

簡易な手書き譜面を利用した演奏システム Gocen の設計

馬場 哲晃^{1,a)} 菊川 裕也¹ 串山 久美子¹ 青木 允²

受付日 2012年7月2日, 採録日 2013年1月11日

概要: 幼少時期に限らず, 我々は楽譜を読むには一定の学習と練習が必要であり, このことは音楽を楽しむうえでの障害になる場合がある. 本稿では五線と符頭により構成される簡易な手書き譜面を, リアルタイムに画像処理を行うことで, 記譜された譜面を即座に演奏可能にするシステム “Gocen” に関して報告する. 1960年代以降, 光学楽譜認識 (Optical Music Recognition) に関して多くの研究が発表された. 現在においては製品に応用される等, 一般ユーザが手軽に享受できるほどになった. 一方で手書き譜面を対話的に演奏可能なシステムについてはこれまであまり事例が報告されていない. そこで我々は記譜と演奏を同時に楽しめるシステム Gocen を提案し, 演奏手法の設計と実装を示すことで, 音楽学習や演奏パフォーマンスへの応用を提案する. また, ユーザテストを通じて, 簡略化された譜面において 97%の精度で初学ユーザが任意の音高を再生可能であることを示した.

キーワード: 演奏インタフェース, 光学楽譜認識, 画像認識, 演奏インタラクションデザイン

Gocen: Design for Musical Interface Based on Handwritten Notation

TETSUAKI BABA^{1,a)} YUYA KIKUKAWA¹ KUMIKO KUSHIYAMA¹ MAKOTO AOKI²

Received: July 2, 2012, Accepted: January 11, 2013

Abstract: Not only in childhood but also adulthood, we need some trainings to read music scores, which sometimes make music hard to learn and enjoy. In this article, we shall propose the system that enables users to play their handwritten musical notations by our musical interface. Since 1960's, Optical Music Recognition (OMR) has become matured in the field of printed score. In recent years, some products were released on a market that uses OMR for music composition and playing. However, few researches on handwritten notations have been done, as well as on interactive system for OMR. We combined notating with performing in order to make the music more intuitive for users and give aids to learn music to users. We shall report that the recognition rate of the pitch of note is about 97% and we can play some songs with our Gocen system.

Keywords: musical interface, optical music recognition, computer vision, musical interaction design

1. はじめに

電子音源の登場から, 電子楽器インタフェースは多様化し, 身体動作を利用したものやウェアラブルなもの等, 現在においても継続的に様々な電子楽器が発表されている. 古典楽器の模倣から, まったく新しい電子楽器等, これらは演奏者や一般ユーザにとって音楽演奏パフォーマンスに

新たな可能性を提供している. 一方で, それらパフォーマンスを行うためには, 事前に作曲を行うのが一般的である. 作曲には記譜を用いるのが一般的である. 17世紀頃から利用され始めたといわれる五線譜による記譜法が広まり, 現代においても一般的な記譜には五線譜が用いられている. このほかデスクトップミュージックではピアノロールやタイムライン上に音声ファイルを貼り付ける等の, いわゆる「打ち込み」を利用した作曲も広く利用されている. しかし音楽教育の教科書では必ず五線譜記譜法が利用されるように, 一般的な記譜には五線譜を利用することが, 音楽家

¹ 首都大学東京大学院
Graduate School of Tokyo Metropolitan University, Hino,
Tokyo 191-0065, Japan

² 株式会社ウイズコーポレーション
With Corporation, Adachi, Tokyo 120-0005, Japan

a) baba@sd.tmu.ac.jp

本論文に関する動画コンテンツを以下で視聴できる.
<http://www.youtube.com/watch?v=OjFNZD-bC5Y>

や作曲家の基礎となる。作曲家が五線譜を読むのは当然であるが、演奏家が五線譜を読めない場合がしばしば見受けられる。それは、五線譜に記載された楽曲を読むためには、ある程度訓練を必要とするためである。幼少期に限らず大人になってから音楽を楽しもうとする場合、この“五線譜を読む”行為が音楽を楽しむうえでの障害になることがある。そこで、本研究では記譜と演奏を合わせたインタラクションを提案する。楽曲を記述することと、演奏することをつなげることで、音楽教育における展開が望める。

音楽教育の1つに読譜学習があり、読譜力を身につけるためには簡単な譜面を繰り返し練習するのが一般的であり、それを支援するシステムもいくつか報告されている [1], [2], [3]。しかしながら依然として幼少期の音楽教育を経てもなお、読譜能力が身につかないこともしばしばある。本研究ではこのようなユーザやこれから音楽学習を始めるユーザにとって、コンピュータインタラクションの観点から新しい記譜体験を提供することで、ユーザに五線譜に対して親しみや楽しみを持たせることを狙いとする。

このほか、演奏パフォーマンスへの展開も視野にいれている。近年のラップトップコンピュータに代表される音楽パフォーマンスは従来の楽器では不可能な音表現ができる一方で、演奏時におけるパフォーマンスの行為と音出力の対応関係把握が困難で、鑑賞者への理解性が欠如している場合がある。古典楽器やそれらを模した電子楽器類は演奏手法と音出力がほとんどの場合1対1で結び付いているため、理解が容易である。一方で、実験的な電子楽器やコンピュータ・ミュージックパフォーマンスの場合、音と演奏手法の関係が必ずしも1対1ではなく、さらに対応関係を変更することで、音出力と演奏手法の組合せを自由に変わることができてしまう。たとえば、ボタンやつまみ、スライダ等の汎用的なインターフェースが用いられた場合、初学ユーザにとってインターフェースそのものを操作/動かすことはできるが、それを操作するまでどのような音が鳴るのか、といった対応関係を即座に理解することが難しい。Normanはこのようなユーザがインターフェースを操作する際に描く操作イメージをメンタルモデル [4] と呼び、使いやすいインターフェースにとって重要な概念であることを示した。

そこで本システムでは鑑賞者に対して記譜の様子やスキャナ画像を演奏時にプロジェクションすることで、演奏パフォーマンス時における理解性向上を狙う。光学楽譜認識だけでなく、OCR機能を用いることで、いま何が起きているのかを鑑賞者は画像情報（楽譜情報）から容易に知ることができる。

具体的にはその場で紙に記述した五線および音符情報をイメージセンサでスキャニング操作をすることで、記述された楽曲を演奏可能にするシステムを提案する。スキャン操作はユーザが手動で行うため、再生速度や逆再生、好きな場所からの再生等の直感的操作が可能となる。さらには

スキャニングデバイスに身体的な操作を付け加えることで、ベンドやビブラート等の変化をリアルタイムに付加させることも可能となる。本システムの使用時のスケッチイメージを図1に示す。パフォーマンスは簡易な五線譜法によって楽曲を記述し、それをデバイスによってなぞることで即座に楽器音を出力する。

後述するとおり、現状における手書き楽譜に関する OMR 技術では、その認識率や処理速度の面から五線譜をインタラクティブな楽器として扱うことは困難である。そこで本研究では、グレゴリオ聖歌等に利用されたネウマ譜（図2参照）を参考に、五線と符頭から構成されるさらに簡易な手書き楽譜を対象とすることで認識率と処理速度の問題を解決し、ユーザインタフェースやインタラクションの設計を行う。なお符頭とは、図3に示す部分であり、本稿では特にこれを塗りつぶしたものを符頭と呼ぶ。

1.1 紙とディスプレイ

紙とディスプレイに関する議論は主に読みやすさに関する先行研究が多く報告されている。古くは1982年にGouldら [5] が印刷物と CRT ディスプレイにおける文章校正速度および正確性について実験を行い、CRT ディスプレイよりも印刷物の方が20-30%ほど校正が速いことを示した。近年では電子ペーパーやタブレット端末の登場により、紙と

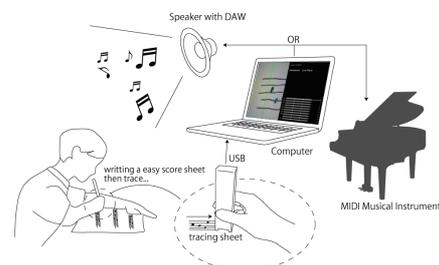


図1 システム概要のスケッチ

Fig. 1 A sketch of our system abstract.

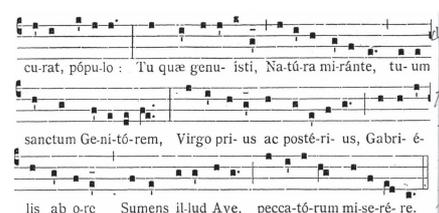


図2 グレゴリオ聖歌楽譜の一例

Fig. 2 A sample score of gregorian chant.

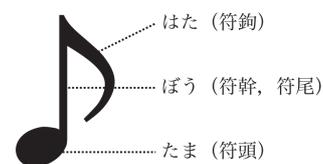


図3 音符の構成

Fig. 3 Composition of a musical note.

従来ディスプレイでは条件を統一できなかった携帯性や閲覧姿勢等をユーザ評価に取り込んだ研究も報告されている [6], [7]. さらに一定の条件下において紙と同等の視認解像度を持つ高解像度ディスプレイの登場により, ディスプレイの表現能力は着実に紙に近づいている.

本システムは紙への記述を前提としているが, タブレット端末を利用し, 記譜をタッチパネル上で行うことでユーザの手書きデータをオンラインで取得可能になるだけでなく, 修正作業も容易にできる. 利便性や効率性においてはすべてのシステムをこのようなデバイスとソフトウェア上で構築することも有効な手段である. 一方で, 紙メディアでは折り曲げや破く等の操作が可能であるほか, 折り紙のような立体物構成やこれら操作をした結果, 元には戻せない不可逆性質を持ち, これらは紙が持つ大きな特徴である. ディスプレイの解像度に制限されることなく, 好きな場所に紙を並べることができたり, 手書きを行ったりすることで, デジタルでは表現が困難な素材の質感を得ることができる. ユーザに対しての制約が少ない点やその作業過程において, 様々な「気付き」をユーザに提供する可能性が大きい点から, 紙と手書きによる作業は創造的な作業に向いていると考えられる. 作曲においても紙と手書きが古くより利用されてきた. Tsandilasら [8] は作曲支援システムの制作を通じ, 作曲家の手法を観察したところ, コンピュータを利用している作曲家においてもその多くが作曲初期段階で紙への手書きを利用していることや, 様々なアイデアや楽曲構成を書き留める際には, 紙の自由な記述可能性がユーザの創造性支援に成功していることを報告した.

創造的作業には明確な目的が定まっておらず, その過程において様々な気づきを得ることで作り手は新たな表現や作品制作が可能になることがしばしばある. 本システムは音楽演奏や学習を目的としたシステムであるため, なるべくユーザを拘束しない, ユーザの創造性を妨げないインタフェースであることが望ましい. そこで, 紙と手書きによる人間が古くより慣れ親しんできた手法をそのまま利用できることを, 設計指針の1つとした.

本システムはデバイスの操作方法によって, 1つの旋律を逆再生方向や上下反転方向でも音符の再生が可能であることから, 一意に楽曲再生を限定することができない. これは楽曲の再現性の点から, 本システムの問題点といえるが, 同時にユーザに対して操作性の自由度を与えていることになる. 本研究では後者を重視し, ユーザの対話的な記譜体験を通じて, 本システム価値を議論する.

2. 関連研究

2.1 光学楽譜認識

手書き楽譜を扱う際, オンライン入力とオフライン入力に大別される. オンライン入力は, コンピュータディスプレイ等にタッチペン等を利用して記号を入力する手法で

あるのに対し, オフライン入力は紙面等に記述された記号を読み取る手法である. オンライン入力の場合, ユーザのストロークデータが取得できるため, 入力方式に関してこれまでいくつかの検討がなされてきた [9], [10]. 特に Forsbergら [11] による Music Notepad は, ストロークを簡略化することで, 手軽にオンライン手書き入力を可能にしたシステムとして広く知られている. これに対して本研究ではオフライン手書き入力になる.

楽譜認識において, これまで最も活発に研究されてきた領域は, オフライン印刷入力による光学楽譜認識 (OMR: Optical Music Recognition) に関する研究である. 1963年に Kasslerら [12], 1966年に Pruslin [13], 1970年に Pre-rau [14] らが自動認識システムに関する論文を発表して以来, 現在に至るまで依然として活発な研究領域となっている. 国内では, 1980年代以降, 青山ら [15] や, 大照ら [16], 松島 [17] らを皮切りに, 楽譜認識技術に関する研究が活発に行われてきた. 楽譜認識の試みは数多く行われているが, 画像画質, 記号重なりや多様性等の理由から, すべての楽譜を完全に自動認識することは容易ではない. しかしながら認識精度はすでに実用レベルにあり, いくつかの市販ソフトでは自動認識機能を有している. また, 市販ソフトに限らず, Audiveris [18] や OpenOMR [19], Gamera Framework [20] 等, 開発者や一般ユーザが利用できる OMR ソフトウェア (ライブラリ) も提供されてきた.

リアルタイムにおける演奏を考慮した事例として, 最近ではスマートフォンアプリを利用した楽譜認識ソフトウェアが発売され [21], より手軽に五線譜から楽曲を再現することが可能になった. 河合楽器製作所の製品「PDF ミュージシャン」[22] では, PDF 楽譜を手でなぞって演奏ができる演奏設計を提案しており, 直感的な演奏体験をユーザに提供している. ただし, 事前にテキスト PDF と呼ばれる, 音符情報が埋め込まれたデータを利用しており, 即時的かつ高精度な楽譜認識が実現できているわけではない. 山本ら [23] は楽譜画像特徴量を利用し, 対応する楽譜情報をデータベースから取得することで, 楽曲の再現を可能としたシステムを提案している. これらは単なる楽譜認識にとどまらず, より楽譜を身近に楽しめる直感的なインタラクションを実現している例といえる. これら先行事例において, 「楽譜を利用して演奏する」という設計は非常に本研究と類似している. これらに対して本研究では, 既存楽譜ではなく手書き楽譜を対象とし, スキャニング操作によって楽譜には記載されていない音長や音高を動的に制御できる点が異なる.

Hoerterら [24] による “The Music Wand” は, 本研究と同様にユーザがスキャニング操作をしながら印刷された楽曲を演奏可能とした事例である. 手持ち型のデバイスのスキャン操作により, 印刷楽譜をリアルタイムに再生するものであるが, 音高認識が70%の認識率であったり, 各種臨

時記号の認識に関してはそれ以下であったりすること等の問題点が報告されている。我々の初期プロトタイプ [25] において、Hoerter らのようにラインイメージセンサを利用していたが、Hoerter らと同様の認識率程度であった。このほか、Hoerter らのシステムでは、2 値化画像取得のためにトレースボックスを必要としたり、ユーザのスキヤニング操作補助用に定規を利用したりしている等、ユーザビリティに問題があるといえる。本研究では手書き譜面を対象とするほか、操作方法によって各種アーティキュレーションや楽器選択ができる等、Hoerter らの研究と比較し、より高い認識率や、多種のインタラクションを提案する。

OMR の多くは印刷楽譜を対象としたものが多いが、手書き楽譜についても研究は行われており、1970 年代以降、手書き楽譜に関してもいくつかの研究事例が報告されている。Bulis [26] らは縦横の 2 値画像スペクトラムによるテンプレートマッチング手法や、Yadid-Pecht ら [27] によるニューラルネットワークを利用した手書き楽譜認識手法で 80-90% 程度の認識率を示した。近年では動的時間伸縮法を利用した事例 [28] 等が報告されているが、インタラクティブな操作を考慮した場合、即時性や認識率に関して依然として問題がある。

2.2 インタラクティブシステム

工学やメディアアート分野において、多くの楽器作品がこれまで発表されている。なかでも楽譜をモチーフとすることで、演奏者の学習支援やパフォーマンス時の聴衆への理解性を高めている作品がある。手書きではないが、五線譜上に音符記号オブジェクトを置くことで楽曲演奏を行うシステムとして、Bottello による Tangible user interface for music learning [29] がある。あらかじめマーカをつけた実物体オブジェクトを印刷された五線紙上に配置することで、楽曲演奏が可能になる。実物体オブジェクトを置くだけでなく、オブジェクトの引き伸ばしや回転動作に対応しており、それぞれの動作によって拍の長さや臨時記号に対応できる。Bottello らはこれらタンジブルな電子楽器を応用して、音楽教育への可能性をふまえている点で、本研究に類似している。また、オブジェクトをト音記号や四分音符等の形に成型し、それらを五線譜が描かれたテーブルに並べることで演奏を可能にする Noteput [30] は、より直感的に五線譜による演奏を実現している。このような実物体をタイムライン上に配置し、それに応じて楽曲を再生する手法は Levin [31] や岩井による「音楽のチェス」[32] 等多くの先行事例がある。

Overholt [33] によるスキヤナデバイスは楽譜情報ではなく、画像情報を音響生成のパラメータとすることで、作曲や演奏に応用可能なシステムを報告した。指先の圧力センサによる音量制御や色情報の利用、サンプリング機能等、本システムの対話設計指針において参考にできる。

Zachary による作品「Drawn」[34] は、筆で手書きされた記号や絵を、コンピュータに取り込み、ユーザが取り込まれた画像オブジェクトに対して手で触れることで、それらを動かしたり、音を鳴らしたりすることのできるアート作品である。手書きによる“アナログ感”と、データ化による“デジタル感”を美しく融合させた作品である。本研究では、ユーザの手書き楽譜（アナログ）を即座にデジタル化し、ユーザが対話的に演奏可能になる点が類似しているため、操作設計指針において参考にした。

3. システム概要

本章では制作したシステムについて、操作方法を軸に概略的に解説する。システムは図 1 に示したとおり、主にデバイス、コンピュータ、音源の 3 つからなる。デバイスはカメラモジュール、スイッチ群、マイクロコントローラ、小型振動モータからなる。内部の構成を図 5 に示す。スキヤナデバイスの重量は 163 [g] (ケーブル含む)、高さ 94 [mm]、覆い部分の外形 (最大) が 55 [mm]、手持ち部分の外形寸法は 34 [mm] である。デバイス外観の成形には Dimension 社製 uPrint を使用し、成形材質は ABS 樹脂である。

ユーザのスキヤン操作によって取り込まれた画像を処理し、その結果を適合する音程や楽器音で出力する。カメラ画像から取得する画像サイズは、ユーザへの快適な操作感を考慮し、120 × 160 px (横 × 縦) とし、更新速度は 30 fps である。

コンピュータソフトウェアは Xcode 上で作成し、主

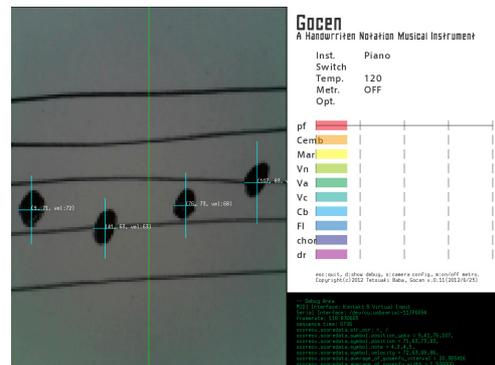


図 4 PC 上に表示されるアプリケーション画面

Fig. 4 A screenshot of Gocen application on PC screen.

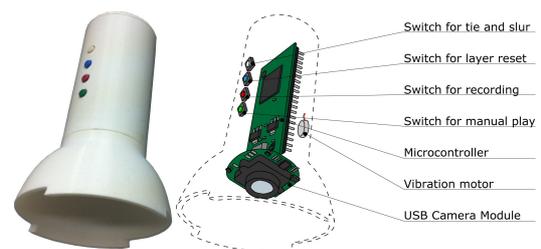


図 5 デバイス内部

Fig. 5 Inside of our device.

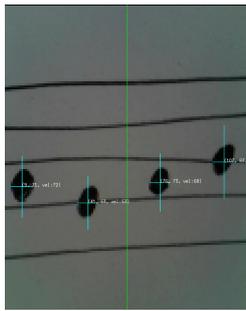


図 6 符頭を認識しているときの画面
 Fig. 6 A screenshot of recognized notes.

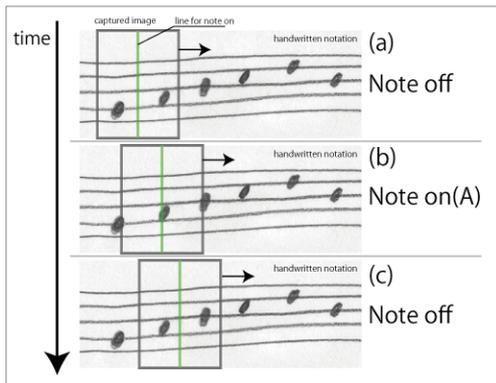


図 7 発音手順
 Fig. 7 Flow to make sound.

に openFrameworks [35], OpenCV, OpenGL, portmidi, Ocrad 等のライブラリを使用した。アプリケーションからは任意の MIDI デバイスを設定できるが、現行システム環境下において、Native Instruments 社の Kontakt Player [36] を使用した。

3.1 発音方法

ユーザはデバイスを手に取り、図 4 に表示される画面を参考に、演奏を行う。ユーザが記述した手書き譜面上にデバイスを接地することで、図 6 に示される画像がコンピュータ画面に表示される。符頭位置が十字にマークされ、ユーザがデバイスを操作し、画面中央の緑色の演奏ラインに重畳させることで、符頭の音高を出力できる。なお十字の縦はその符頭の音量情報を、横は符頭認識領域幅を示している。ユーザは図 4 に示す画面を利用して演奏を行う。図 4 にはカメラ画像、選択楽器名、オプション選択、タイムライン確認、デバッグのそれぞれの項目が確認できるようになっている。演奏ボタンを押しながらスキャニング操作を行うことで、ユーザは楽器音の発音が可能になる。

3.2 OCR 機能

五線譜を利用した符頭再生のほか、ユーザは指定された文字列を記述し、デバイスに読み込ませることで楽器選択、転調、音域等の様々な設定が可能になる。

たとえば楽器選択の場合、一般的に記譜上の楽器を示す場合、ピアノであれば pf, ギターであれば gt といった省略文字列が利用される。そこで本システムにおいても同様に、手書きで記述した文字列をカメラで取得、処理することで任意の楽器選択機能を実現した。OCR ライブラリを利用しているため、事前に設定した楽器であればどのような名前でもある程度正確に認識が可能である。記述の仕方や、文字に決まりがあるため、乱雑な文字等も認識するわけではないが、事前の教示を受ければ比較的正確に楽器切替えができる。外部 OCR ライブラリ [37] を使用しているため、認識精度に関する問題は本稿では扱わない。

本機能は、取得画像上に五線譜エリアが 1 カ所も認識されない場合、自動的に OCR モードに切り替わるため、ユーザは任意のボタンを押す等の操作を必要としない。これら設定文字列を記述する箇所は、紙面上の空き領域を適時使うことができる。現状のシステムでは、五線譜の演奏と同時に楽器変更、転調を行うといった、OCR 機能の併用はできない。

3.3 シーケンス演奏

本システムには対話的な演奏方法に加え、ユーザの演奏をタイムライン上に録音し、それらを重ね合わせることで複雑な演奏を可能にするシーケンス演奏機能を実装した。シーケンス演奏を行うには、デバイス上の録音ボタンを押しながら、演奏を行えばよい。各楽器に 1 レイヤを設定し、ユーザは楽器を切り替えると同時にシーケンス演奏用のレイヤを切り替えることになる。これによりユーザはレイヤ概念を意識せずに演奏を楽しめる。ユーザは対話的に切替え可能なメトロノーム音を基準に、録音時間に応じて伸縮されるタイムライン上に発音ポイントを作成できる。録音長は $4 \times n$ 小節長を基準に伸縮できる。なお録音機能にはクオンタイズ機能が実装されているため、ユーザが乱雑に発音を行った場合におけるわずかな発音ポイントの揺れを修正できる。1 度録音したタイムラインはデバイス上のリセットボタンを押すことで消去することができる。一方で、1 度録音したタイムラインにさらに同じ楽器音で録音を行う場合、再度録音ボタンを押しながら演奏することで録音ポイントが追加できる。

4. 認識手順

4.1 処理の流れ

五線および符頭に関する認識手順を図 8 に示す。デバイス上の CMOS センサから取得された画像が USB ケーブルを通じてコンピュータに送信される。取得した画像(図 9-1 参照)にガウシアンぼかしをかけ、判別分析 2 値法 [38] により 2 値化する。ただし、ここで正確な 2 値画像を得るために、本デバイスでは 4 カ所に白色 LED を取り付け、周辺減光の比較的少ない CMOS モジュールを使用

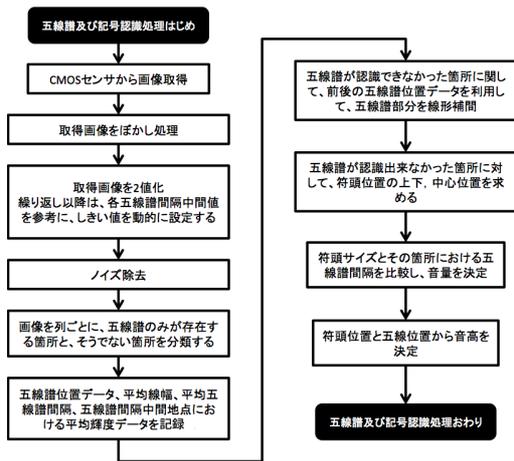


図 8 五線および符頭の認識手順

Fig. 8 Process for recognition of staves and simple note.

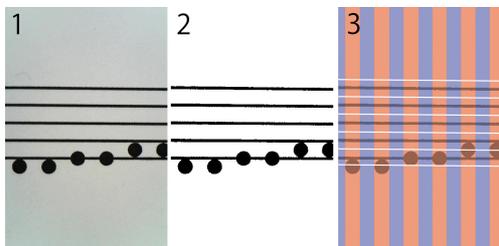


図 9 左：取得画像，中央：2 値化画像，右：領域判別画像
Fig. 9 Left: original, center: binalized, right: domain.

している。得られた画像 (図 9-2 参照) に、簡単なノイズ除去を行う。

次に五線位置の検出を行う。まず、五線 1 本あたりの線の太さ平均値 l_{aw} を事前に設定しておく。ただし、 l_{aw} はカメラ画像を更新する際、そのつど得られた五線の平均の太さに更新される。比較対象線の太さを l_{cw} とし、 $|l_{aw} - l_{cw}| \leq 3$ [px] である場合、その線分箇所を五線の線分とし、これが画素列において、ちょうど 5 カ所検出された場合を「五線検出」とし、それ以外を「五線不検出」とする。図 9-3 に五線を検出した箇所を青色、不検出の位置を赤色で示す。なお画素列とは、カメラ画像を画面下から上に走査した場合の 1 ピクセル幅の画素列を指す。

五線不検出箇所に関しては、その箇所の右および左隣にあたる五線検出箇所 (図 9-3 における青色領域) から各五線位置を線形補間する。ただし隣に五線検出箇所がない場合は、線形補間はされず、五線位置は未定義となる。なお、各五線位置を基に、それぞれの中間値を求めた補助線 (以下 6 線) も同時に求めておく。6 線は図 9-3 に白線として図示する。この処理により五線が等間隔直線にならなくとも、五線位置を正確に検出できる。

4.2 五線上における符頭位置の決定

これまでの手順によって取得した情報を基に、音高を決定する符頭位置を求める。符頭の音高を決定するにあつ

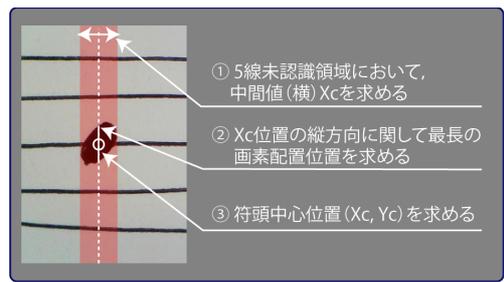


図 10 符頭音高位置の決定方法

Fig. 10 How the horizontal position of simple note is detected.

て、手書き譜面においては必ずしも符頭は綺麗な円を描かない。さらに、本研究では符頭サイズを音量として扱う (後述) ことから符頭サイズがそもそも一定のものではない。各符頭画像における縦横のスペクトラムから重心位置を推定する等、いくつかの手法を試したが、縦横の中間値をとる手法が認識精度が明らかに高かった。そこで、図 10 に示すように、五線未認識領域 (赤色) に関して、その領域の中間位置を縦方向に走査し (図 10 の点線部分)、最長の黒ランの中心位置を符頭の中心 (X_c, Y_c) として扱うこととした。得られた各 Y_c と各 6 線位置を比較し、最も近い五線または 6 線位置を音高位置とした。ただし図 10 において、横軸を x 、縦軸を y とし、 x は右方向に、 y は上方向に値が増加する 2 次元座標系である。

通常の記譜では、加線を用いて音域を広げることができる。加線の数に制限はないが、あまり多くなると読みにくくなる。その場合、加線を描く手間を省くため、オクターブ記号が用いられる。特に本システムの場合、加線が増えすぎると読みにくくなるだけでなく、スキャナデバイスの画像からはみ出す恐れがある。そこで加線は上下第 1 線までとし、それ以上もしくは以下に符頭を記述した場合は上下第 1 線の音高になるよう設定した。本システムはプロトタイプであるため、加線をいくつまで許可するかは今後の議論とする。

5. 演奏方法の設計

ここまでで、五線上に記述された任意の符頭位置を推定でき、Note On/Off については制御可能となったが、演奏に必要な様々な手法を十分に再現できてはいない。音楽を演奏する際、音の強弱、音高のゆらぎ (ピッチベンド、モジュレーション、ビブラート)、和音、発音長さ等のアーティキュレーションによって、演奏に表情付けを行う。また、1 行の五線では、音域に限界がある。そこで、このような表情付けや機能拡張を実現するための設計と実装を行った。

5.1 発音長

図 6 に示される緑色の発音ラインと符頭を重畳することで、任意の音程を発音できる。ユーザのスキャニング操作

速度によってノートの制御が可能であるため、スラーやスタッカートが表現可能になる。しかし実際に演奏をしてみると、手元の制御が難しく、符頭間の音をスムーズにつなげて演奏することが困難であった。そこで、タイ・スラー機能を付加した。タイ・スラー機能は図 11 に示すように、各符頭間にタイやスラーに似た形状のグラフィックを画面の上に重ねて表示し、各符頭の Note Off のタイミングを次の符頭発音もしくは演奏ボタンを離すときまで継続させる。これはデバイスの手元ボタンで切替え可能である。

5.2 音の強弱

譜面上で音量を指示する場合、フォルテやピアノ、アクセント記号等を用いるのが一般的である。本研究ではこれに対し、より直感的な操作インタフェースを提供するために、符頭の大小を音量として扱う。図 12 に示すとおり、符頭の大小に合わせてベロシティ（音量）が異なっているのが分かる。

5.3 音高のゆらぎ

主に弦楽器を利用する場合、発音後に音程を一定に保つだけでなく、音程にゆらぎをつけることで、ビブラートや（逆）ターンといった音高のゆらぎによるアーティキュレーションを表現する。本研究において、ユーザは発音後、デバイスを上下に動かすことで音程にピッチベンド効果を与えることができる。これによって前述のアーティキュレーションを表現可能である。発音時の符頭位置を基にカメラの上下動作量を検出している。発音時の符頭縦位置を Y_1 、発音後の符頭縦位置を Y_2 、とした場合、 $|Y_1 - Y_2|$ をピッチベンド変数に割り当てる。ただし、デバイスの移動が下方 ($Y_1 - Y_2 > 0$) のときは、ベンド音高を下降方向に、デ

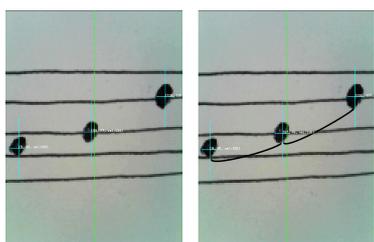


図 11 タイ・スラー機能。右が機能適用時のスクリーンショット
 Fig. 11 Tie and slur function. The right figure shows the screenshot that the function is applied to.

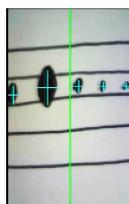


図 12 符頭の大小とベロシティの関係
 Fig. 12 Relation of velocity and size of notes.

バイスの移動が上方 ($Y_1 - Y_2 < 0$) のときはベンド音高を上昇方向に変化させる。

5.4 音域の設定

制作したプロトタイプシステムにおいて、認識された五線は初期状態において、ト音記号とハ長調が設定された五線としている。そのままでは音域が C4 - A5 となる。一般的な記譜法では補助線やオクターブ記号 (8va や 15va) を利用することで音域を拡張可能である。本システムでは音域を設定する手法として、楽器選択と同様に OCR 認識を利用する。具体的にはユーザが記述した “+15”, “+8”, “0”, “-8”, “-15” 等の文字列を読むことで、オクターブの音域移動を可能にした。

5.5 調の設定

上記の説明のとおり初期状態において、ハ長調の設定がされている。先の音域の設定と同様に、これについても OCR 認識機能を利用して解決する。一般的な記譜に倣い、“A”, “Am”, “B”, “Bm” 等の文字列を読み込ませることで調の設定を可能にした。

5.6 転調

シーケンス機能で演奏中の楽器音に対し転調操作を行うことができる。OCR 機能で文字列 “key” を読み込んだ後、転調したい音程となる符頭を再生操作すればよい。

6. ユーザテスト：音高認識

制作したプロトタイプを利用して、ユーザが比較的自由に記述した手書き譜面をどの程度正確に再生可能かを知るため、ユーザテストを実施した。

6.1 目的

本システムにて実装した、記述した符頭の Note On/Off のみの機能を利用し、ユーザが意図した音高をどの程度正確に再現可能かを明らかにする。

6.2 実験手順

被験者に対し、事前に Gocen システムおよび記譜法に関して実験シート 1 枚を配り、説明を行う。その後本記譜法に倣い、被験者にきらきら星を記述、演奏してもらう。事前に用意した本システム用の楽譜を提示した後、被験者にキラキラ星を記述してもらった。ただし手書きの場合、被験者の記述方法に影響を与える恐れがあるため、コン

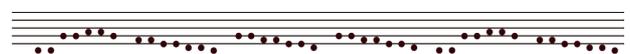


図 13 ユーザに提示したきらきら星の譜面例
 Fig. 13 An example of score of “Twinkle, twinkle, little star” that is shown to users.

コンピュータ上で作成したもの(図 13 参照)をユーザに提示した。手順を以下にまとめる。

- (1) 実験説明 (5分程度)。
- (2) きらきら星を手書きで記述。
- (3) 被験者にデバイスを利用して演奏してもらう。
- (4) 記譜に誤りあった場合や、意図しない音高があった場合に記述した譜面を修正してもらう。
- (5) 再度演奏し、正しく再生できるか確かめてもらう。もし正しく再生できない場合は(4)に戻る。
- (6) 自由記述アンケート回答時間。

被験者には2本のペンを用意し、好きなペン(0.7mm油性ボールペン, 0.5mm油性ボールペン)で記述してもらった。紙は一般的なコピー用紙(再生紙)を使用した。また、きらきら星の旋律は図 13 に示す音高の連なりとし、合計で42音とした。実際にユーザが42音中、どれだけ音高を再現できたかで、認識精度を示す。

6.3 被験者

被験者は合計23人、有効回答は20人である。これはユーザが事前説明を理解できていないまま、実験を始めてしまった場合を除外したためである。20人中、男性9人、女性11人、平均年齢は21.9歳、記譜経験については、「日常的にする」:0人、「たまにする」:3人、「めったにしない」:9人、「ない」:8人であった。「きらきら星」が既知楽曲であることを実験前に全被験者に対して確認した。

6.4 実験結果

1度目の試行において、42音すべて正確に再現できたユーザは11人、修正を加えた後、正確に再現できたユーザが9名であった。表 1 に結果を示す。1度目の記述において、全被験者における平均符頭認識率は97.4%であり、その後間違い箇所を修正し、2度目の試行では残りのユーザも42音すべて正しく再生できることを示した。なお、実際にユーザが記述した譜面を図 14 に示す。

自由記述アンケート

実験後、自由記述で本システムのユーザの感想をまとめ



図 14 ユーザが記述した譜面例

Fig. 14 Examples of scores that users wrote.

た。「おもしろかった」、「とても画期的で実際に実用化したらかなり便利」、「すごかった」、「これがあればもっとピアノの練習が頑張れた」、「線がゆがんでいても読み取れるところが良い」等、好意的な感想の一方、「持ちにくい」、「丁寧さが必要」、「色を使ってオクターブを操作できたらもっと良い」、「最初は難しく感じる」、「下のドに横線(補助線)がないことに違和感」等といった不満があげられた。

6.5 考察

記述に関する実験において、対象とする譜面を非常に簡略化したため、期待どおりの高い正確性を実現できた。ユーザの記述ミスについては、表記ゆれによる誤認識のほか、ユーザの音程誤認識も含まれており、このヒューマンエラー率が下がればさらに高い認識率を期待できる。また、操作時においても30fps程度の画像更新速度を維持したため、ユーザからは「思った以上にレスポンスが早かった」といった意見があがった。

本ユーザテストではユーザに基本的な機能しか紹介しなかったため、機能付加に関する意見を多くもらった。多くの機能についてはすでに実装済みではあったが、「色を利用したい」、「左利きには利用しにくい」、「和音が使えるようになると楽しみ」等、今後の発展において重要な意見を集めることができた。

7. 操作方法に関する考察

これまで本システムを利用し、国内で3度、海外で1度の展示を行った。体験者に本システムの説明および記譜と演奏体験をしてもらい、その際のユーザ観察および、記録されたノートから本システムについて考察をする。なお、記述されたノートは現在のところ90ページ、体験者は約300人である。また展示の際、本研究室内で十分に楽曲を練習し、実際に合奏を行うパフォーマンスも実施した。そこでの考察もあわせて述べる。

7.1 発音長

発音長制御は、符頭位置から発音ラインの重畳で可能であるが、これだけでは2つ以上の音符を切れ目なしに演奏する際に問題が生じる。そこでタイ・スラー機能を追加し、次の音符を発音するまで、それ以前の音符をノートオフしないようにした。この場合では、休符における発音長制御ができなくなるが、ユーザがデバイスの再生ボタンをoffにすることでその問題を解決した。実際の展示においては

表 1 実験結果

Table 1 Result of the user test.

試行回数(人数)	音符数	認識音符数(平均)	認識率
1(20)	42	40.9	97.4%
2(9)	42	42	100%

記譜を行い、その音程を再生する程度であったため、音長制御についての細かな意見を集めることはできなかった。合奏演奏パフォーマンスの際は、必ずタイ・スラー機能を on にし、音長制御は手元の再生ボタンで行った。単音の演奏については問題がないが、今後和音認識を行った場合、現状の操作手法だけでは図 15 に示す二分音符の音長を再現できないという問題がある。

7.2 音の強弱

本稿で実装した強弱認識については、見た目の縦サイズがそのまま音量になるため、体験したユーザは即座にその内容を理解できていた一方で、演奏パフォーマンスの際は強弱については固定とし、OCR 機能で、mf や mp 等の音量調整を事前に行った。それは現状のシステムでは微妙な強弱はつけることが難しく、演奏時においてその強弱操作が困難であったためである。強弱の解像度はカメラ画像中の符頭ピクセル数に依存するため、カメラ画像の解像度をあげることでこの問題は解決可能であると考えている。本手法はユーザにとって理解しやすいものである一方、1 度記述した音符の音量を動的に操作することができない。そこで「スキャナデバイスの移動加速度に応じて音符の強弱を制御してみてはどうか」という意見を体験者との議論から得ることができたため、今後の展望としたい。

7.3 音高のゆらぎ

音符の発音中（ただし発音ラインと符頭が重畳している場合のみ）にデバイスを紙面に水平に上下操作することで、ピッチベンドを実現した。展示および演奏時は、発音高から上下に長 2 度分変更可能であり、それぞれ上下方向に 64 分解能を持つ。これらは設定値によって様々に変更可能であるが、実際の演奏時において調整を行った結果、現在の設定値としている。これによりピッチベンドだけでなく、Mordent 表現が比較的容易に演奏可能となった。

7.4 音域の設定

楽器演奏経験を持つ体験者の多くが本システムで記譜を行う際、音部を確認していた。本システムでは OCR 機能でト音記号とヘ音記号を選択可能であるが、ハ音記号を実



図 15 現状の操作方法だけで和音に対応した場合、このような二分音符の音長を再現できない

Fig. 15 Our system cannot apply this kind of half note with our interaction method.

装していなかったため、体験者から実装の要望を得た。加線についても、第 1 線までだけでなく、それ以上の第 2、3 線を記譜するユーザが多くいたため、今後対応予定である。

7.5 調の設定、転調

現在までの展示や演奏において、事前に調の設定をすることはあっても、転調を有効に利用してこなかった。今後は演奏パフォーマンスにおいて、コード進行等の代用として、転調を利用する予定である。

7.6 その他

手持ちのスキャナデバイスを譜面上に接地することで、デバイスの覆い部分によりユーザが記述した譜面部分が遮蔽されてしまう。そこで本デバイスのカメラ画像による表示面積を実際に計測したところ、 $21 \times 32 = 672$ [mm²] であり、デバイスの覆いによる譜面遮蔽面積（デバイス覆い部分の面積）は覆い部分の直径が 55 [mm] であることから、 $22.5^2 \times \pi \approx 1590$ [mm²] となり、ユーザが譜面上にデバイスを接地した際、デバイスが遮蔽した面積のうち、カメラ画像によって $672/1590 = 42\%$ の面積しか見えていないことになる。しかし実際の展示の中で、遮蔽部分に関する意見はあげられなかった。これは、演奏をする際は画面上のカメラ画像だけを見ているため、譜面上の遮蔽箇所を気にすることがなかったと考えられる。

8. まとめと課題

楽譜を極力簡略化することで、ユーザが手書きした譜面をスキャニング操作により再生可能なシステム Gocen を提案した。従来の OMR 領域において、複雑な楽譜に対してもある程度の認識率を実現できていたが、これらを手書きや、インタラクティブな操作に実現するためには、依然として認識率や処理速度に対して困難な箇所が多々あった。そこで本研究では楽譜を簡略化することで認識精度を大幅に向上させ、30 fps のカメラ画像速度を維持できた。画像更新以外の処理回数は毎秒 150 程度（測定コンピュータスペック：1.6 GHz Intel Core 2 Duo, 4 GB メモリ, MacBook Air 11-inch, Late 2010）であるため、カメラ更新速度の向上で、さらなる操作性向上が見込める。複雑な記譜ができないことで制限されてしまう演奏表現については、対話的手法によってその問題のいくつかを解決できた。特に OCR 機能を利用することで、様々な演奏設定を紙面記述のみで実現できた。結果として、複雑な音高表現や高速な再生をともしなう楽曲を再現することは依然として困難であるが、たとえば「カノン」や「主よ人の望みの喜びよ」等の臨時記号を用いないシンプルな楽曲をリアルタイムで演奏することができるようになった。しかしながら、さらなる演奏表現のためにはより多くの機能追加が望まれる。

8.1 今後の展望

デバイスについては、操作性の向上を狙い、より画像更新速度の速い CMOS モジュールの選定を行う。本研究目的でもある、パフォーマンス用途の楽器としては一定の完成をみたが、教育用途としては、ハードウェア面に関して多くの改良が必要である。特にモジュールの一体化を行うことで、より多くのユーザに対し本システムを体験する機会を提供できるようになる。これは今後の課題としたい。

最後に具体的な演奏手法に関して、現在予定している機能追加項目を以下に述べる。

8.1.1 臨時記号

本機能対応について、主にシャープ、フラット、ナチュラル記号について、プロトタイプ初期段階から、画素スペクトラム比較やパターンマッチング処理によって実装を行っている。しかし、本ユーザテストで示した音高認識精度と比較し、大きくその精度が劣っているため、現段階においては、システムに組み込んでいない。今後は SVM 等別の認識手法を試し、精度を向上させることを予定している。

8.1.2 和音

現在和音を再現する場合、録音機能を利用して1音ずつ重ねるほかないが、本機能を実装することでより簡単に多くの音符を同時に再生可能となる。しかし図 16 に示すような、2度や1度による和音を読み取る際、本システムにおいてこれを同時に再生することは困難である。左から再生する場合は B が、右から再生する場合は A が先に再生されてしまう。

8.1.3 録音テンポと拍子の設定

シーケンス機能は、ユーザが1人で楽曲を演奏する際には必要不可欠の機能である。現状では可変小節長を自由に録音できるが、基本的なテンポ設定や拍子の設定が未実装である。様々な楽曲に対応するために本機能を実装する。

8.1.4 既存記譜にはない演奏表現

ここまで、既存の記譜法を参考にいくつかの設計を行ったが、近年のデスクトップミュージック等はこれら要素のほか、様々な効果を表現できる。パンやリバーブ、バンドパスフィルタやイコライザ、フェーザやフランジャといったエフェクト等を利用することで、より多種の演奏表現が可能となる。これらの操作設計については今後の課題とし



図 16 本システムでは同時発音が困難な和音例

Fig. 16 An example of a chord that is difficult to play at the same time on our system.

たい。

参考文献

- [1] 兩宮聡子, 金子敬一: 音高と音価に着目した読譜学習システムの設計と実現, 情報処理学会研究報告, コンピュータと教育研究会報告, Vol.2006, No.46, pp.7-14 (2006).
- [2] 千国岳広: ドレミファトレーナー, 入手先 <<http://www.vector.co.jp/soft/win95/edu/se420733.html>> (参照 2012-10).
- [3] みなぞう: おとめちゃん, 入手先 <<http://www.vector.co.jp/soft/win95/edu/se314997.html>> (参照 2012-10).
- [4] ドナルド・ノーマン (著), 野島久雄 (訳): 誰のためのデザイン? (1990).
- [5] Gould, J.: Doing the same work with hardcopy and with cathode ray tube (CRT) computer terminals, *Human Factors*, No.26, pp.323-337 (1982).
- [6] 小清水実, 津田大介, 馬場和夫: 電子ペーパーに求められる形態的特性の研究 (特集: 紙面とディスプレイ画面に関わる先端技術動向), 映像情報メディア学会技術報告, Vol.25, No.10, pp.19-24 (2001).
- [7] 岡野 翔, 面谷 信, 中田将裕, 前田秀一: 読書作業性に対する電子媒体呈示条件の影響評価: 電子ペーパーをめざす読みやすさに関する検討 (エレクトログラフィー特別研究会) (特集: 紙面とディスプレイ画面に関わる先端技術動向), 電子情報通信学会技術研究報告, EID, 電子ディスプレイ, Vol.104, No.666, pp.13-16 (2005).
- [8] Tsandilas, T., Letondal, C. and Mackay, W.E.: Musink: Composing music through augmented drawing, *Proc. 27th International Conference on Human Factors in Computing Systems, CHI '09*, New York, NY, USA, pp.819-828, ACM (2009).
- [9] Anstice, J., Bell, T., Cockburn, A. and Setchell, M.: The design of a pen-based musical input system, *Proc. 6th Australian Conference on Computer-Human Interaction (OZCHI '96)*, Washington, DC, USA, pp.260-267, IEEE Computer Society (1996).
- [10] George, S.E.: Clustering on-line dynamically constructed handwritten music notation with the self-organising feature map, *Proc. 16th International Conference on Developments in Applied Artificial Intelligence, IEA/AIE 2003*, pp.93-103, Springer Verlag Inc. (2003).
- [11] Forsberg, A., Dieterich, M. and Zeleznik, R.: The music notepad, *Proc. 11th Annual ACM Symposium on User Interface Software and Technology, UIST '98*, New York, NY, USA, pp.203-210, ACM (1998).
- [12] Kassler, M.: *An Essay Toward Specification of a Music-reading Machine*, Princeton University, Department of Music (1963).
- [13] Pruslin, D.H.: Automatic Recognition of Sheet Music, Ph.D. thesis, Doctoral Thesis MIT Cambridge MA (Jan. 1967).
- [14] Prerau, D.S.: Computer pattern recognition of printed music, *Proc. Fall Joint Computer Conference, AFIPS '71* (Fall), New York, NY, USA, pp.153-162, ACM (Nov. 1971).
- [15] 青山 宏, 棟上昭男: 印刷楽譜の自動読取り, 画像電子学会誌, Vol.11, No.5, pp.427-435 (1982).
- [16] 大照 完, 松島俊明, 金森克洋: 楽譜の自動認識 (小特集: 自動演奏), 日本音響学会誌, Vol.41, No.6, pp.412-415 (1985).
- [17] 松島俊明: 楽譜の自動認識システム (wabot-2 の視覚系), 日本ロボット学会誌, Vol.3, pp.354-361 (1985).

- [18] Bitteur, H.: Audiveris: Open music scanner (2000–2012), available from <http://audiveris.kenai.com/> (accessed 2012-06).
- [19] Desaedeleer, A.: Openomr, available from <http://sourceforge.net/projects/openomr/> (accessed 2012-06).
- [20] Droettboom, M., Fujinaga, I., MacMillan, K., Chouhury, G.S., DiLauro, T., Patton, M. and Anderson, T.: Using the gamera framework for the recognition of cultural heritage materials, *Proc. 2nd ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '02*, New York, NY, USA, pp.11–17, ACM (2002).
- [21] 河合楽器製作所：楽譜カメラ (2011).
- [22] 河合楽器製作所：PDF ミュージシャン (2012).
- [23] 山本祐介, 内山英昭, 笈 康明：紙楽譜を用いた演奏メディア onnote のためのマーカレス楽譜認識の提案, No.EC20 (2011-03).
- [24] Hoerter, N. and Chatt, T.: The music wand: Real-time optical scanning of sheet music (2008), available from http://people.ece.cornell.edu/land/courses/ece4760/FinalProjects/s2008/tjc42_nah35/tjc42_nah35/index.html (accessed 2012-06).
- [25] 馬場哲晃, 青木 允：ラインイメージセンサを利用した簡易な手書き五線譜読取りアルゴリズム, 情報処理学会研究報告, EC, エンタテインメントコンピューティング, Vol.2011, No.9, pp.1–4 (2011).
- [26] Computerized recognition of hand-written musical notes, *Proc. International Computer Music Conference*, pp.110–112 (1992).
- [27] Yadid-Pecht, O., Gerner, M., Dvir, L., Brutman, E. and Shimony, U.: Recognition of handwritten musical notes by a modified neocognitron, *Mach. Vision Appl.*, Vol.9, No.2, pp.65–72 (1996).
- [28] Fornés, A., Lladós, J. and Sánchez, G.: Graphics recognition, recent advances and new opportunities, chapter *Old Handwritten Musical Symbol Classification by a Dynamic Time Warping Based Method*, pp.51–60, Springer-Verlag, Berlin, Heidelberg (2008).
- [29] Bottello, L.: Tangible user interface for music learning, 2010, available from <http://cloverleafdesign.weebly.com/> (accessed 2012-06).
- [30] Heuer, J.F.: Noteput, available from <http://www.jonasheuer.de/index.php/noteput/> (accessed 2012-06).
- [31] Levin, G.: The table is the score: An augmented-reality interface for real-time, tangible, spectrographic performance, *Proc. International Conference on Computer Music 2006 (ICMC'06)* (2006).
- [32] 岩井俊雄：音楽のチェス (1997).
- [33] Overholt, D.: The sonic scanner and the graphonic interface, *Proc. 2004 Int. Computer Music Conf.* (2004).
- [34] Zachary, L.: Drawn, available from <http://thesystemis.com/projects/drawn/> (accessed 2012-06).
- [35] openFrameworks community: openFrameworks, available from <http://www.openframeworks.cc> (accessed 2012-06).
- [36] Native Instruments, Kontakt player, available from <http://www.native-instruments.com/> (accessed 2012-06).
- [37] Ocrad – the gnu ocr, available from <http://www.gnu.org/software/ocrad/> (accessed 2012-06).
- [38] 大津：判別および最小 2 乗基準に基づく自動しきい値選

定法, 電子通信学会論文誌 D, Vol.63, No.4, pp.349–356 (1989).



馬場 哲晃 (正会員)

1979 年長野県中野市生まれ。2005 年九州芸術工科大学大学院芸術工学府修了後、2008 年九州大学大学院芸術工学府博士後期課程単位取得退学。博士(芸術工学)九州大学。2008 年 4 月より公立大学法人首都大学東京助教。所属はシステムデザイン研究科インダストリアルアート学域。ACM 会員。



菊川 裕也

1985 年鳥取県生まれ。2008 年一橋大学商学部経営学科卒業、2010 年首都大学東京大学院システムデザイン研究科インダストリアルアート学域に入学、現在に至る。インタラクティブアートおよびインタラクティブデザインの研究に従事。



串山 久美子

1986 年筑波大学大学院芸術研究科修了。2005 年東京農工大学特任准教授、JST さきがけ研究員。2007 年首都大学東京教授、現在に至る。メディア芸術およびインタフェースデザイン研究に従事。



青木 允

1980 年東京都生まれ。2002 年亜細亜大学法学部卒業。商社勤務を経て、2005 年株式会社ウイズコーポレーション入社。首都大学東京の馬場助教考案のフレクトリックドラムスの商品化を手がける。現在は主に OID (Optical Identification) を活用した商品化の開発を担当している。