

SoundPond：音を可視化し直感的に操作できる環境の提案

SoundPond：Making Sound Visible and Intuitively Manipulable

村田 直紀† 外村 佳伸‡
Naoki Murata Yoshinobu Tonomura

1. はじめに

現実世界では人は様々なモノを扱っている。道具であればその用途はもちろん、道具を何かに見立てる、または複数配置することで何かを表現したり、様々な別のモノと組み合わせることができる。しかしコンピュータ上で扱う情報はシステムのアイコンなどを除くと、モノとして触って扱うことはなくなり、異なる種類の情報を組み合わせたり、相互に干渉させることもない。そこで情報にも形を持たせれば、情報を物のように扱い、情報と情報の様々な新しい関わりやインタラクションが生まれるのではないかと考えた。

その中でも本研究では現実世界においても一度発せられると消えていき、目に見えず、触れられない音に着目し、発音された音声が形を持ち目に見える形で残り続け、ほかの音や情報と干渉し、置かれた場のルールに従って様々な動作する環境 SoundPond を提案する。

今回基本コンセプトの検証のためプロトタイプシステムを構築し、実験を行った。

2. SoundPond の提案

従来、音声情報は録音してファイルにしたものを再生したり、編集を行う。また可視化として時間領域上で波形として表す方法、または周波数領域上でスペクトルとして表す方法が多い。またどこでどのくらいの大きさの音が鳴っているかを空間分布を可視化する音響ホログラフィ[1]などもある。これらに対し、本論文では信号や波形としてでなく、あたかも音をモノとして扱うことができるように形を与える SoundPond を提案する。

SoundPond では体感的かつ直感的な操作を目的としているため、利用者にボタンやメニューなどのコマンド選択方式の操作はさせないことを基本方針に、以下のコンセプトで設計した。(図 1)

1. 音に形を与え仮想的にモノ化する
2. 単体の音オブジェクトを体感的に操作できる
3. 複数の音オブジェクト間の相互作用と、さらに人とのインタラクションの実現
4. 場に特性を与え、音オブジェクトに影響を与える

システムの応用例としては、パーソナル用途として、音楽制作環境、音楽再生環境、音楽演奏パフォーマンスの道具といった創造的な利用が可能と考えられる。

また大画面ディスプレイを用いた場面として、複数人での利用による、メッセージのやり取りや、様々な情報を残せる落書き掲示板、子供の遊び場的な利用が可能と考えられる。

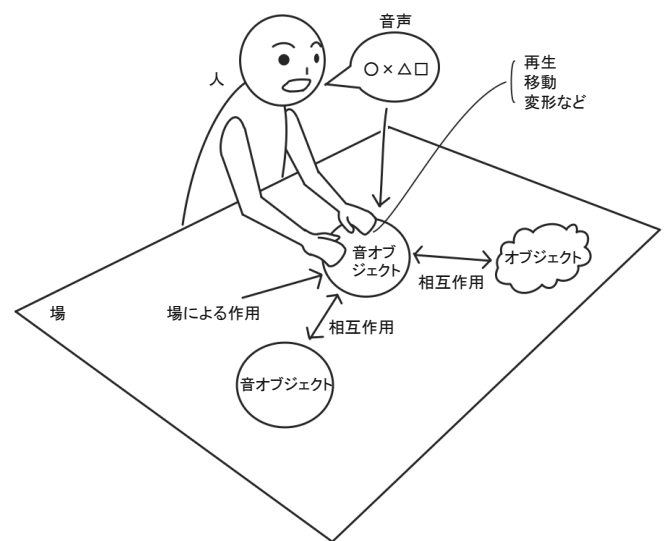


図 1. SoundPond のコンセプトのイメージ図

2.1 課題

上記の設計方針とコンセプトに基づいて、システムを構築する上での主な課題を以下に示す。

- (1) 入力音の自動区分化 (音区間検出)
- (2) 音の基本要素の可視化表現
- (3) 単一の音オブジェクトの体感的な操作
- (4) 複数のモノ間や場との相互作用
- (5) (3), (4)に伴う音の変化の処理

2.2 課題へのアプローチ

- (1) 入力音の自動区分化

コマンドによって明示的に録音することなく、自動的に発生した音が形をもって残るには入力音から、連続音からなるかたまりを検出する必要がある。プロトタイプシステムでは、主に利用する人の声を素材として用いることを考え、シンプルに音量によって有声区間を検知し、その長さによって扱うべき音かどうかを判定することとした。

†龍谷大学大学院, Graduate school, Ryukoku University.

‡龍谷大学, Ryukoku University

(2) 音の基本要素の可視化表現

音のオブジェクトが音の基本的な特徴を直感的に表現するために、音の大きさを縦の長さ、音の長さを横の長さとする楕円形とし、音の高さを色で表すこととした。

今回のプロトタイプシステムは図 2 のように音を可視化した。

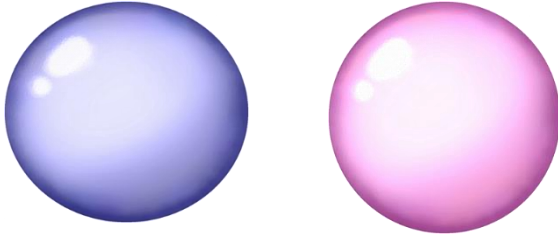


図 2. 131hz の音 (左, 青) と 294hz の音 (右, ピンク)

(3) 単一の音オブジェクトの体感的な操作

タッチ操作のみに操作を限定し、シングルタッチによる操作で音の再生, 移動, また場に絵を描くなどを実現し, マルチタッチ操作による音オブジェクトの拡大縮小などを実現する。また, それらによって音に効果が与えられ, 音が変化する。

(4) 複数のモノ間の相互作用

場に様々な特性を与えることで音オブジェクトに影響を与えることができるが, プロトタイプシステムでは場上の音オブジェクトの位置座標の上下によって再生音の高さが変化, 左右によって再生音の左右の定位が変化するものとした。また音オブジェクトと他の音オブジェクトや描いた絵との衝突, 反射を考えて実装した。衝突の起きた時は, 衝突位置の条件により音が再生されるようにした。

本研究ではこのモノの組み合わせによる相互作用の実現が重要と考えており, 今後の実装の可能性として, 衝突した際に条件に応じて反射以外にすり抜ける, くっつく, 変形するといった動作を検討している。また近くにあるモノ同士の距離や配置パターンによって自律的な移動の方向やパターン, あるいは形の変化, といった作用も検討している。

(5) 音の処理

プロトタイプシステムでの音の処理はピッチシフト, タイムストレッチ, 音量の変化, 左右のバランスの変化, 逆再生のみであるが, 残響や反響の付加, フィルターによる音質の変化などの様々な効果が直感的な形で起こることを検討している。

3. プロトタイプシステム

提案した環境の基本機能をもったプロトタイプシステムを構築し, 提案コンセプトの検証を行った。

3.1 ハードウェア構成

プロトタイプシステムは, デスクトップタイプの利用として 27 インチのマルチタッチディスプレイを使用したものと, 複数人利用のため大画面ディスプレイを用いた壁型のもを構築した。

実際のシステムの外観を図 3, 4 に示す。図 4 の環境は大画面を用いた汎用の実験環境であるアンビエントウォールを用いている[2]。

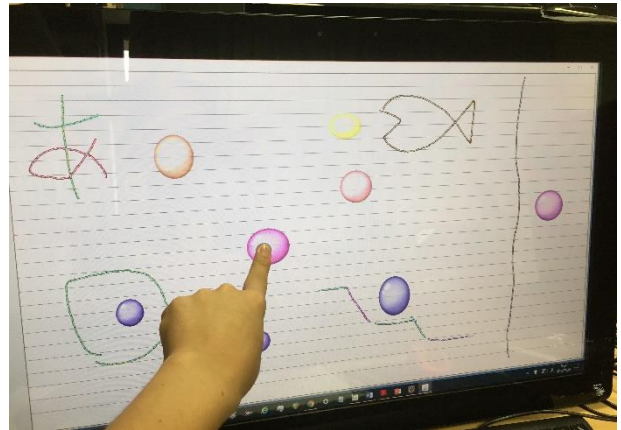


図 3. システム外観



図 4. システム外観 2

3.2 システム構成

システム全体の構成図を図 5, ソフトウェアの構成図を図 6 に示す。

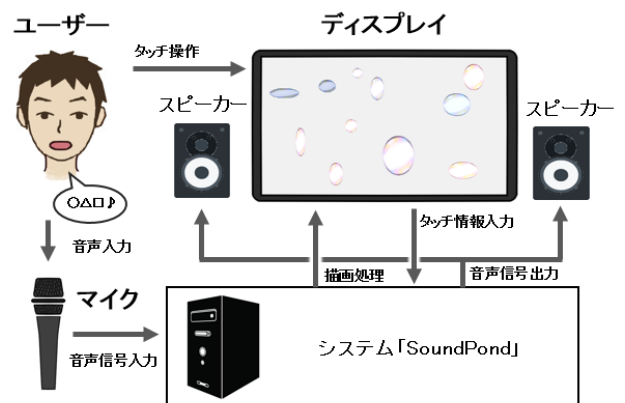


図 5. システムの構成図

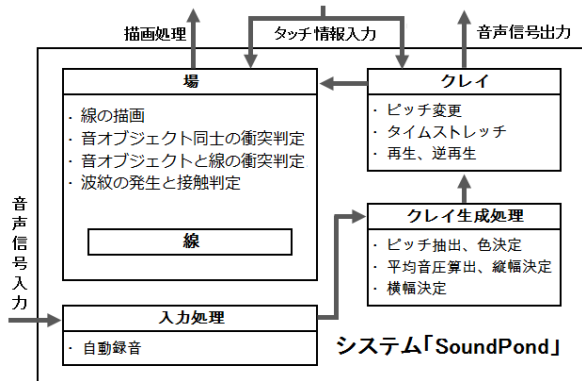


図 6. ソフトウェア構成図

3.3 プロトタイプシステムの機能

プロトタイプシステムは以下の機能を有している。

- ・声（音）を出すとそれを可視化したオブジェクトを生成する。
- ・オブジェクトを指でタッチして音を再生することができる。
- ・オブジェクトを指でタッチしスライドしてオブジェクトを移動することができる。
- ・オブジェクトを指ではじいて速度を持たせて移動させることができる。
- ・場に指で自由曲線を引き、絵や字などが描ける。
- ・タッチ長押しで音を逆再生させることができる。
- ・場をタッチして波紋を起しそれに触れたオブジェクトを再生させることができる。
- ・音の高さによってオブジェクトの色が決まる。
- ・音の大きさによってオブジェクトの縦幅が決まる。
- ・音の時間の長さによってオブジェクトの横幅が決まる。
- ・オブジェクト同士が衝突することで音が再生され、オブジェクトは反射する。
- ・描いた線とオブジェクトが衝突することで音が再生され、オブジェクトは反射する。
- ・場の左右の位置と左右の音の定位が対応している。
- ・場の上下の位置によって音の高さが増える。

4. ソフトウェア処理

4.1 録音の有意区間検出

よって音をモノとして扱うために有意な音と雑音を区別し取捨選択し、連続したかたまり単位で扱う必要がある。

プロトタイプシステムでは単純に音量で有意区間を検出している。

最初に 5000ms 音を録音し、恒常的に発生しているノイズとする。50ms ごとの RMS 値をデシベルで記録し、その中央値の音量を閾値の基準とした。

入力中は常に RMS 値を取り続け、予備実験により閾値 +12db 以上の音が 250ms 以上続き、閾値+6db 以下が 500ms 続くまでを有意区間とした。

4.2 音の高さの推定

基本的にマイクから録音する音声は複合音であり、純音であることはない。そこで音の高さを求めるには基本周波数を求める必要がある。検出した基本周波数がそのままオブジェクトの色表現に対応するので、できる限り誤検出の少ない手法を用いる必要がある。

プロトタイプシステムでは自己相関関数に NSD (Normalized Square Difference) 関数というものを適応し、-1~1 に正規化する MPM アルゴリズム[3]によって基本周波数を求めた。信号波形 $v(t)$ についての自己相関関数 $A(j)$ は以下の式にて定義される。

$$A(j) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^{\infty} v(t) \cdot v(t+j) dt$$

これにより信号波形 $v(t)$ とそれを j ずらした信号波形 $v(t+j)$ の近似の度合いが求まる。音の波形は 1 周期ずらすと近似するのでこの式によって推定できる周期から基本周波数を求めることができる。音の高さは短時間 (4.3 のブロック単位) で求めた。

4.3 音の効果

本システムでは音の高さと長さを変化させることができる。通常音は再生速度を変えると高さも変化してしまう。

それを回避するために再生速度を変えずに音の長さを変える手法を利用した[4]。

本システムでは音を 1028 サンプルごとに分割し 1 ブロックとし、ブロックを抜去、またコピーして挿入することで音の長さを変化させた。またブロックを増減させた部分では音の信号の連続性が損なわれノイズが発生する。それを避けるためにブロックとブロックのつなぎ目をクロスフェードさせた。またクロスフェードの位置を前後のブロックの相関をとり、最大になる位置とすることでクロスフェード部分に音の重なりや不自然な音量変化が極力起きないようにしている。

また、音の高さを変えるには再生速度を変化させ、その際に変化する音の長さの分だけ上記の手法で音の長さ調節することで音の長さを変化させずに音の高さを変化させることができる。

4.4 衝突の判定

音に形を持たせることにより衝突するという物理現象が起こることを考えた。本システムでは描いた絵との衝突や音のオブジェクト同士の衝突を考えた。

正円同士の衝突は比較的単純に判定できるが、今回は楕円や自由曲線を扱うため、視覚的にも不自然でない衝突、反射を実現する必要がある。次の 2 つを行った。

音オブジェクトは楕円であり、利用者の描く絵は自由曲線である。自由曲線の衝突判定は非常に難しく、計算量も

多くなる。そこで本システムでは線を長さ 20 ピクセルの直線の集まりに分割し、直線と楕円の衝突として考えた。

楕円と線の衝突判定は図 7 のように楕円を円に直し座標を回転させて考える。座標の回転は以下の公式にて求められる。

$$X' = (\cos \theta)X - (\sin \theta)Y$$

$$Y' = (\sin \theta)X + (\cos \theta)Y$$

また、反射の方向は回転させたベクトルの y 方向の値を反転させ、逆回転させた。また本システムでは回転は考慮していない。

音のオブジェクト同士の衝突は楕円の衝突と反射を見る必要がある。座標 (ax, ay) にある幅 aw 、高さ ah の楕円 A と座標 (bx, by) にある幅 bw 、高さ bh の楕円 B の衝突を調べるには座標 (ax, ay) にある幅 $aw+bw$ 、高さ $ah+bh$ の楕円と座標 (bx, by) の接触を考える。さらに楕円が円になるように縦横比を変換することで円の半径と 2 点間の距離の比較で接触が判定できる。

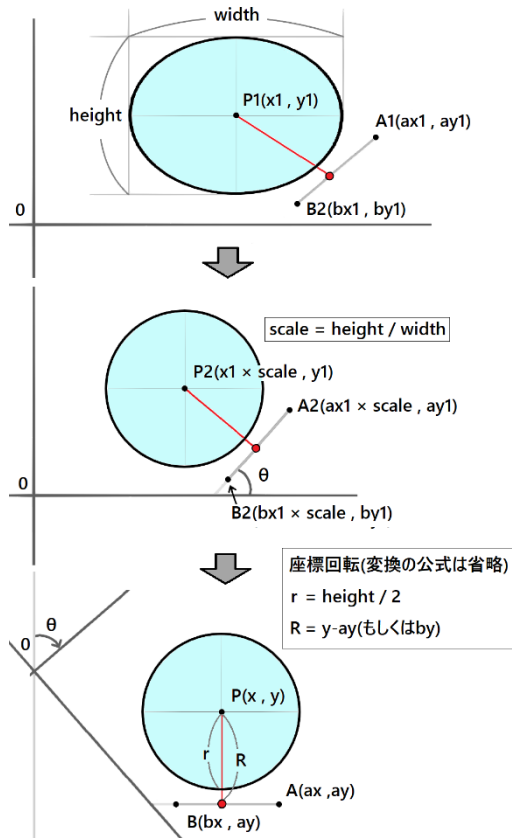


図 7. 楕円と直線の衝突判定のイメージ図

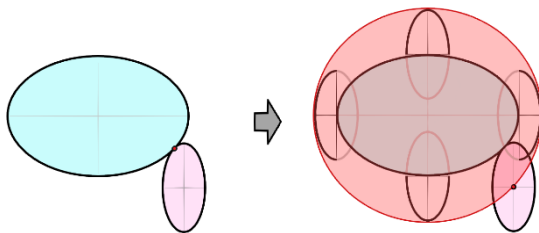


図 8. 楕円同士の衝突判定のイメージ図

5. 実験

作成したシステムの直感性を評価するために 12 人に対して主観評価実験を行った。その内容と結果について以下に述べる。

5.1 操作に関する実験

システムの操作が直感的でわかりやすいかどうか、システムの機能やオブジェクトや場の表現の意味について直感的に理解できるかどうかを調べる実験を行った。

システムの操作の調査項目を以下に示す。

- 項目 1. オブジェクトを指でスライドして移動
- 項目 2. オブジェクトを指ではじいて動かす
- 項目 3. オブジェクトをタッチして音を再生
- 項目 4. 場に指で自由曲線を引く
- 項目 5. タッチ長押しで音を逆再生させる操作
- 項目 6. 場をタッチして波紋を起こす操作

次に、可視化表現、オブジェクト同士の相互作用、場の特性に関する調査項目を以下に示す。

- 項目 1. オブジェクトの色と音の高さ
- 項目 2. オブジェクトの縦幅と音の大きさ
- 項目 3. オブジェクトの横幅と音の時間の長さ
- 項目 4. オブジェクト同士の衝突による音の再生
- 項目 5. 線とオブジェクトの衝突による音の再生
- 項目 6. 波紋とオブジェクト接触による音の再生
- 項目 7. 場の左右の位置と左右の音の定位
- 項目 8. 場の上下の位置と音の高さの変化

被験者に詳しいシステムの機能を説明せずに 3 分間システムを使用してもらい、その中で上記の項目について理解できたか調べた。その実験の結果を図 9 に示す。

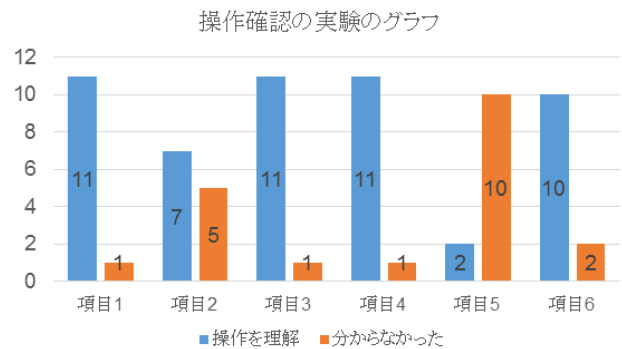


図 9. 実験 1 の結果

この結果から項目 5 以外の操作は理解できた方が過半数であり、直感で行える操作であると言えるだろう。

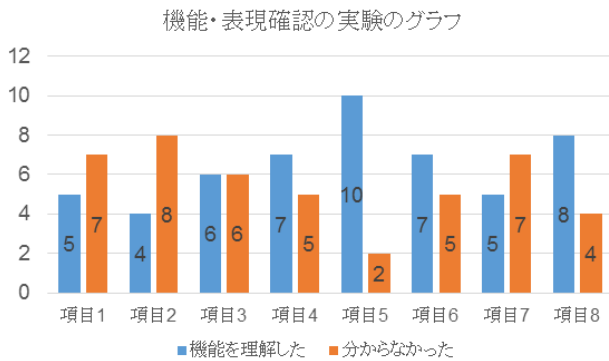


図 10. 実験 1 の結果 2

この結果をみると、環境としての機能は理解できた人とできなかった人がおおよそ半々であり、十分に直感的とは言えない。考察では主にこの原因を推測する。

6. 考察

実験 1 の図 9 の結果から項目 2 の結果が理解できなかった人が比較的多いことが分かる。これは素早くはじかなければオブジェクトが動き続けないためであると考えられる。指が離れる直前の指のスライドの速度が一定以上であればはじいたとみなしているので予備実験を行い適切な閾値を設定すればより直感的になると考えられる。また項目 5 の機能はほとんどの人が分からなかった。このことから長押しを伴う操作が直感的でないと考えられる。

次に実験 1 の図 10 の結果から、音の高さによる音オブジェクトの色の变化を理解できない人が多かったが、これは実験の際に色々な高さの音を出してみるということを行わなかった人が多く、色の变化があまり出なかったためだと考えられ、実験内容を見直す必要があると考えられる。色々な高さの音を出すことを試した人は全員が音の高さと音オブジェクトの色の関係に気付いていた。

音の大きさと音オブジェクトの縦幅の関係はほとんどの人が気付かなかった。これはマイクの指向性が狭くマイクとの距離やマイクの向きによってマイクに入力される音量が大きく変わってしまうことが考えられる。解決策としては指向性の広い環境音などを録音することに特化した専用のマイクを用いることで簡単に解決できると思われる。

場にある音オブジェクトの左右の位置によって左右のスピーカーからの出力バランスが変わることは多くの人が気付いていなかった。これはスピーカーの左右の配置のレンジが狭いためだと考えられる。実験環境としてもう少し左右の間隔をあけてスピーカーを配置すれば改善すると考えられる。

7. 結論

本研究では音を目で見て触ることができる環境を目指し、音を自動でかたまりごとに分割して録音し、指で触れることで音に様々な変化を加えることができるシステム SoundPond を提案し、その操作の直感性、わかりやすさを検証する実験を行った。

実験から、直感的でわかりやすい操作感に加え、音が目に見えて直接手で触れているという感覚が得られることを示すことができた。

一方、情報を物のように扱い、情報と情報の様々な新しい関わりやインタラクションを生み出すという点に関しては指ではじいた方向に動き出し、ぶつかって反射するということができなかつた。今後はぶつかった際の反応のルールを作る、情報同士の位置関係によるインタラクション、直線移動以外の自律動作を考える、写真や文字といった情報を取り入れたインタラクションを研究するといったことを中心に取り組んでいきたい。

参考文献

- [1] 林卓郎,"音響ホログラフィによる音場の可視化",騒音制御,Vol.15,No.4,pp166-169,1991
- [2] 外村佳伸,"インタラクティブ・アンビエンスをめざして",映像情報メディア学会技術報告,Vol.37,No.56,pp.31-34,2013.
- [3] P. McLeod and G. Wyvill," A smarter way to find pitch.",In *Proceedings of International Computer Music Conference, ICMC*, 2005.
- [4] あつきー,"タイムストレッチ, ピッチシフトのアルゴリズム",<http://ackiesound.ifdef.jp/tech/timestretch.html>