

深度センサを用いたバーチャル楽器演奏システム

村上 裕美† 中島 克人†

東京電機大学大学院 未来科学研究科 情報メディア学専攻†

1 はじめに

安価でコンパクトになり、普及が進む深度センサ付きカメラを用いたバーチャル楽器を開発している。深度センサ付きカメラとパソコンを組み合わせることで、場所を選ばず、手軽に演奏を楽しむことができるようになる。今回は机上に敷かれたペーパー鍵盤を打面とするバーチャルな鍵盤楽器(バーチャルピアノ)を試作した。紙をタッチするユーザの指先の位置を深度センサで検出することにより、対応する鍵盤の音を鳴らす。これは、電子楽器の一種となる。バーチャル楽器としての今回の要件は、音色ではなく、タッチの検出精度、および、反応時間である。本稿では、試作したバーチャルピアノのシステム構成とその反応精度等を示すとともに、バーチャル楽器であるがゆえにできる拡張機能についても述べる。

2 関連研究

コンピュータを用いたバーチャル楽器演奏に関する研究として、本物の楽器を使い、その楽器の演奏技術向上を支援するシステムの研究[1]や、CG映像を用いたエア楽器の演奏システム[2]などがある。また、HMD(Head Mounted Display)を用いてAR表示させた楽器を演奏するシステム[3]なども提案されている。しかし、演奏場所の制約、機器の設置やそれらの費用などの面で手軽に利用できるものは少ない。

3 提案システム

3.1 システムの概要

今回試作したバーチャルピアノでは、鍵盤の描かれた紙(ペーパー鍵盤)を机の上に置き、鍵盤部分をタッチすることで、演奏することができる。

タッチ判定にはペーパー鍵盤の斜め上に設置した深度センサ付きカメラのインテル® RealSense™ F200を用いる(図1)。本システムでは、ペーパー鍵盤上の指先位置の認識がカギとなる。そのため、近接位置から深度情報を取得できるこのカメラを選択した。RealSenseは深度画像に加え、カラー画像と赤外線画像を取得することができるが、今回は赤外線画像を利用しない。解像度は640×480画素、フレームレートは30fpsである。本体は150×30×5.8(mm)であるため、持ち運びに不便のないサ



図1 バーチャルピアノの利用イメージ

イズと言える。また、本システムの実装には、持ち運びが容易なノートパソコンを使用する。

3.2 処理の流れ

バーチャルピアノの処理の流れを図2に示す。

初期化ではカメラから見たペーパー鍵盤の各キー位置を調整すると共に、ペーパー鍵盤の深度情報を背景深度画像として保存する。楽器の利用中は、この背景深度画像と現在の深度画像との背景差分により、次節で述べるように、鍵盤上の指の侵入とそのタッチの判定を行い、タッチ位置の鍵盤に対応した音を出力する。

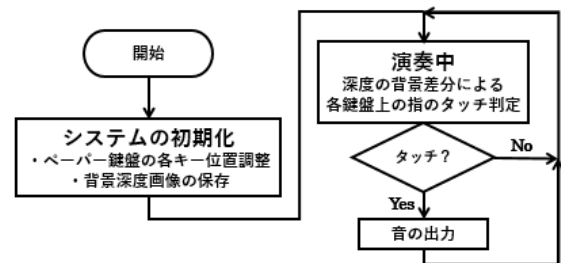


図2 バーチャルピアノの処理の流れ

3.3 指先侵入とタッチ位置の認識

背景深度と現在の深度の差分値 d が、非ゼロであるが小さい値の部分はある物体(以下、指のみを想定)がペーパー鍵盤にタッチしようとしていると判断できる。そこで、深度センサのノイズを考慮し、図3に示すように、2つの閾値 d_{min} と d_{max} を設ける。背景差分値 d がそれらの間となる深度画像の画素を前景、即ち、侵入した指とみなす。

Virtual Musical Instruments based on a Depth Sensor
†Hiromi Murakami · Tokyo Denki University
†Katsuto Nakajima · Tokyo Denki University

ペーパー鍵盤と各鍵盤に対応してシステムがタッチを検知すべき領域(今回は鍵盤幅の正方形)との位置合せは、RealSenseのカラー画像と深度画像を見比べながら人手で行う。図4はタッチ判定された際のカラー画像と前景を白で表現した深度の背景差分画像である。タッチ検知領域内の前景の比率がある閾値R以上の時、対応する鍵盤をタッチしたと判定する。

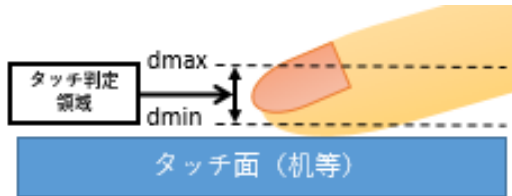


図3 タッチの際の指先深度(横)



a) カラー画像 b) 差分画像

図4 タッチ判定された時の様子

3.4 音の出力

楽器それぞれに合わせて設定した音源を出力する。今回は、ピアノ鍵盤ごとにあらかじめ設定したWAV形式の音源ファイルを再生する。

4 評価

4.1 設置環境と諸パラメータ

今回の評価ではカメラの高さを机面から 20cm, カメラの俯角を約 30° とした。ペーパー鍵盤はドからシの7鍵盤で白鍵の幅は 30mm とした。

タッチ判定の閾値 d_{min} および d_{max} は事前実験の結果を受けて、それぞれ 3mm と 10mm に設定した。

タッチ検知領域内のタッチと判定する前景比率の閾値 R は、それぞれの鍵盤位置で8回ずつタッチし、その際のタッチ判定用領域内の前景画素の占める割合が13~26%であったことから15%とした。

4.2 曲演奏によるタッチ判定精度

3名の被験者に「きらきら星」の冒頭14音を $\text{♩}=60$ (四分音符毎秒60回), 80, 100, 120 の4つのテンポで各々5回ずつ演奏して貰い、タッチの誤判定(判定漏れ, 1回タッチで複数回の判定, 隣接鍵盤と誤判定した場合)の比率を評価した。図5に示すように、 $\text{♩}=120$ という速いテンポでも判定ミスはわずかであり、鍵盤位置にも依存しないことが分かった。しかしこのテンポでは出力される音の遅延が感じられるという感想を被験者から得た。判定漏れには、タッチ検出領域内への指の侵入が不十分である場合や、タッチしている指が他の指により隠れてしまった場合があり、カメラ位置の角度の改善は検討課題である。

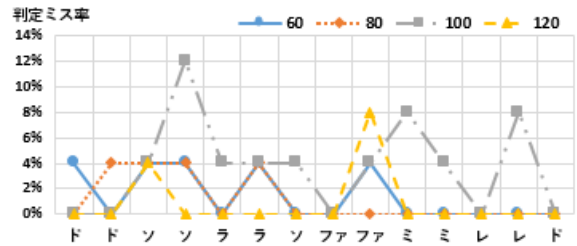


図5 曲演奏によるタッチ判定精度

4.3 タッチ反応速度

それぞれの鍵盤をタッチしてから音が再生される時間を、 $\text{♩}=120$ で各鍵盤を10回ずつタッチすることにより計測した。表1のように鍵盤ごとに若干のばらつきがあるものの、全体としては約59.0msと十分に高速であった。しかし、連続してタッチする中で音の出力が追い付かず、不自然な間を感じてしまうことがあった。

表1 鍵盤ごとの反応速度 (単位: ms)

	ド	レ	ミ	ファ	ソ	ラ	シ	全体
平均	57.5	58.7	58.3	59.2	59.3	61.0	59.1	59.0

5 今後の課題

今回の試作で、簡単な曲を弾いて遊ぶ程度なら使用に耐えるタッチ反応の精度を得ることができた。しかし、早いテンポでの演奏ではタッチと音に幾分かのタイムラグによる、違和感が生じる事があるため、原因の特定と改善を目指したい。また、多様な音の発出のために音源をWAVファイルからMIDI等に置き換えることも望まれる。鍵盤とタッチ判定領域の位置合せの画像認識を用いた自動化も、手軽さを求める上での課題である。

深度センサを用いたバーチャル楽器ならではの拡張も今後の課題である。例えば、ピアノ以外の鍵盤楽器や打楽器等への拡張、複数の楽器の同時動作、深度を利用したエフェクト(抑揚・音程の変化など)などの機能の追加が考えられる。また、幼児教育や英語教育などでの活用を想定した、音が鳴る絵本や英語の本などにも発展させることができると考えている。

参考文献

- [1] 竹川佳成, 寺田努, 塚本昌彦, “運指認識技術を活用したピアノ演奏学習支援システムの構築”, 情報処理学会論文誌, pp.917-927, 2011
- [2] 岩谷亮明, 澤田秀之, “VRエンタテイメントに向けたエア楽器演奏システム”, 情報処理学会インタラクシオン2014, pp.587-592, 2014
- [3] 一之瀬元喜, 坂東哲平, “3次元動画処理による仮想楽器システムの提案”, 人工知能学会全国大会, 1G1-4, 2010
- [4] A. D. Wilson, “Using a depth camera as touch sensor,” International Conference on Interactive Tabletops and Surfaces (ITS) 2010, pp.69-72, 2010