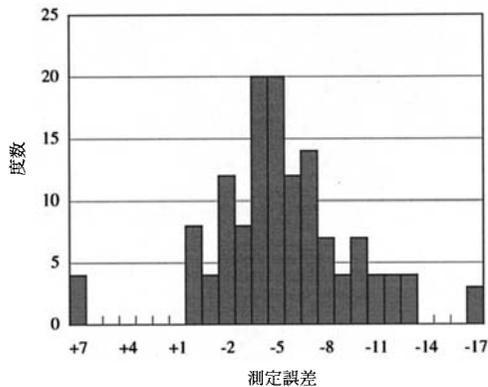


図8. 高さの測定誤差

測定結果をテーブル面の領域ごとに分析したものが図9である。グラフの横軸はカラーボールの実際の高さを示し、縦軸はシステムによって判定された高さを示す。グラフの凡例に図示した通り、領域は外周に沿って4つに分割した。

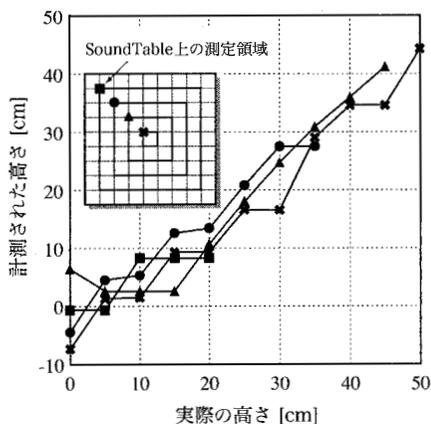


図9. カラーボールの実際の高さとシステムが判定した高さの関係

グラフより、システムはカラーボールの高さを実際より低く見積もる傾向があると分かる。システム内の基準値を調整することにより、誤差を抑えることができると考えられる。

6. システムの利用

3次元空間操作ならびに平面空間上で音を再生できる特徴を利用した具体的な対話系を実装した。

6.1. 音の操作

音を仮想的なオブジェクトとして捉え（以下、サウンドオブジェクトと呼ぶ）、これに操作を加えることができるようにする。このような機能を実現するにあたり、サウンドオブジェクトを図10のように四角形としてSoundTable上に視覚化する。



図10. サウンドオブジェクトの操作

サウンドオブジェクトに以下の3つの基本的なジェスチャ操作を定義する（図11）。

- drag and drop
- turn over
- dive

drag and dropは、選択したサウンドオブジェクトを異なる位置へ移動させる操作である。サウンドオブジェクトにボールを被せると選択が指示されたときサウンドオブジェクトを、そのままボールを平行移動することで対応するサウンドオブジェクトを移動させることができる。任意の位置でボールを取りさえすれば、その位置にサウンドオブジェクトが固定される。

turn overは、サウンドオブジェクトにメッセージを送る操作である。サウンドオブジェクトの表面を撫でるように右から左、もしくは左から右へ手を動かすと、それらの動作にあらかじめ設定されたメッセージがサウンドオブジェクトに送られる。今回の実装では再生される音の種類を変更する。またサウンドオブジェクトが裏返る視覚効果によって利用者はメッセージが送られたことがわかるようになっている。

diveは、サウンドオブジェクトの属性値を変更する操作である。今回の実装ではサウンドオブジェクトの音量を調節するようにした。利用者がサウンド

オブジェクトに手を重ねてしばらく待つとサウンドオブジェクトの外側に枠が表示され、dive操作モードに入ったことがわかる。そのまま、枠内でのボールの上げ下げによって音量が調節される。また音量に比例してオブジェクト表示の大きさが変化する。dive操作モードを終了するには枠外へ手を外す。

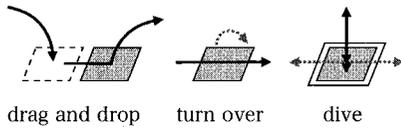


図11. 3種類の基本のジェスチャ



図12. ジェスチャ操作の例 (dive)

当然ながら、操作と音の関係づけはこれだけではない。例えばサウンドオブジェクトの位置と音程を対応づけてもよい。その場合、ボールを転がすことで音の連続的変化を楽しむことができる。

6.2. ハードウェア・センサの併用

前節に説明したインタラクションの他に、別途ハードウェアセンサを併用することで、より多彩なインタラクションの実現を図った。この目的のために任天堂のWiiRemoteコントローラを使用した。これは家庭用ビデオゲーム向けの製品であるが、Bluetooth接続の安価な傾斜/加速度センサとして使用することができる[14]。カラーボールと組み合わせた実装を図13に示す。

このデバイスを3次元空間内の自由な位置に持っていき、次第に傾けると、その直下にサウンドオブジェクトが現れる。デバイスが高くにある場合ほど音は大きく、デバイスを大きく傾けるほどオブジェは短い間隔で現れる。傾けた状態でデバイスを平行

移動させると、ちょうど水滴が落ちるような効果が得られる(図14)。

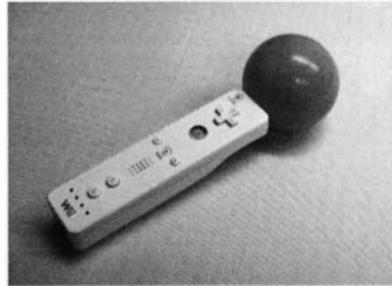


図13. 傾斜センサを併用したデバイス

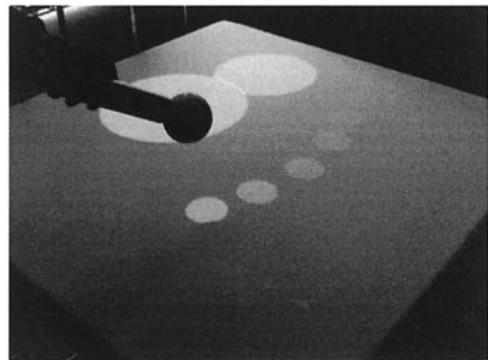


図14. MusicDrops

この他にも、加速度の変化を利用してデバイスを振り下ろす動作を検出した。利用者はあたかもテーブル上の仮想的なドラムを叩くように音を作り出すことができる。また、叩くと同時にWiiRemoteに内蔵された振動機構を動作させることで、叩いた際の物理的な反動が再現されるようにした。

7. 結論

本論文では、3次元空間上でのジェスチャ動作によって直感的に操作できる音楽演奏システムを提案した。

システムの中核としてSoundTableという16基のスピーカを内蔵したテーブルがあり、スピーカの位置に関わらず、テーブル上の任意の位置から音が発生しているような制御を実現した。加えて、テーブル上の3次元空間内での利用者の操作に応じたコンピュータ生成映像をテーブル面へ投影することで、

利用者は聴覚と視覚のフィードバックを伴った「演奏している感じ」を得ることができた。

ジェスチャ操作の獲得に際してはカラーボールを用い、具体的に3種類のジェスチャ操作を実現した。また、傾斜/加速度センサデバイスを併用して、より楽器の演奏に近いインタラクションの実現を図った。

評価実験の結果、SoundTable上に定位する音の聴こえ方については概ね良好であったと言えるが、音楽演奏システムとしてのインタラクションに関しては、まだ限定的な実装であり、今後の更なる検討が必要である。

謝辞

本システムの実装と評価実験にあたり協力していただいた平川研究室の板倉稔氏（現、NTTデータカスタマーサービス(株)）に謝意を表する。

参考文献

- [1] M. Hamanaka, M. Goto, H. Asoh, N. Otsu, "A Learning-Based Jam Session System that Imitates a Player's Personality Model," Proc., Int'l Joint Conf. on Artificial Intelligence (IJCAI 2003), pp.51-58, 2003.
- [2] <http://www.instantcity.ch/>
- [3] <http://www.chrisoshea.org/lab/drop-spin-fade/>
- [4] <http://www.participationsystems.com/jigsaw.htm>
- [5] <http://www.sonosphere.com/>
- [6] J. Patten, B. Recht, and H. Ishii, "Interaction Techniques for Musical Performance with Tabletop Tangible Interfaces," Proc., ACM Conf. on Advances in Computer Entertainment Technology, 2006.
- [7] S. Jorda, G. Geiger, M. Alonso and M. Kaltenbrunner, "The reacTable: Exploring the Synergy between Live Music Performance and Tabletop Tangible Interfaces," Proc., ACM Conf. on Expressive Character of Interaction, pp.139-146, 2007.
- [8] H. Newton-Dunn, H. Nakano, and J. Gibson, "Block Jam," Proc., ACM SIGGRAPH2002 Conf., p.67, 2002.
- [9] H. H. S. Ip, B. Kwong, and K. C. K. Law, "BodyMusic: A Novel Framework Design for Body-driven Music Composition," Proc., ACM Int'l Conf. on Advances in Computer Entertainment Technology, pp.342-345, 2005.
- [10] J. Zigelbaum, A. Millner, B. Desai, and H. Ishii, "BodyBeats: Whole-Body, Musical Interfaces for Children," Proc., ACM CHI Conf., pp.1595-1600, 2006.
- [11] S. Mann, "The AndantePhone: A Musical Instrument that You Play by Simply Walking," Proc., ACM Multimedia Conf., pp.181-184, 2006.
- [12] T. Ogi, T. Kayahara, M. Kato, H. Asayama, and M. Hirose, "Immersive Sound Field Simulation in Multi-screen Projection Displays," Proc., Eurographics Workshop on Virtual Environments, pp.135-142, 2003.
- [13] J. Rekimoto, "SmartSkin: An Infrastructure for Freehand Manipulation on Interactive Surfaces," Proc., ACM CHI2002, pp.113-120, 2002.
- [14] <http://www.wiili.org/>