

KAON(顔音) : 顔認識を利用したインタラクティブ・コンテンツ

太田高志[†] 田中潤[‡]

[†] 東京工科大学 メディア学部

[‡] 東京工科大学大学院 バイオ・情報メディア研究科

KAON: An Interactive Content using Face Detection Technique

Takashi Ohta[†] Jun Tanaka[‡]

[†]Tokyo University of Technology, School of Media Science

[‡]Tokyo University of Technology, Graduate School of Bionics, Computer and Media Science

takashi@media.teu.ac.jp, junpluse@gmail.com

アブストラクト

多人数が同時に関わることができるインタラクティブなメディア・コンテンツとして KAON (顔音) を作成した。この作品は、鑑賞者の顔を顔認識技術によりリアルタイムに取り込み、音符へと加工して画面上の楽譜に表示するものである。楽譜は繰り返し演奏されるが、表示されている楽譜は、鑑賞者の顔の位置によって演奏の間にも刻々と変化する。多人数が同時に鑑賞できるようにするため、マルチディスプレイ構成で稼働するような仕組みとした。複数のディスプレイの使用では、単に領域を拡大するだけではなく、合奏への対応など複数の異なる楽譜構成を行うことも可能とした。作品の評価として、画像認識の性能評価を行うと共に、作品内容の評価を展示会における鑑賞者の観察とアンケートにより行った。本稿では、作品の概要や機能の説明と、評価結果について報告する。

Abstract

KAON, is an interactive-art work that aims to offer a fun of collaboration and improvisation. It captures real-time images of players' faces by employing face detection technique, displays them as musical note symbols, and simultaneously plays the resulted score. The placement of musical notes alters dynamically as players move, and the tune changes accordingly. In order to let multiple-player to participate together in composing a score, KAON employs multi-display system over network for providing a sufficiently large display area and a playing space. By mapping different group of PCs to different musical instruments, it is possible to play an ensemble. In this article, we explain the concept and system design of KAON. We also give a discussion for the evaluation of system performance and the result of the questionnaire taken at an exhibition.

Keywords: Face Detection, Multi-Display, Interactive Art, Sound, Interaction, Video-capture

1. はじめに

KAON(顔音)は、鑑賞者の顔を音符になぞらえ楽譜を構成し、それに従って音を鳴らすインスタレーション作品である。この作品は五線譜が表示されている PC ディスプレイ前に立った人の顔を認識し、その顔を音符として楽譜上に表示する。そして、その時々に表示されている楽譜に従い音が奏でられるものである。音符は顔の上下左右の位置が移動するのに併せて五線譜上の位置も変わるようにしているために、それに併せて旋律が刻々と変化する。一人の顔が一つの音符に対応しているため、複数の鑑賞者(遊戯者)が同時に参加することによって旋律を構成することも可能である。我々がKAONの制作で意図したことは、鑑賞者がコンピューターとしての操作性を過度に意識することが無いようにすることである。従って、音符を利用してはいるが演奏を意図したものではない。また、鑑賞者自身が何らかの形で作品に現れるようにもしたいと考えた。大勢の人が一度に楽しめるようにするためにはマルチディスプレイを利用し、その台数構成を自由に変更できるようなシステムとすることも試みたいことであった。

コンピューターを用いたインタラクティブな作品は、Web 広告やアート作品などの多様な領域で見られ、今日ではその地位を確立したと言えるだろう。特に、コンピューターの特性を生かした人の働きかけに動的に反応するものは、絵画や彫刻などの静的な作品や映画のように受動的に鑑賞する性質のものとは異なり、状況や働きかけに応答した変化が魅力である。しかしながら、コンピューターの万能性が認知されて来ている昨今では、作品とのインタラクションにおいて過度にコンピューターの操作を意識するようなことがあると、作品自体の鑑賞よりもコンピューターの操作に人の注意が奪われてしまうことがあるように感じられる。我々は、将来的な人とコンピューターの関わりは機材としてコンピューターの存在を意識する事無く、その機能のみが生活の様々な場面に組み込まれて現れることを予想し、模索している。例えば、カクテルを飲む場を想定した作品の“bAR-Table”は、テーブルの上にカクテルのグラスが置かれるとグラス近辺のテーブル表面にカクテルや原材料に関する情報が提示されるものである[1]。この作品では、そのような場で自然に行われる「グラスをテーブルに置く」という行為が情報を引き出すきっかけとなっている。本作品は実用性を考えたものではないが、上の例と同様にインタラクションが意識的なコンピューターの操作とはならないものとしたと考えた。そこで、インタラクションを起こすきっかけとして採用したのは、人が作品の方を見る行為や、見ている状態である。顔認識技術を利用することにより、作品に顔を向けると顔の存在を認識して反応するようにした。また、顔の認識を入力のきっかけとするだけでなく、認識した顔のリアルタイム映像を同時に作品の表示素材としても利用している。自身の映像が作品に取り込まれることで鑑賞者の作品に対する興味の増大が期待できるとともに、リアルタイムなインタラクティブ性を感じることができるようになるだろう。また、同じような考えから、作品内容がゲームのように達成すべきゴールが設定された上で意識的に関

わるものではなく、場を創る環境の一つとしてそこにあるようなものを目指した。

システムは、ディスプレイ一台の構成でも基本的な機能が実現されるものとなっている。しかしながら、多人数が同時に参加することや、楽譜の形成をメタファーとして扱っていることから、作品と鑑賞者のインターフェースとなる画面の規模を大きく(長く)することが必要であると考えた。そこで、複数のPCがネットワークを通じて連携するマルチディスプレイの環境を利用できるようにソフトウェアの拡張を行った。マルチディスプレイ・システムは主にビジュアライゼーションの研究分野において、大領域で高精度なディスプレイを実現することを主目的としたタイル・ディスプレイやウォール・ディスプレイと呼ばれるものが多く研究されている[2-4]。我々は、表示の拡大だけではなく、ディスプレイの動的な構成変更や、ユーザーとのインタラクションを含めてより豊かな表現形態を模索できるソフトウェア環境を考案したいと考えてきた[5-7]。本作品はそのような試みの一環として、具体的な課題を認識するためにも重要であると考えている。

本論文では、作品コンセプトとシステムの実装について説明し、作品としての機能面と内容面の評価について報告する。

2. 関連研究, 作品

本作品が関連する従来の研究や作品を以下に紹介する。鑑賞者の顔や体の存在に対して反応をするコンセプトは、インタラクティブアートと呼ばれるものが作成されるようになった初期のころから試みられている。Krugerの“Videoplacé”[8]は、カメラで取り込まれた鑑賞者の輪郭を塗りつぶした映像がスクリーン上に同時に投影され、鑑賞者の動きによってコンピューターグラフィックスによる抽象的な図形とインタラクションを起こす作品である。以来、鑑賞者自身の体や顔を入力として反応する作品は他にも製作されており、“Oops!”という作品[9]は鑑賞者の映像をスクリーンに写しその周辺に様々なグラフィックス映像を表示するものであるが、カメラで取り込まれた鑑賞者の画像そのものが表示に使用されている。このように、自身の姿や形を作品中に取り込むことで、鑑賞者に作品へのより大きな興味を持たせ楽しませることが期待できる。特に顔を利用したものとしては、例えばウェブの広告である“Who is the hero of the world”[10]では、画像をアップロードすることで広告として流れる映像の一部に閲覧者の顔画像がストーリーに反映される形で現れる。このように、自分自身が作品に取り込まれて表示されることは鑑賞者に新鮮な驚きを与えるだろう。インタラクティブな作品では、顔認識の技術を利用し来場者の顔を素材として利用するものがある。D’Souzaらの作品[11]は、鑑賞者の顔画像自身を画素としたフォトモザイクを作成するものである。また、“Face to Face”[12]と“15 Seconds of Fame”[13]という作品はどちらも顔認識を用いたメディアアート作品である。前者は、鏡のような正面の視点から撮られたのではない自身の顔が表示されることで、普段見ることのない自分の顔を再発見するという試みである。後者は、場に居る人たちの顔を認識し、

その画像を絵画風に加工したものを、ランダムに一人ずつ順番に肖像画の展示風に表示するものである。

他の側面における関連研究として、養毛らの人の影を利用した作品[14]は、鑑賞者が特に操作をするのではなくその場にいるということ自体がインタラクションを引き起こすという点で本作品と同じような試みと言えるだろう。また、顔の認識から音に結びつける作品としては、Valenti らの”Sonify Your Face”[15]がある。これは、表情の変化を顔認識の技術によってパラメーター化し、それによって音を生成する作品である。遊戯者は音を鳴らすために様々な表情を作ることになる。

顔認識の技術を用いている多くの作品は、作品中の表示要素に利用する顔の画像を切り出すために顔認識を行っている。このような作品では、一旦顔が認識された後にはそれが加工された結果が提示されるのを見ることで一つのプロセスが完結する。いわば一方的なインタラクションを繰り返すものと言えるだろう。一方、最近のゲーム機などでコンテンツ内での操作を行うために遊戯者の体の認識を、顔もその一部として行っているものがある。制御のためには、体や顔の位置や動きの方向を、コントローラーや矢印キーの上下左右方向に割り当てるような対応がとられることが殆どである。本作品では、顔の認識をコンテンツの制御に利用することは同じであるが、その位置や動きを一旦コントローラーの操作にマッピングし、そのメタファーによって制御するような二重構造にしているのではなく、楽譜上の音符というメタファーによって顔の位置が直接意味を持つようになっていることがそのようなアプローチとの違いである。顔が認識されただけで楽譜上のいずれかの位置に音符として配置され、作品の入力として充分なものとなる。顔の位置は意識して動かさずとも良いし、意識的に音の位置を決定するように動くこともできるが、そうした意識的な働きかけがなくても作品として成立する。

また、顔画像を表示に利用することにおいても、位置の特定のためにリアルタイムに認識を行っていることを利用して動画として画面に現している。音符として加工された表現となっているが、インタラクションのリアルタイム性が表情の変化として表示にも反映している。

まとめると、本作品では、顔認識をリアルタイムに行い、その結果を同時に表示と制御のために利用しており、またそれらの利用が、楽譜上の音符というメタファーを利用することで、自然に一体化されていることが特徴である。また、本作品は、鑑賞者自身の顔のリアルタイム画像を作品に取り込んで使用することで作品への自己の関わりを明確に意識できるものでありながら、演奏と言う程の強い目的意識を持たない作品として仕上げている。作品を媒介として同時に参加している鑑賞者の間の連携の意識がなんとなく感じられるようなコミュニケーションの場を創るものとしての期待もしており、そうした多人数が同時に関わる「環境」としての位置づけが、他の作品と異なる部分と言えるだろう。

3. 作品コンセプト

本作品の作成にあたって、前章で記した研究背景に従い、以下の三つの項目をデザインの指針とした。まず、操作のためにキーボードやマウス操作をしたりコントローラーなどの使用をしたりすることで、コンピューターを操作していることを強く意識することが無いようにすることである。本作品はその場の背景として存在し、人がただ近づいたりその前に立ったりすると反応するような作品を目指した。もう一つは鑑賞者自身が作品に反映されるようなものにするということである。具体的には、鑑賞者自身の顔が作品内に映し出されることによって、自身の作品への関与を認知できるようにした。最後に、ウェブ上のコンテンツやTVゲームなどのPCとモニター一台に對峙して行うアプリケーションとは異なり、人々が行き来する空間に環境の一部としてあって、多くの人が順番を待つことなく好きなタイミングで作品に関わることを可能であるものにするということである。同時に複数の人が関わることを想定すると、作品の体験が個人的なものとなるのではなく、個々の行動が連携して、作品を通じて他を意識することがあるようにもしたいと考えた。

以上のような設計指針を基にして、鑑賞者自身の顔を楽譜の記号として取り込み、それによって作成された楽譜に従って音が演奏されるという案を考えた。基本となる作品コンセプトは図1に示すように、作品の前にたつ鑑賞者の顔をカメラで認識し、その画像を音符として楽譜に表示してその楽譜に従って音が奏でられるというものである。鑑賞者の顔をインタラクションのきっかけとして使用した理由は、作品（の方向）を見るところという行為を捉えて反応することを考えたからである。作品を見ると必然的に顔が作品の方を向くため、それをきっかけとすることによって、作品の存在を認識してからそれに対する働きかけを行うような意識的な段階を踏むのとは違った側面を創り出すことができると期待した。つまり、人の方から能動的に働きかけることが無くとも反応が起これ、逆に、その反応が初めて作品に気づき興味を抱かせるきっかけとなることである。

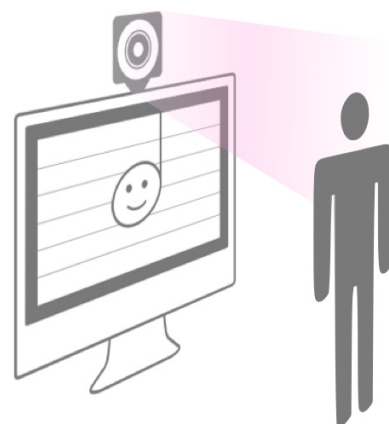


図1. 作品の基本概念

作品表現としての面白さは、自身の顔がリアルタイムに音符として作品中に表示されることと、顔と作品の相対的な位置関

係という単純なインタラクションのみによって音程や音量などの変化が起きることや、多人数で同時に遊ぶことによって複数の音による広がり生まれることにあると考えている。楽譜を形成するのは音符のアナロジーを活かすための側面が主な目的であり、実際にメロディを創りあげることや演奏することを重要な要素として考えているのではない。

4. システム構成

4.1 システムの概要

前章で述べた作品のアイデアを実装するにあたり、システム的设计を以下のように行った。システム全体の構成を図2に示す。

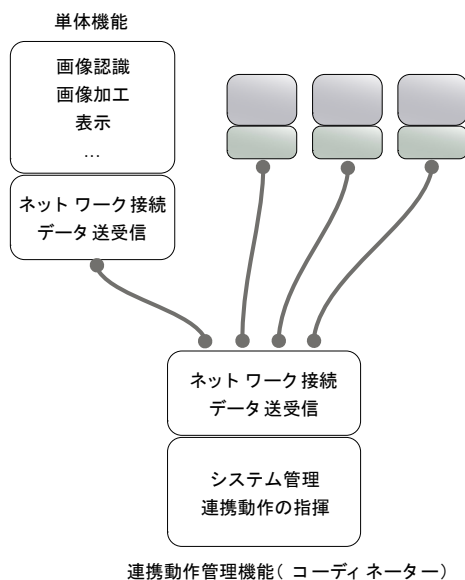


図2. システム全体の構成概念図

システムは大きく分けて二つの機能要素があり、それぞれ独立した実行プロセスとして実装している。一つは、顔を認識し楽譜として表示してそれを演奏するという単体としての機能である。図3に示す画面は実行時のイメージである。鑑賞者の顔がそのまま画像として取り込まれ、音符の一部として加工されて表示される。その場に同時に複数の鑑賞者がいれば人数分の顔が、同時にかつそれぞれの位置関係に対応した楽譜上の位置に表示される。楽譜は一定間隔でループ状に繰り返し演奏されるようになっており、その時々タイミングで楽譜上に配置された音符に従った音が鳴る。もう一つの機能は、今述べた単体の機能を連携し、マルチディスプレイとしての表示を行うためのネットワーク処理や管理のための機構である。この機能が、複数の単体機能の登録管理や実行において必要な管理信号を送信し、全体でマルチディスプレイとしての連携した動作を実現している。それぞれの機能要素について以下の節で詳しく説明する。

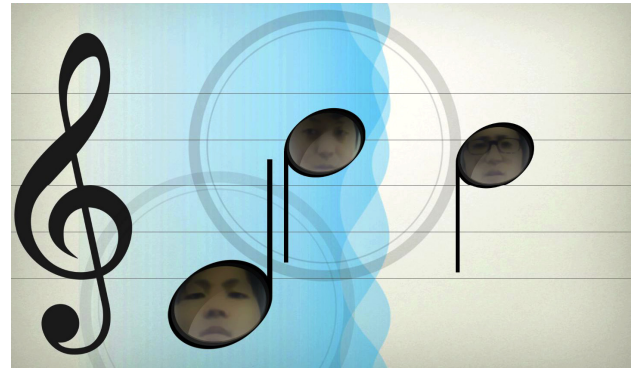


図3. 実行表示画面

4.2 基本表示機能

PC単体で実行する、KAONの基本的なインタラクションを提供するプロセスでは以下の処理を行う。

1. カメラで画像を取り込み、顔認識を行う
2. 認識された顔を画像として切り出す
3. 顔画像から音符の表示を作成する
4. 顔に位置により、楽譜上に音符を配置する
5. 楽譜上の音符に従い音を鳴らす
6. 音と同時に画像的な効果を表示する

上記項目の1から3の、鑑賞者の顔を認識し楽譜として加工して表示するまでの処理の流れを図4に示す。まず、カメラでリアルタイムにキャプチャーしている画像より人の顔を認識する。その結果から顔の部分を選択し、音符の形でマスクした画像をさらに切り出す。その間、作品の表示画面の質感となじませるために、そのままの画質ではなく多少画像処理による画質の加工を行っている。最後に、音符の形に抽出された画像を音符の形に合わせて楽譜上に配置し、表示している。

作品と遊戯者（の顔）の位置関係には上下左右方向と、奥行き方向が存在する。上下左右の位置関係は、楽譜上の音符の場所として反映される。すなわち、上下に移動することによって、音符が示す音程が変化することになり、左右に移動することで楽譜中の位置が変わる。一方、奥行き方向に顔を移動させることは、キャプチャーされる画像における顔の大きさが変化することとなる。これは表示される音符の大きさとして反映され、音の大きさを変化させる。

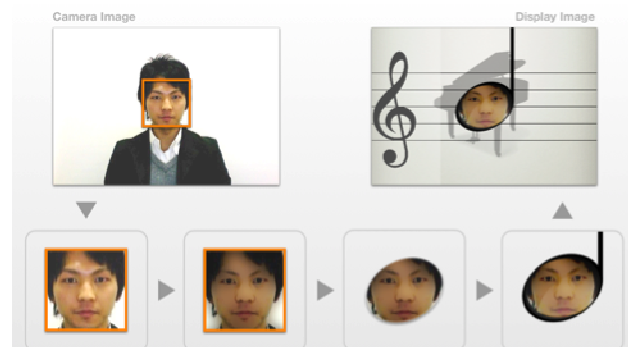


図4. 音符画像の合成手順

音を鳴らすにあたっては、画面の左端から右に向かって一定の時間間隔で音符の有無を調べる走査線が流れ、その線の一定の間隔内に音符が存在することを認識した時点で対応する音程の音を鳴らしている。走査線は、表示画面の右端についたところで左端に戻り処理を繰り返す。走査線が過ぎた間に音符が移動していれば、その都度鳴らされる音は対応して変化することになる。音を鳴らすタイミングは、音符の横方向の座標値と、音符の有無を調べるための走査線の位置座標が一致したときに行うのが基本的な考え方である。ただし、完全に座標が一致したときを認識しようとする、処理実行の時間刻み毎に走査線が移動する距離が有限値であるため、音符の位置を走査線が飛び越えることが頻出する。そのため判定に際して走査線の前後に一定の範囲を設け、その内部に音符があるときに音を鳴らすこととした。図5は、音符の横方向の座標（x座標）を0としたときに、それより以前の位置に走査線が来た時点で音符を認識している場合を示している。音程については、座標の上下方向の位置と、画面上の楽譜の線の位置関係から決定するが、これについても個々の音程について画面上下方向の一定の範囲を割当ててことで、無段階の音程ではなく、音符で表すことができる音だけとなるように量子化をしている。ただし、表示位置を量子化した結果に合わせて調整することは行わないこととした。音源としては楽器の音のオーディオファイルを利用し、音階毎に再生している。

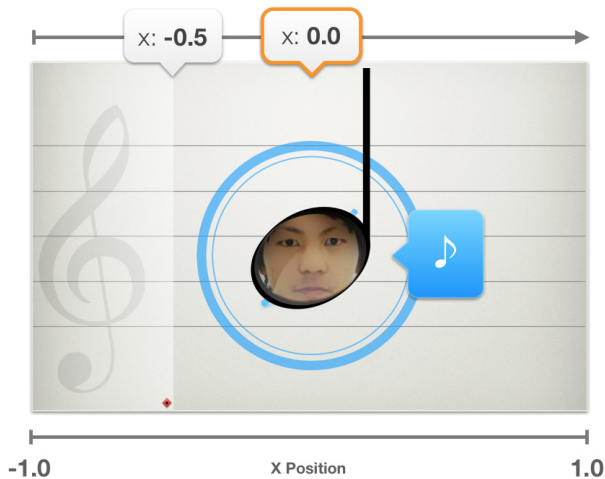


図5. 音を鳴らすタイミング

最後に、楽譜中のどの部分が演奏されているかを鑑賞者に分かるようにするため、音が発せられるときに合わせて、音符の周りに水面の波紋に模して円が表示されるようにした。これによって、どの音符が演奏されているかを視覚的に確認することができる。

4.3 ネットワークによる連携表示

一つのディスプレイだけでは、楽譜の表示としては短いため、まとまった旋律を構成するような発想へと展開しにくいと考えられる。また、複数人が同時に作品と関わる必要からも、そのインターフェースとしてより広い表示画面が必要である。そこ

で、複数のディスプレイを隣接して配置し、それらの表示動作を連携することにより楽譜の表示面積を拡大することを考えた。そのためには前節で説明した単体の機能をネットワークで連携し、楽譜の演奏が複数のディスプレイをまたがって連続に行われるようにする必要がある。すなわち、前節で述べたような楽譜上で音符を走査する処理が、複数のディスプレイをまたがって実行されることが必要である。そのために必要な機能として、無線ネットワークによる通信機能と複数のプロセスの連携管理機構を用意した。

マルチディスプレイを実現するためのアプローチについては、我々はマルチディスプレイのための専用プログラムを全体として用意するのではなく、まず単独で動作する独立したプロセスを用意し、それらが情報をやりとりすることで動的にマルチディスプレイ環境を形成する方法を追求して来た[2, 3]。そのためのプロセスの設計指針は、全体の表示領域の情報や位置関係を共有するのではなく、互いの位置関係を境界条件として扱うことと、個々のディスプレイにおける音符や走査線などの表示オブジェクトはその担当プロセスがローカルに責任を持って処理をすることである。今回の作品では、情報の送受信を管理する中心制御機構（コンダクター・プロセス）を用意したが、上述の設計指針を踏襲している。すなわち、個々のプロセスは単独でKAONの機能を実行するものとなっており、その連携は演奏実行制御の走査線の受け渡しを境界で行うことで実現している。

マルチディスプレイによるシステムを形成する具体的な手順は、まず単体プロセスが互いの認識を行う必要がある。その管理のために、前節で述べた基本機能の実行するプロセスとは独立した、連携管理のためのコンダクター・プロセスを用意した。各システムは、公開された通信ポートを自動的に発見できる Bonjour プロトコル[16]を利用して、プロセスが起動されると自動的にコンダクター・プロセスに登録される。演奏される順番については、発見されたプロセスの物理的な配置を認識して、手で演奏される順番に相当するディスプレイの相対位置情報を設定する必要がある。連携したシステムとしての動作は、コンダクターシステムから登録されたプロセスに対して、動作を制御する信号を送ることによって実現する。連携動作の実現には、並んだ画面を担当するプロセスの間で楽譜走査線の受け渡しタイミング情報などの送受信が必要であるが、本作品では、通信プロトコルとしてOSC (Open Sound Control) [17]を利用し、表1に示すような種類の情報の通信を行っている。

表1. ネットワークで送受信する情報例

	アドレス	型	候補値
再生/停止	/play	Boolean	TRUE, FALSE
音符の記録開始/停止	/recording	Boolean	TRUE, FALSE
演奏速度	/bpm	Int	50, 100, ...
演奏楽器	/instrument	Int	0, 1, 2, ...
小節番号	/section	Int	0, 1, 2, ...
音符を転送する譜面端末の追加	/link/add	String	My-Computer.local
音符を転送する譜面端末の削除	/link/remove	String	My-Computer.local

本作品は、個別のPCで動作する基本機能のシステムは、起動されると個々で独立して顔の認識と楽譜の表示を開始する。この処理は、他と連携しようがしまいが変わりにくく個々のプロセスで独立に行われるものであり、コンダクターからの制御を必要としない。連携動作は、個々のプロセスで表示している楽譜を、画面にまたがって順番に演奏する部分である。この連携の実現は、まず始めのプロセスにコンダクターが演奏開始の信号を送ることから始まる。演奏地点を示す走査線がそのプロセスの画面の最終部まで来たら、そのプロセスに演奏停止の信号を送ると同時に、次のプロセスに演奏開始の信号を送る(図6)。以降、これを繰り返し、楽譜の最後まで到達したら、始めのプロセスに信号が戻るようにして全体をループして演奏が続いていく。

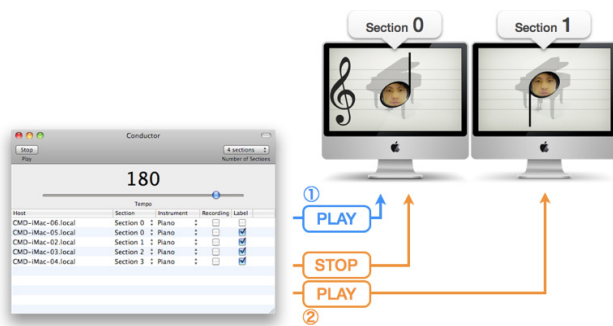


図6. マルチディスプレイ構成の制御

コンダクター・プロセスで画面の組み合わせを管理するにあたっては、参加しているディスプレイ全体を一つの連続した楽譜として機能させる以外にも、全プロセスを同数ずつに分けて、それらが合奏する構成も可能である。現在は、音色としてギターやピアノなど複数のものを用意しており、それらを個々の組にあてはめることが可能である。

5. システムの機能評価.

作品を評価するにあたって、機能面とコンテンツの内容自体についての評価を行った。機能面については、顔のカメラからの距離による認識率と、同時に複数の顔を認識し表示する場合の遅延についての性能評価を行った。評価検証を行ったシステムの仕様は表2に示すものである。カメラはPCに内蔵されているものである。

表2. 評価に仕様したシステム仕様 (iMac)

CPU	3.06 GHz Intel Core 2 Duo
GPU	NVIDIA GeForce GT 120
実装メモリサイズ	4 GB 1067 MHz DDR3
画面解像度	1920 x 1200 pixels
カメラ解像度	1280 x 1024 pixels
OS	Mac OS X 10.6.2

まず、顔の認識性能について、認識処理を行った上での描画性能を調べることで評価した。本システムの顔認識はOpenCVを利用しており、顔を画像中から検出する最小サイズをパラメー

ターとして設定できる。顔は音符として画面の楽譜上に配置するが、現在の設定では画面の幅と音符の楕円部の幅の比率が1/13のときに、ちょうど音符が五線譜の二本の線の間に収まるサイズとなっている。認識できる顔の最小サイズを画面幅の1/16もしくは1/32までの大きさまでできるようにした上で、インタラクティブな作品としての十分な反応性を確保できる描画速度の性能を考慮した結果、Webカメラの解像度が640×480、もしくは320×240とすることが適当であるとの測定結果を得た。これ以上の高解像度の設定の場合、描画性能が非常に低くなるためインタラクティブ性を十分に確保することが難しい。また、これよりも低解像度にする、検出可能な顔の最小サイズの大きさとして、五線譜のサイズに合致する1/13のサイズを確保出来ないため、それも不適切であった。

そこで、640×480と320×240のカメラの解像度設定において、顔の検出最小サイズのいくつかの設定値に対する描画性能を測定したものを表3にまとめた。また、表の一番下の段は、検出する顔がそれぞれのサイズとなるようなカメラからの距離を測定したものである。

表3. 顔の認識性能の評価

		検出できる音符の最小サイズ (音符幅 / 譜面幅)				
		1 / 4	1 / 8	1 / 13 (五線譜標準)	1 / 16	1 / 32
画像解像度	1/2 (640x480)	15.0 fps	12.0 fps	7.5 fps	6.6 fps	3.7 fps
	1/4 (320x240)	19.5 fps	14.7 fps	11.8 fps	10.0 fps	x
カメラと顔の距離		0.8 m	1.8 m	2.2 m	3.5 m	6.0 m

この結果を見ると、解像度が320×240の設定では一番小さな1/32の顔のサイズは認識出来ていない。しかしながら、このサイズを認識出来ている640×480の設定においても、3.7fpsの描画性能しか実現出来ておらず、実際の使用に適しているとはいえないだろう。実際、画面上の顔が譜面幅の1/32のサイズとなる場合は、カメラから6m離れた場合に相当するが、そこまで離れたときには、小さなサイズの音符内に自身の顔を認識することも難しいと考えられる。そのような観点からも、1/16までのサイズの認識が出来れば良いとした。実行性能も併せて考えた上で、320×240のカメラの解像度を採用することとした。

ところで、この状態で描画性能は10fps以上であるが、複数の顔認識を行う場合にはさらに性能が低下することとなり、作品の反応性に相当影響を与えることが懸念された。そこで、顔を認識する画像処理部分と、音符の描画更新をする処理を異なるスレッドに分けて実行する設計とした。その上で、同時に複数の顔認識をした場合の描画性能を測定した結果をグラフに示す(図7)。結果として、5名までであれば15fps以上の描画性能が得られた。非常に滑らかなアニメーション作品を目的としているわけではないため、それで十分な性能であると判断した。6名が同時に一つの画面で遊ぶ場合でも、10fps以上の精度が得られている。7名でも許容範囲と考えられるが、一つの画面でそこまでの人数が写るように位置をとるのは難しいと考えられ

るため、この程度の性能で作品の目的には十分と考えている。

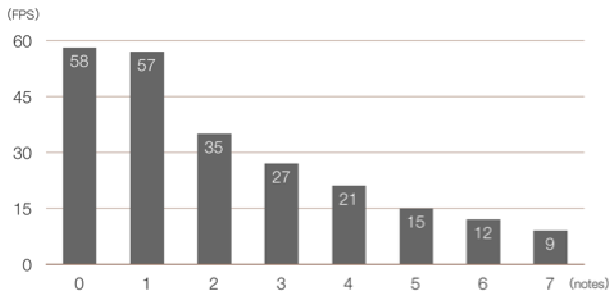


図7. 音符数による描画性能

6. 作品の評価

コンテンツとしての評価を確認するために、東京工科大学で2010年3月に行われたメディア・コンテンツ展にKAONを出展し、観客が作品を体験する様子を観察すると共に、アンケートによる調査を行った。当日のシステム構成として九台のPCを使用し、そのディスプレイを軽くカーブのかかった配置で並び、自由に観客が作品を試す事ができるような展示を行った。マウスやキーボードは置かず、ディスプレイだけが並ぶようにしてPCとしての意識が大きくなるように心掛けた。図8、9に、当日の展示の作品構成の様子を紹介する。



図8. メディア・コンテンツ展における展示（1）



図9. メディア・コンテンツ展における展示（2）

ディスプレイを円弧状に並べたのは、それらが連携して全体で一つの作品となっていることが意識されやすくするためであると同時に、一度に全画面を眺め易くすることを意図した。また、福次的効果として、一人の顔が複数のカメラに同時に認識されるようになり、少ない人数でも複数の画面に音符を出現させることができることもなった。ディスプレイの構成は、九台全体で一つの楽譜とするもの以外に、三つずつ三種類の楽器を割り当てる設定も行った。また、八台構成として四つずつ二種類の異なる楽器に割当てたり二つずつ四種類の楽器の構成にしたり、構成をいくつか変えて反応を観察した。

観客が初めに作品を見たときに、誰も遊んでいない場合は画面に楽譜が現れていない状態である。このときは楽譜（五線譜）だけが画面に表示されている。観客が作品にある程度近づけば自動的に顔が認識され画面に楽譜が現れるのだが、その際に何かが起こった事を認識しても、自分の顔が楽譜として表示されていると気づかない場合があるようであった。しかし、一旦自分の顔が音符として表示されていることに気づくと驚き、非常にその効果を楽しんでいる様子が観察された（図10）。



図10. 鑑賞する人の様子

自分の顔が音符となっていることが分かった後も、顔を上下左右に移動することで音程を変えることができることには、説明をしてから気づく人が多かったようであった。音と自身の対応は非常に簡単なインタラクションで分かりやすいものをデザインしたつもりであったが、その関連性を理解するにはある程度の間、意識的に作品と関わる必要があるかもしれない。簡単な作品解説のボードなどがあれば良かったように思われたが、一方で、何も知らない状態で作品と対峙したときの驚きを与えたい希望がより勝っていたため、今回は解説の設置は控えた。仕組みを口頭で説明された後では、皆様に顔や体の位置を動かして楽しんでいた。

アンケートは展示を体験してもらった人を対象に実施し、41名からの有効回答を得た。回答者の構成は高校生から父兄世代の大人に渡って男女が混ざったものである。質問として以下の三つの設問を用意し、表現に関連した意識調査を行った。

- 隣り合う画面に表示された譜面が繋がっていることに気づいたか？
- 譜面に表示される他の人の顔を意識して自分の顔を動かしたか？

- 他の人と協力してメロディを演奏しようと思ったか？

これらの設問に対して、図11のグラフに示す回答を得た。

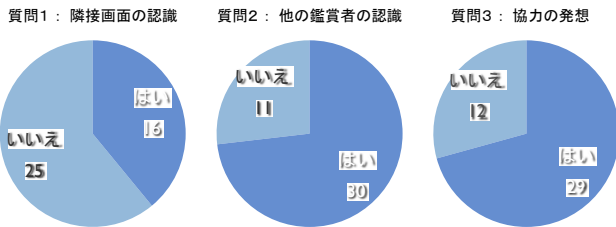


図 11. アンケートの結果

出展では複数のディスプレイで連続した楽譜を表す形態にしたのだが、アンケートの結果からは、それに気がつかなかった人が多数を占めた。一方で、他の鑑賞者が存在していることを意識し、他の人と協力してメロディを形成することの可能性については半数以上の人が気づいたとの結果となった。複数のディスプレイが楽譜として連続していることを気づきにくかった要因としては、画面で楽譜を操作していることを示す表示効果が淡く、気づきにくいものだったことがあるように思われる。個々の音符に対応する音が鳴る際には、その音符から波紋に見立てた円が表示されるのだが、連続して流れていることを意識させるには弱かったのかもしれない。他の鑑賞者と組んで特定のメロディを演奏しようとする試みは、見学に来たのが2、3名の少人数のグループばかりであったため明示的には観察出来なかった。アンケートでは、そのような発想に気づいてはいたようであるため、大人数のグループなどでの鑑賞でどのような反応があるのか、今後観察してみたい。

アンケートでは、自由記述による作品への感想と意見も記述してもらった。それらの意見では、自分の顔が音符として表現されることについて非常に面白く新鮮に感じている人が多くあった。また、顔の位置で音程や音の位置を変えることができることも面白さを感じていた。多くの人が同時に参加できることへの好感も見られたようであり、作品のコンセプトは通じたように思われた。先に述べたように作品の内容にすぐに気づかない人もいたが、それらの人を含めて作品自体は非常に好感を持たれたことが、自由記述やインタビューの結果から伺うことが出来た。

7. おわりに

鑑賞者自身の顔を音符として表示し演奏するインタラクティブ作品 KAON を作成し、その機能を説明すると共に、展覧会での展示への反応から作品の評価と考察を行った。作品として動作的にはある程度の完成度に達し、機能的に想定したものは滞りなく動作している。作品内容としては、以下に述べるように改善の余地が大分残ってはいるが、現時点での評価を得ることによって今後の展開に結びつけて行きたいと考えている。

鑑賞者自身の顔がリアルタイムに作品中に現れるのが本作品

の重要なコンセプトであるが、本人が「人」としての表示や機能を持って扱われるのではなく、音符という記号に変換されて現れるのが本作品の面白さであると考えている。このように、顔認識を利用した試みとしては、表示と操作性の両方を、楽譜上の音符というメタファーを用いることによって同時に上手に扱うことが出来たのではないかと考えている。

音は、顔の位置が音符として変換されて演奏される。人が作品の方向を見るだけで作品が反応するが、作品内容に気づいた後は意識的に顔の位置を動かして音を変えるような積極的な関わりかたも観察できた。現状では音の変化を与えるインタラクションは顔の位置や近さによっているが、さらに、顔を傾けることで音色や音の長さをコントロールするような拡張も考えることができるだろう。技術的には表情の違いによる反応を組み込むことなども可能かもしれない。しかしながら、我々が目指す環境として溶け込むようなデジタル作品というコンセプトを考えたときに、そこまで細かな演奏性を持たせることが有効なのかどうかの疑問がある。あまりに複雑な操作は、大勢で楽しむための敷居を高くすることや、鑑賞することをも複雑にってしまう恐れがある。また、現在はただ顔が音符となり音を出すのを楽しむだけであるが、指定のメロディを実現するなどの、よりゲーム性を高めた方向とすることもアイデアとしては考えられる。しかしながら、これも上記の操作性の拡張と同様に作品のコンセプトが変わることとなるため採用していない。すぐに作品内容について気づかない人がいることも、環境としてあまり強く主張しない作品と考えたときに、それで良いのではないかも感じている。一方で、他の人との連携を認識したり、複数画面の連続性を意識したりするような工夫はさらに考察する余地があるだろう。例えば、ある程度以上近い距離に現れた音符を連桁付きの音符として提示するような表示効果を考えることができる。ただし、そもそも環境的な位置づけの作品とのコンセプトを掲げたときに、むしろ楽譜と音符というような曲を構成することが具体的に意識されるようなアナロジーではなく、もっと抽象的な表現でも良かったのかもしれない。機能と作品コンセプトのバランスは、今回のような作品展示を重ねて今後も考察していきたい。

一方、マルチディスプレイによるメディア表現という点では、単純にディスプレイを長く繋げた表示以外にも、いくつかのディスプレイずつでグループを形成し、複数のグループを異なる場所に設置した上で、互いの音符を含めて表示することなどいくつかの試みを行った。これまで、個々の独立したプロセスを連携することによってマルチディスプレイとしての動作の実現することを目指して基本機能の試作を行ってきたが、具体的な作品への適用が出来たことは良かった点である。しかしながら、マルチディスプレイを利用した表現の拡張の可能性を示すには、複数のディスプレイの使用が、作品内容の質を変えるようなアイデアに結びつくことを今後さらに追求していく必要があると感じている。また、コンテンツ作成者が、マルチディスプレイの環境を利用した作品をより簡易に作成できるような基盤を用意することも、今後の大きな目標である。

参考文献

- [1] 出島大樹, 太田高志, 商品を介したデジタル情報の提示, 情報処理学会第71回全国大会, 滋賀, 2009.
- [2] Tao Ni, et al., “A Survey of Large High-Resolution Display Technologies”, Techniques, and Applications, Proceedings of IEEE Virtual Reality Conference, 2006.
- [3] T. v.d. Schaaf, M. Koutek, D.M.German and H.E.Bal, “ICWall: a Calibrated Stereo Tiled Display from Commodity Components”, VRCIA '06, Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications, 2006.
- [4] Sungwon Nam, et al., “Multiapplication, Intertile Synchronization on Ultra-High-Resolution Display Walls”, MMSys '10, Proceedings of the first annual ACM SIGMM conference on Multimedia systems, 2010.
- [5] 太田高志, 「動的に構成可能な CG 表現のためのマルチディスプレイ環境」, 日本バーチャルリアリティ学会第 13 回大会, 2008.
- [6] 太田高志, 若林尚樹, 高橋里奈, 「メディア表現プラットフォームとしてのマルチディスプレイ環境」, 第 13 回計算工学会講演会, 2008.
- [7] Takashi Ohta and Jun Tanaka, “Automatic Configuration of Display Ordering for Multi-Display Environments”, Conf. Advances in Computer Entertainment Technology (ACE2010), 2010.
- [8] Myron W. Krueger, Thomas Gionfriddo and Katrin Hinrichsen., VIDEOPLACE—an artificial reality, CHI '85 Proceedings of the SIGCHI conference on Human factors in computing systems, 1985.
- [9] Marcio Ambrosio, Oups!, 第12回文化庁メディア芸術祭大賞受賞作品.
- [10] “Who is the hero of the world”, <http://en.tackfilm.se/>
- [11] Daryl D'Souza, Vic Ciesielski, Marsha Berry and Karen Trist, “Generation of Self-Referential Animated Photomosaics”, ACM International Conference on Multimedia (MM2007), 2007.
- [12] Yasuto Nakanishi, “Face to Face – A Media-art Using a Face Detection System and its Exhibition”, ACM International Conference on Multimedia (MM2005), 2005.
- [13] Borut Batagelj, Franc Solina and Peter Peer, “15 Seconds of Fame – An Interactive, Computer-Vision Based Art Installation”, ACM International Conference on Multimedia (MM2004), 2004.
- [14] Yugo Minomo, Yasuaki Kakehi, and Makoto Iida, “Transforming Your Shadow into Colorful Visual Media - Multi-Projection of Complementary Colors”, Conf. Advances in Computer Entertainment Technology (ACE2005), 2005.
- [15] Roberto Valenti, Alejandro Jaimes and Nicu Sebe, “Sonify Your Face: Facial Expressions for Sound Generation”, ACM International Conference on Multimedia (MM2010), 2010.
- [16] Bonjour, <http://www.apple.com/support/bonjour/>
- [17] OSC, <http://opensoundcontrol.org/>

太田 高志



1985年慶応義塾大学理工学部物理学科卒業。1990年東京大学大学院工学系研究科修了。同年日本アイ・ビー・エム株式会社入社。2004年東京工科大学メディア学部助教授, 2009年 同学部准教授, 現在に至る。マルチディスプレイによるメディア・コンテンツ, アンビエント・コンピューティングに関する研究に従事。ACM, 芸術科学会, 他会員。博士(工学)。

田中 潤



2010年東京工科大学メディア学部卒業, 同年東京工科大学大学院, バイオ・情報メディア研究科入学。メディアサイエンス専攻修士課程在学。