

CrestMuse Symposium



CrestMuse

project

Symposium

October 26th, 2006@K.G.Hub Square



Introduction ごあいさつ 06

Symposium Program プログラム 08

Panel Discussion パネルディスカッション 10

Guest Discussants ゲスト紹介 12

Outline プロジェクト概要 15

プロジェクト概要

「時系列メディアのデザイン転写技術の開発」 CrestMuneプロジェクト 研究代表者
片寄グループ:片寄 晴弘(関西学院大学)

研究概要

「演奏表現と音楽聴取」 片寄グループ:長田 典子(関西学院大学)
「音楽デザイン転写・音響信号理解に基づく音インタフェース」
後藤グループ:後藤 真孝(産業技術総合研究所)
「聴覚情報表現に基づく音楽デザイン転写」 河原グループ:河原 英紀(和歌山大学)
「数理的アプローチに基づく音楽デザイン転写」 嵯峨山グループ:嵯峨山 茂樹(東京大学)
「AIアプローチに基づく音楽デザイン転写」 奥乃グループ:奥乃 博(京都大学)

Intracative Session 29

インタラクティブセッション

後藤グループ(産業技術総合研究所)

「MusicRainbow: A New User Interface to Discover Artists」

Elias Pampalk and Masataka Goto **体験デモ**

「Musiccream:

楽曲を流してくっつけて並べて思い出すことのできる音楽再生インタフェース」**実演デモ**

後藤 真孝, 後藤 孝行(元産総研)

河原グループ(和歌山大学システム工学部)

「歌真似と声真似, STRAIGHTのモーフィングは,どこまで人間に迫るか?」**体験デモ**

森勢 将雅, 畑 宏明, 田原 佳代子, 生駒 太一, 河原 英紀, 豊田 健一(元関西学院大),

片寄 晴弘(関西学院大)

「STRAIGHTの原理とrealtime STRAIGHT,そして,応用へ」

河原 英紀

嵯峨山グループ(東京大学情報理工学系研究科)

「音楽音響信号をMIDIデータへ変換する」

亀岡 弘和, 齊藤 翔一郎, 西本 卓也, 小野 順貴, 嵯峨山 茂樹

「MIDIデータを楽譜へ変換する」

武田 晴登, 山本 遼, 西本 卓也, 小野 順貴, 嵯峨山 茂樹

「自動伴奏システム」**体験デモ**

武田 晴登, 嵯峨山 茂樹

奥乃グループ(京都大学情報学研究科)

「Drumix:ポピュラー音楽CDのドラムパート編集機能付きオーディオプレイヤー」**体験デモ**

吉井 和佳, 後藤 真孝(産総研), 奥乃 博

「Instrogramによる類似楽曲検索」**実演デモ**

北原 鉄朗, 後藤 真孝(産総研)・奥乃 博

「混合音中の歌声と歌詞との時間的対応付け」**実演デモ**

藤原 弘将, 後藤 真孝(産総研)奥乃 博

「混合音中のパートの音量を操作可能なオーディオプレイヤー」**実演デモ**

糸山 克寿, 吉井 和佳, 奥乃 博

片寄グループ(関西学院大学工学部情報科学科)

「名演奏追体験インタフェース:iFP」**体験デモ**

福岡 浩行, 酒造 祐介, 片寄 晴弘

(奥平 啓太, 豊田 健一, 野池 賢二 元関西学院大)

「音楽を体験している脳をはかる」

風井 浩志, 藤澤 隆史, 森 悠太, 山岡 晶, 八田原 慎悟, 須田 一哉(東京大),

高橋 理宇真, 長田 典子, 片寄 晴弘

「演奏表情付けシステム:Pop-E」**実演デモ**

橋田 光代, 長田 典子, 片寄 晴弘

Laboratories 研究グループ紹介 59

Members メンバー一覧 65



Introduction

はじめに

ごあいさつ

CrestMuseプロジェクトでは、メディアアートやデザインの成立前提である共通理解性、および、音楽に代表される時系列メディアの認知特性に着目し、既存事例中のデザインの転写によってコンテンツ制作を支援する方式の開発に取り組んでいます。本研究によって、アマチュア、プロの双方が使用可能な直感的な作品制作支援環境を提供すること、能動的芸術鑑賞と新エンタテインメントの形成を通じて、我が国の当該領域の競争力確保に貢献することを目指しています。

このシンポジウムでは、実際にさわって頂けるデモや実演デモを多数用意しています。楽しんで、本グループの目指す所を感じて戴ければ幸いです。

また、音楽教育学の村尾忠廣氏、作曲家の菅野由弘氏をゲストディスカッサントとして迎え「音楽の楽しみ」と題してパネルディスカッションを実施します。

お楽しみください。

片寄晴弘(研究代表者)

関西学院大学工学部教授

大阪大学基礎工学部制御工学科卒。大阪大学基礎工学研究科博士後期課程修了。工学博士。オーグス総研、イメージ情報科学研究所、和歌山大学を経て、現在、関西学院大学工学部教授。ヒューマンメディア研究センターセンター長。音楽情報処理、コンテンツ Design & Creation、心理計測の研究に従事。CREST「時系列メディアのデザイン転写技術の開発 (Project)」研究代表者。

元さきがけ研究 2-1「協調と制御」領域研究者。



プログラム 司会:内谷佐和子

13:00~14:10 オープニング~全体説明

プロジェクト概要
「音楽デザイン転写・音響信号理解に基づく音インタフェース」
「聴覚情報表現に基づく音楽デザイン転写」
「数理的アプローチに基づく音楽デザイン転写」
「AIアプローチに基づく音楽デザイン転写」
「演奏表現と音楽聴取」

研究統括:片寄 晴弘(関西学院大学)
後藤 真孝(産業技術総合研究所)
河原 英紀(和歌山大学)
嵯峨山 茂樹(東京大学)
奥乃 博(京都大学)
長田 典子(関西学院大学)

14:30~16:00 インタラクティブセッション

プロジェクトの研究内容をより身近に知っていただくために、研究者によるインタラクティブセッションを用意しています。この機会に、体験版デモ、研究員とのトークをお楽しみください。

16:30~18:00 パネルディスカッション「音楽の楽しみ」

Guest Discussants

菅野 由弘 (作曲/早稲田大学)
村尾 忠廣 (音楽認知/愛知教育大学)

ポジショントーク

研究者の立場から「音楽の楽しみ」を語る
音楽家の立場から「音楽の楽しみ」を語る
音楽学の立場から「音楽の楽しみ」を語る
「音楽を愉しんでいる」脳をはかってみると…

後藤 真孝
菅野 由弘
村尾 忠廣
片寄 晴弘

全体討論

18:30~ 懇親会

連携イベント:第67回音楽情報科学研究会
10月27日(金)同会場にて



Panel Discussion パネルディスカッション

テーマ：「音楽の楽しみ」

ポジショントーク

後藤 真孝：研究者の立場から「音楽の楽しみ」を語る
菅野 由弘：音楽家の立場から「音楽の楽しみ」を語る
村尾 忠廣：音楽学の立場から「音楽の楽しみ」を語る
片寄 晴弘：「音楽を愉しんでいる」脳をはかってみると…

全体討論

Guest Discussants

村尾 忠廣（音楽認知／愛知教育大学）
菅野 由弘（作曲／早稲田大学）

Guest Discussants

菅野由弘

●音楽家の立場から「音楽の愉しみ」を語る

音楽の「愉しみ」という表現は、様々な事柄を言い得た言葉です。良く言われる、「音楽」は「音」を「楽しむ」から音楽だ、というのは素敵な間違いで、器楽が「楽」で、歌が「曲」、つまり楽器で演奏するものが音楽で、歌が音曲なわけです。同様に、ポップスとクラシックという分類は、音楽のジャンル分け、スタイルの違いを表しているのですが、この分類によって誤解が生まれています。本来は「シリアス・ミュージック」と「コマーシャル・ミュージック」に分類し、そのスタイル、もしくはジャンルは何に属する、と認識されるべきものです。「音」が出れば何でも「音楽」である、というのも一面の事実ですが、文化や人間を背負っていないものは、音楽ではありません。特に研究者の皆様には、この「文化」の部分大切にしたい、単なる材料として扱うことは、音楽への冒瀆であることを再認識して頂きたいと思っています。

菅野由弘(かんの・よしひろ) プロフィール 作曲家

東京芸術大学大学院作曲科修了。79年「弦楽四重奏曲」がモナコ・フランス・ピエール作曲賞。94年、電子音楽「時の鏡I一風の地平」がユネスコ主催、IMC推薦作品、2002年ラジオドラマ「アウラ」でイタリア放送協会賞受賞。オーケストラ、室内楽などの洋楽器、雅楽や聲明などの日本の伝統的な楽器、コンピュータによる電子音楽素材の3つの柱を駆使して作曲活動を行っている。

主要作品は、国立劇場委嘱の雅楽、聲明、古代楽器のための「西行一光の道」(春秋社刊)、聲明とパルサー波によるコンピュータ音楽「虚空星響」(DENON)、NHK交響楽団委嘱のオーケストラのための「崩壊の神話」、NHK委嘱の、オーケストラと雅楽、能管、太棹のための「蜘蛛」など。他にNHK大河ドラマ「炎立つ」、NHKスペシャル「フィレンツェ・ルネサンス」の音楽や、サッポロビール北海道工場の環境音楽、サッポロ・ファクトリーの環境音楽なども作曲。CDは菅野由弘作品集 Resounding Sphere I「星の死」(FOCD3128), II「砂の都市」(FOCD3184), III「崩壊の神話」(FOCD3449)。



写真(c)木之下晃

Guest Discussants

村尾忠廣

●音楽学の立場から「音楽の愉しみ」を語る

～サビとサビ形式の面白さを

「暗意—実現プロセス」理論によってひも解く～

「サビのない音楽なんて山葵のない寿司のようなもの」

すし屋では山葵のことをサビと言いますし、サビを効かせるといいます。「効かせる」は“聴かせる”。ノミと槌から左手に持つノミが“飲み”から酒飲み＝左党になったように、サビを“効かせる”から“聴かせる”，すなわち“聴かせどころ”が歌のサビになっていったのでしょうか(村尾の自説)。

そのサビは、意外なことに60年代までの日本歌謡曲にはありませんでした。「いい日旅立ち」の中間部(ア、ア日本のどこかに)などサビのように思われますが、これは伝統的な唱歌形式のAA'BAという二部分形式の中間で変化する部分、つまり、英語でいうブリッジにあたります。では、今日のJ.ポップスの「サビ」に近いものが登場してくるは、いったい何時頃なのでしょう。そして、その構造とはどういうものなのでしょう。

一般的にはサビを「盛り上がる」とすることが多いのですが、それでは単純に音が高いとか、長いとか、大きいなどといったレベルのことになってしまいます。サビはサビ形式として歴史的、様式的にとらえる必要があります。ただし、それだけではサビの面白さ、高揚感が説明できません。したがって、定量分析と共に、特定の作品の特異構造を情動の理論(暗意—実現プロセス)によって定性的に説明する必要があります。今回は「サビ無し」歌謡曲から「サビ曖昧」、「サビ有り」、「完サビ」にいたるまで、私たちの「サビ」定義にしたがいがら区分、定量分析しながら、サビの面白さの分析事例を展開して説明してみたいと思います。

村尾忠廣(むらお・ただひろ) プロフィール 愛知教育大学教授

1969年 東京芸術大学チェロ科卒

1971年 東京芸術大学大学院音楽学専門課程修了 芸術学修士

1982～1983年 ペンシルヴァニア交換研究員としてL. Meyer, E. Narmourのもとで研修。認知的音楽分析、暗意—実現プロセスの理論研究にあたる。

専門: 認知音楽学, 音楽教育学

現在: 愛知教育大学教授

ISME(国際音楽教育学会)リサーチ委員会委員長(1998-2000)

ISME理事(2000-2004)

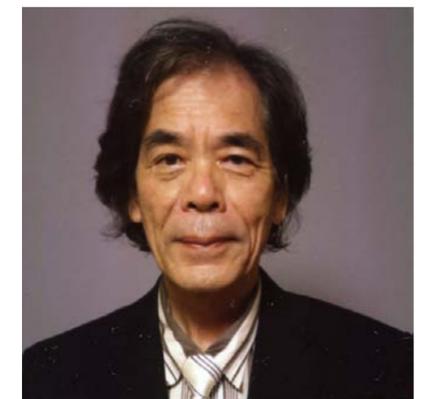
日本音楽教育学会 会長(2002-2005)

日本音楽知覚認知学会 副会長(1996-2000)

RSMEオーストラリア音楽教育学会(学会誌編集委員)

APJAEアジア太平洋芸術教育学会誌編集委員

日本音楽学会(全国役員, 学会誌編集委員 1996-2000)





プロジェクト概要

「時系列メディアのデザイン転写技術の開発」

CrestMuseプロジェクト研究代表者
片寄グループ 片寄 晴弘
(関西学院大学工学部 教授)

本プロジェクトの狙い

音楽とは、本来、時間進行の中で存在し、その場に居合わせたヒトのみが体験できる時系列芸術である。音楽を担う音そのものは時間とともに消失してしまう。音楽はヒトの記憶、進行の予測能力のもとに成立している。その性質からは、芸術の中でも主体的な要素が強い芸術であるといえる。一方で、音楽はテクノロジーの影響を最も大きく受けてきた芸術領域の一つであり、その処理や流通、享受のされ方に視点を移せば、主体的な要素とは別の芸術としての質的変遷を観ることができる。

現在、「音楽鑑賞（CD、ラジオ、音楽会、コンサート等）」は、映画・ビデオ 鑑賞、外食に続いて、現代人の余暇の過ごし方の第3位にランクされている¹。最近では、「演奏する（弾く）」ことを趣味としてあげる人も増えつつある。このように、音楽は一般の人にとって最も身近な芸術であり、娯楽の対象としても成立している。現在の「音楽鑑賞」の形態や音楽の制作や流通形態については、当たり前のものとして深く考えられることは少ないかもしれないが、この形はテクノロジーとその時代の価値基準の影響を多分に受けた結果である。iPodに代表される大容量携帯音楽プレイヤーが登場し、無尽蔵のタイトルを聴取できる状況に至り、改めてそのことを考えさせられることとなった。

長らく「音楽鑑賞」は余暇の過ごし方の代表的なものとしてランクインしているが、それには、レコードやCDを聞く、音楽放送を聞く、という聴取形態の普及が背景にある。19世紀末に、相次いでレコード、無線通信の発明がなされ、20世紀に入ってこれらの技術は一般生活にも深く浸透していった。レコード技術によって「その時に居合わせないとその音楽は聞けない」、放送技術によって「その場所に居合わせないとその音楽は聞けない」という制約が外れた。その結果、「音楽鑑賞」はいつでもどこでも誰もがができるものとなり、音楽は大量消費の対象の一つとなっていった。20世紀はとみに音楽文化が普及した世紀である。その一方で、「聴き方」の視点からは、他人の用意した音楽を受動的に聞く、あるいは、聞き流すスタイルが過分に定着した時代であるということもできる。

本プロジェクトでは、とすれば「受動的」あるいは「聞き流す」ことに傾斜しがちな音楽の愉しみを本来の主体的な形へと回帰させるシステムの開発を課題としている。美空ひばりが宇多田ヒカルの曲を歌うとどうなるか。思いついたメロディをブーニンならどのように演奏するのか。ビートルズをP.メセニーのスタッフがリミックスしたらどうなるのか。芸術作品を体験したり、鑑賞者自身の意図を反映できるようなインタフェースを利用することによって、鑑賞者は眺めているだけではとどろけない「理解」を得られるようになる。このことを言い表したキーコンセプトが「能動的芸術鑑賞」である。「能動的芸術鑑賞」の実現は、デザインの支援、新たなエンタテインメントの創出にも直結している。本研究の推進によって、我が国のコンテンツ産業の発展にも寄与していきたい。

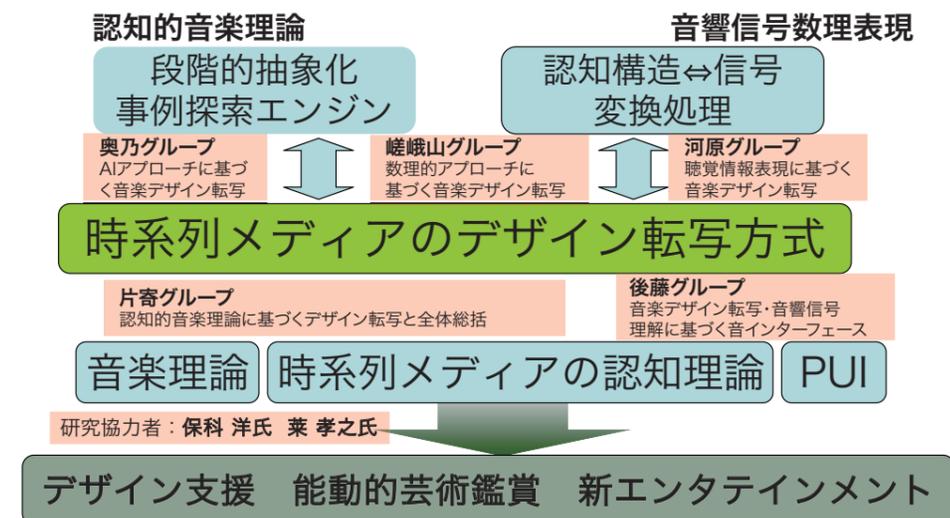
研究の進め方

芸術領域に関する研究においては「オリジナリティ」「新規性」が取り沙汰されることが少なくないが、これらの議論は様式やスタイル、あるいは、どのように感じるかの共通理解があってはじめて成立するものである。デザインに視点を移せば、その実施においては、芸術以上に「共通理解」を前提とした作り込みが意識されている。「共通理解」が形成されていく仕組みの解明はこれからの脳科学領域のグランドチャレンジの一つである。簡単に取り扱える課題ではないが、我々は、既存事例の活用、その目的に際したデザイン記述のためのデータ設計、デザインの制御・転写機構の開発によって、その一端を担っていきたく考えている。

音楽は、最終的には音の波に投影される時系列芸術であるが、作曲、編曲、演奏の表情付け、ミックスダウン、セッション、伴奏などのようにそれぞれのタスク毎に分業体制が確立している。また、それぞれのタスクレベルにおいて、例えば楽譜やMIDIのように、規範に準拠した形でのデザイン記述形式が存在する。これらの記述形式はデザイン上の意図的な逸脱（=elaboration）を記述するためのデータ記述として利用され得る。音楽には、これら客観的データ記述に加え、拍節構造、和声、フレーズなどのように、受容者や創作者のナイーブな知覚・認知にかかわる記述レベル、もしくは、抽象化構造（=reduction）が存在する。「作曲家は音楽構造をきわめて明確に意識して作曲を行い、また、演奏家は自身の理解した音楽的構造を明確化するという意図をもって演奏表現を行う」と言われている。一方、聴取者は、演奏表情の付いたメロディを聞けば、微細な時間的位置を含めて後続音が予測でき、さらに、和声やフレーズ構造が捕らえられればその確信度が上がる。つまり、音楽には、複数のデザイン記述レベルにおいて、1）未来のelaborationは、観測された当該elaborationの振る舞いと因果性を持つ、2）elaborationは、reductionの階層的な組み合わせに基づいて付与されるという性質がある。これらの性質に着目した時系列メディアの解析手法と段階的探索によるelaborationの転写モデルの構築、ならびに、音の波から音楽の抽象化レベルまでの一貫したデータの変換技術の構築が本チームの技術的研究課題である。

科学技術研究プロジェクトの多くは、核となる理論や技術を幹に据え、その高度化と応用利用という基軸で研究を実施している。核となる理論や技術を縦軸とするなら、本研究で取り扱う対象音楽は横軸に相当する。我々の取り組みにおいては、幅広い要素技術の集積と統合が不可欠となる。本プロジェクトでは、信号処理、記号処理、インタフェース、心理・脳機能計測、さらに、音楽認識、音楽制作の双方をカバーする研究チームの編成によって「能動的芸術鑑賞」の実態に迫る。

片寄グループ（関西学院大学）は、認知的音楽理論に基づくデザイン転写と心理・脳機能計測に基づく評価研究、全体統括を担当する。後藤グループ（産業技術総合研究所）は音楽デザイン転写・音響信号理解に基づく音インタフェース、河原グループ（和歌山大学）は聴覚情報表現に基づく音楽デザイン転写（歌唱特徴の転写）、嵯峨山グループ（東京大学）は数理的アプローチに基づく音楽デザイン転写（音楽解析変換と自動編曲）、奥乃グループ（京都大学）はAIアプローチに基づく音楽デザイン転写（楽器音のクラスタリングと音源分離）を担当し、それぞれが連携しつつ研究開発を実施する。



脚注 1. (財) 社会経済生産性本部「レジャー白書2006」より。「音楽鑑賞」については「音楽鑑賞」と「音楽会、コンサートなど」を合算、「映画・ビデオ鑑賞」については「ビデオの鑑賞」と「映画」を合算。

研究概要

演奏表現と音楽聴取

片寄グループ 長田 典子
(関西学院大学理工学部 助教授)

本グループの目的

情緒あふれる演奏に耳を傾け、感動に酔いしれることは、音楽の大きな楽しみのひとつである。しかしその感動の仕方は人それぞれに異なる。「このメロディが好き」という人、「このコード進行がたまらない」という人、あるいは「この演奏の間合いがグッとくる」という人など様々である。

人がどのような仕組みで音楽を認知し、音楽を楽しむか、そうした音楽の認知過程のモデル化を行うことは、音楽デザインや演奏表現を考える上で本質的な課題となる。

音楽のみならず芸術一般に共通して言えることとして、芸術作品の作り手と受け手との間のインタラクションの問題がある。音楽であれば演奏者と鑑賞者が、音楽的な構造やデザインに関してなんらかの共通理解を持つ。この次にはこの音がくるとか、この和音で解決されるといった予期や期待に添ってこそ、音楽を音楽として認知し、安心して楽しむことができる。それと同時に、予期せぬフレーズ展開や意外なリズム変化のようなある種の裏切り(デザインからの逸脱)に出会ったときにこそ、感動が喚起され演奏の醍醐味を感じるものでもある。当グループでは、演奏表現側と音楽聴取側の双方のモデル構築と、モデルに基づく演奏生成・評価を認知心理学的な視点から行っている。また音楽演奏時・聴取時の脳活動を計測して、どんな音楽要素に脳の各部位がどのように賦活(活性化)するかを観測し、音楽の認知メカニズムを解明する研究も進めている。

ここでは次の3つのデモと展示を行う。「名演奏追体験インタフェース:iFP」では、テンポと音量に関する演奏表現モデルに鑑賞者がインタラクティブに働きかけを行うもので、キーコンセプトである“能動的音楽鑑賞”のひとつの実現形を示す。「音楽を体験している脳をはかる」では、和音・調(共感覚)・リズム(太鼓)というプリミティブな音楽要素と脳活動の関係から音楽認知メカニズムの一端を解明する。「表情付けシステム:Pop-E」では演奏表現モデルの中でも複数旋律を扱える仕組みを導入し、人間に近い演奏表情付けを実現する。

1. 「名演奏追体験インタフェース:iFP」体験デモ

iFPは「手振り」という簡単な拍打で、テンポや音量などの操作を行い、指揮的な演奏表現感覚を楽しむための演奏インタフェースである(図1)。情緒あふれる演奏をテンプレートとして利用して、予測制御によってスケジューリングを実施することで、ピアニストを指揮しているような感覚を得ることができる[1]。iFP2ではさらに、ユーザ向けの情報提示など、演奏インタフェースとしての機能を強化している。

「Conducting Program(指揮プログラム)を用いることで、演奏者は指運びを気にすることなく、演奏表現でもっとも重要なフレーズ(テンポ表現、音量表現)に集中できる」とはConducting Programの提唱者であるMax Mathewsの言葉である。今回のデモンストレーションでは、今は亡きコルトーの演奏(別れの曲抜粋)のテンプレートも用意している。この機会に“演奏表現に集中する”能動的芸術鑑賞を体験していただきたい。



図1 iFPを使って演奏をしている様子

2. 音楽を体験している脳をはかる

音楽が何を表現し、音楽から人が何を感じ取るかといった音楽の認知メカニズムの解明は、音楽研究者の長年の夢である。最近の脳機能イメージング技術の発展により、音楽と認知の関係を脳活動分布から解析する研究が急速に進められている。

当グループではfMRI(機能的磁気共鳴イメージング法)やfNIRS(近赤外分光イメージング法)を用いて、音楽演奏時・聴取時の脳活動計測を行っている。この方法は血中の酸素濃度変化を測定し、脳のどの部分が特異的に賦活するかを計るものである。ここでは3つの事例を紹介する。

(1) 和音認知の心理物理モデルと脳活動[2]

長調(M)・短調(m)・減和音(dim)・増和音(aug)の4種類の三和音(転回型を含め計10個)に対する印象を、緊張性とモダリティ(長調らしさと短調らしさ)というモデルでうまく説明できることを示した。またこの時の脳活動計測では、感情を司る前頭眼窩野に賦活が見られ、和音と情動的経験の関わりを明らかにした(図2)。

(2) 共感覚の計測[3]

調性の異なる種々の音楽を共感覚保持者に呈示すると、目を閉じているにもかかわらず、色知覚に関わる紡錘状回(V4/V8)付近で賦活が見られることが確認された。

(3) 太鼓演奏時の脳活動[4]

単調なリズムより複雑なリズム(三三七拍子)を叩いている時の方が前頭部の活動が強いことが示された。また、被験者の年齢によって、脳活動傾向が異なる状況が観測された。

これらの研究事例は、音楽の愉しさという複雑な心理現象を明らかにするための基礎的知見となりうると考えられる。

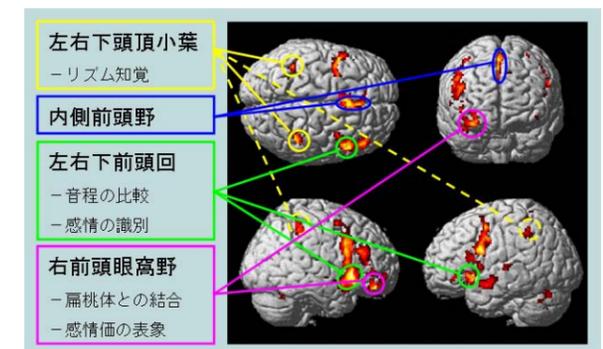


図2 和音聴取時の脳活動

3. 「演奏表情付けシステム:Pop-E」実演デモ

Pop-E(Polyphrase Ensemble)は、演奏デザイン支援を行うため音楽解釈モデルである[5]。演奏表情付けシステムの音楽表現能力は近年大いに向上してきたが、複数の旋律からなる楽曲の演奏表現を扱ったものはない。Pop-Eでは、まず各旋律(声部別)に演奏表現のルールを適用する。しかしこれだけでは、複数の声部間で演奏に必要な占有時間が異なるため、発音のタイミングを揃えたい場合にもズレが生じる。これに対し、グループ構造と演奏上のアテンションパートを手がかりにして必要に応じて同期させる処理を実施する。アテンションパートとは、目立って聞こえてくる音の並びのことで、メロディに限らず、鼻歌などで思わず口ずさんでしまう伴奏のパートやオブリガートをも含む。

3人のピアニストの演奏をPop-Eモデル上で再構築し、聴取実験を行ったところ、Pop-Eが少数のルール群で人間に近い演奏を生成しうる可能性が示唆された。

デモでは、Pop-Eモデルに基づいて生成した「幻想即興曲」を実演する。この曲はNIME-Renconにおける聴き比べコンテストにおいて第一位を受賞したものである。演奏表情付け研究の最前線をお耳にかけたい。

参考文献

- [1] Katayose, H. and Okudaira, K.: iFP A Music Interface Using an Expressive Performance Template, Entertainment Computing 2004, Lecture Notes in Computer Science, Vol. 3166, pp.529-540 (2004).
- [2] 藤澤隆史, Cook, N.D., 長田典子, 片寄晴弘: 和音認知に関する心理物理モデル, 情報処理学会研究報告2006-MUS-66, pp.99-104(2006).
- [3] 高橋理宇真, 藤澤隆史, 長田典子, 杉尾志志, 井口征士: fMRIによる共感覚の計測—色聴者の音楽聴取時の脳活動—, 情報処理学会研究報告 2006-MUS-66, pp.105-108(2006).
- [4] Katayose, H., Nagata, N., and Kazai, K.: Investigation of brain activation while listening to and playing music using fNIRS, Proceedings of 9th International Conference on Music Perception and Cognition (ICMPC2006), pp.107-114 (2006).
- [5] Hashida, M., Nagata, N. and Katayose, H.: Pop-E: A performance rendering system for the ensemble music that considered group expression, Proceedings of 9th International Conference on Music Perception and Cognition (ICMPC2006), pp.526-534 (2006).

音楽デザイン転写・音響信号理解に基づく 音インタフェース

後藤グループ 後藤 真孝
(産業技術総合研究所 主任研究員)

はじめに

CrestMuseプロジェクトにおいて、本グループでは、音楽デザイン転写と音響信号の分析・理解に関する研究を実施すると共に、それらの研究成果に基づいて具体的な音インタフェースの事例を実現することを目指す。一万曲を携帯型音楽プレーヤに入れて、いつでもどこでも手軽に持ち歩けるようになった今日、インターネットや携帯電話を用いた音楽配信も普及し、好きな曲を好きなときに好きな場所で好きなだけ聴ける時代が到来しつつある。そこでは、音楽に対する情報検索サービス等の、楽曲の中身を理解した結果やメタデータ (例えば書誌情報) に基づくサービスが求められている[1,2]。このように、音楽の音響信号を人間のように理解できる計算機システムを実現する音楽音響信号理解研究の重要性がますます高まってきており、その研究も着実に進歩している[3]。

こうした音楽音響信号理解研究の進歩の恩恵をエンドユーザが受けられるようにするためには、日常的な場面で使える音楽インタフェースの研究が重要となる。例えば、音楽家でない一般のエンドユーザは、音楽を創作するのではなく、受動的に音楽を鑑賞することが多い。そこで、我々は2003年に、従来の受動的な鑑賞とは違う、能動的な音楽鑑賞を可能にするエンドユーザ向け音楽インタフェースを実現する研究アプローチを提唱した[4]。我々はこれを、「能動的音楽鑑賞インタフェース」(Active Music Listening Interface)と名付けた[5]。これは、CrestMuseプロジェクトでの本グループの方向性を表す重要なキーワードである。

能動的音楽鑑賞インタフェース

能動的音楽鑑賞インタフェースの実現は、「音楽音響信号理解研究が音楽の聴き方をどのように豊かにできるか」を追求する研究アプローチである。こうした能動的鑑賞において重要なのは、エンドユーザが意図を容易に指定でき、かつ、その意図が適切に鑑賞体験に反映されることである。既にこのアプローチの有効性を示す実例として、以下のような、音楽音響信号理解に基づく具体的なインタフェースを実現してきた。

(1) サビ出し機能付き音楽試聴機 SmartMusicKIOSK [4]

音楽情景記述[6]の一連の研究の中で実現した、サビ区間検出手法RefrainDに基づく音楽再生インタフェースである。図1に実行画面を示す。エンドユーザが音楽地図を見ながら楽曲の再生に能動的に介入して、聴きたい部分だけを選び出すという、音楽との新たなインタラクションを実現した。文献[4]では、能動的音楽鑑賞 (インタラクティブで能動的な音楽聴取) について様々な観点から議論している。

(2) 楽曲を流してくっつけて並べることのできる新たな音楽再生インタフェース Musicream [8]

楽曲間の類似度に基づき、エンドユーザが抱く「何か聴きたい」「思い通りに聴きたい」という欲求を満たせる音楽検索・再生インタフェースである。従来から研究されている単なる類似楽曲検索でなく、様々な楽曲の中からエンドユーザが能動的に類似楽曲群を選んだ上で、それらをザッピング再生して吟味しながらプレイリストを作成することで、音楽との多様な出会い方を可能にした。

(3) ドラムイコライズシステム INTER:D [9], ドラムパートのリアルタイム編集機能付きオーディオプレイヤー Drumix [10]

音楽CDに対するドラム音検出手法に基づく音楽再生インタフェースである。INTER:Dでは、楽曲の音楽音響信号の再生中にエンドユーザが能動的に介入し、バスドラムとスネアドラムの音量を調節したり、それらの音色を好きな音色に差換えることを可能にした。Drumixではこれに加えて、楽曲の元々のドラムパターンもリアルタイムに置き換え、好きなドラムパターンでの鑑賞を可能にした。

(4) アーティスト発見インタフェース

MusicRainbow [11]

未知の音楽コレクションに含まれる多数のアーティストの中から、アーティストを表す言葉 (属性) を見ながらアーティスト間を渡り歩くことで、好みのアーティストを発見できるインタフェースである。楽曲間の類似度に基づいてアーティスト間の距離を計算することで、似たアーティストを近くに配置でき、すぐに見つけられるようにした。さらに、Web検索エンジンを用いて各アーティストに属性を自動ラベリングすることで、そのアーティストを象徴する言葉を表示できる。ダイヤル型のデバイスを回したり押ししたりするだけで、楽曲断片を聴きながら能動的なアーティスト探索をすることを可能にした。

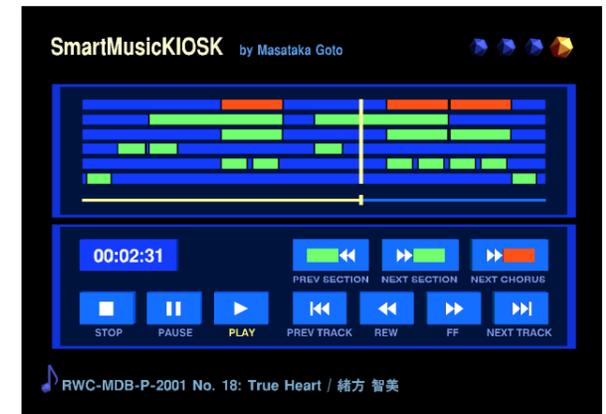


図1: サビ区間検出手法に基づく音楽再生インタフェース SmartMusicKIOSK [4]の画面表示例(RWC研究用音楽データベース[7] RWC-MDB-P-2001 No. 18に対する自動サビ区間検出結果):上側のウィンドウは楽曲内容の視覚化結果(音楽地図)を表す。横軸は時間軸で楽曲全体を表示しており、最上段がサビ区間の一覧、その下の5段が繰り返し構造、その下の横棒が再生位置スライダーを表す。下側のウィンドウは再生操作ボタン群を表し、通常のCDプレーヤ相当のボタンに加え、サビ区間の頭出しをする「サビ出しボタン」等を含む。ユーザはボタンひと押しでサビを聴くことができ、音楽地図を見ながら能動的に聴きたい場所を探することができる。

おわりに

より高度な音楽音響信号理解の実現と、それに基づく能動的音楽鑑賞インタフェースの実現は、いずれも重要な課題であり、本グループの研究の両輪である。今後も、他のグループと連携しながら、世の中に貢献できる研究を展開していく予定である。

参考文献

- [1] 後藤 真孝, 平田 圭二: 解説 "音楽情報処理の最近の研究", 日本音響学会誌, Vol.60, No.11, pp.675-681, November 2004.
- [2] Masataka Goto and Keiji Hirata: Invited review "Recent studies on music information processing", Acoustical Science and Technology (edited by the Acoustical Society of Japan), Vol.25, No.6, pp.419-425, November 2004.
- [3] Anssi Klapuri and Manuel Davy, editors: Signal Processing Methods for Music Transcription, Springer, 2006.
- [4] 後藤 真孝: "SmartMusicKIOSK: サビ出し機能付き音楽試聴機", 情報処理学会論文誌, Vol.44, No.11, pp.2737-2747, November 2003. (推薦論文)
- [5] 後藤 真孝: "音楽情報処理が実世界と結び付く: あなたも使える音楽情報処理", 日本音響学会 音楽音響研究会資料 MA2005-39, Vol.24, No.4, pp.97-100, August 2005. (招待講演)
- [6] Masataka Goto: Music Scene Description, In Anssi Klapuri and Manuel Davy, editors, Signal Processing Methods for Music Transcription, pp.327-359, Springer, 2006.
- [7] 後藤 真孝, 橋口 博樹, 西村 拓一, 岡 隆一: "RWC研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース", 情報処理学会論文誌, Vol.45, No.3, pp.728-738, March 2004.
- [8] Masataka Goto and Takayuki Goto: Musicream: New Music Playback Interface for Streaming, Sticking, Sorting, and Recalling Musical Pieces, Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005), pp.404-411, September 2005.
- [9] Kazuyoshi Yoshii, Masataka Goto, and Hiroshi G. Okuno: INTER:D: A Drum Sound Equalizer for Controlling Volume and Timbre of Drums, Proceedings of the 2nd European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies (EWIMT2005), pp.205-212, November 2005.
- [10] 吉井 和佳, 後藤 真孝, 駒谷 和範, 尾形 哲也, 奥乃 博: "Drumix: ドラムパートのリアルタイム編集機能付きオーディオプレイヤー", 情報処理学会 インタラクティブ2006論文集, pp.207-208, March 2006. (インタラクティブ発表)
- [11] Elias Pampalk and Masataka Goto: MusicRainbow: A New User Interface to Discover Artists Using Audio-based Similarity and Web-based Labeling, Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR 2006), October 2006.

聴覚情報表現に基づく音楽デザイン転写

河原グループ 河原 英紀
(和歌山大学システム工学部 教授)

はじめに

歌い手は演奏表現のために、物理信号から言語に至る様々なレベルに広がる多様なチャンネルを利用する。それぞれの曲に対し、それらにどのような表現を割り当て、演奏の進行の中で演奏履歴と競演者や客席などの反応により、それらをどのように更新して行くかが、歌い手の個性とスタイルとなる。これらをデザインととらえ、自由に好みの曲に転写し操作するという新しい音楽の楽しみ方を創り出すことを狙っている。聴覚の情報表現に基づく音声・音響処理の技術体系を構築する過程で開発された STRAIGHT と、それに基づくモーフィング技術は、この目標を実現するための鍵となる。本シンポジウムでは、まず、基盤となる STRAIGHT の技術と、 STRAIGHT を用いたモーフィングの歌唱デザイン転写への応用について紹介する。

STRAIGHT

STRAIGHT[1]は聴覚における情報表現と機能的に等価な工学的表現に基づく音声・音響処理を実現することを目標として開発されたシステムである。人間にとって重要な音の多くは周期性を有している。STRAIGHTは、この周期性を積極的に利用し、その中に織り込まれている豊かな情報を、再利用可能な形で取り出し操作することを可能にする技術である。

人間は高度に非線形なシステムである。そのようなシステムとのインタラクションでは、インタラクションに用いるメディアの質が結果を大きく左右する。音楽におけるシンボルのレベルは、音声言語における言語情報のように離散的な構造と規則により守られている。しかし、音声言語や歌唱をテキストや楽譜以上のものとしている非言語情報およびパラ言語情報は、メディアの質の影響を強く受ける。歌唱における音楽のデザインは、これらの情報における既定値からの逸脱として実現されている。STRAIGHTは、メディアの質を劣化させずにこれらの情報を操作するための手段を提供する。

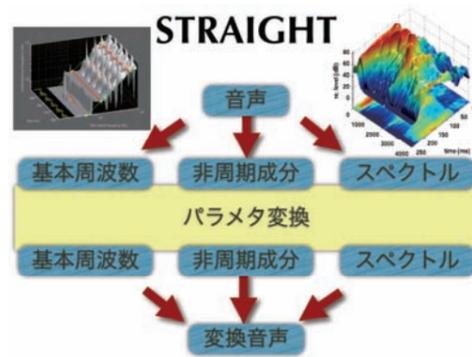


図1 STRAIGHTの概要

アルゴリズムと実装

STRAIGHT は、Channel VOCODER のアーキテクチャに基づき、音声を分解・変換・再合成する。現在の STRAIGHT の実装では、音声波形は、(1) 基本周波数、(2) 非周期成分から構成される音源情報と、滑らかに変化する (3) スペクトル包絡情報とに分解される。操作は、分解されたパラメータ上で行われる。操作された上記の3種類のパラメータから音声再合成される。相補的時間窓と群遅延に基づく処理が、高い品質の変換音声を可能とする STRAIGHT のアルゴリズムの鍵となっている。

スピーカ再生を前提とすれば、元の自然音声と比較して議論できるレベルの品質は、現在の STRAIGHT でも既に実現されている。STRAIGHT パラメータの比例的な操作と接続だけを用いて制作した2004年のRENCONへの応募作品[2]は、この基本的な品質のレベルを示す一つの例として挙げることができよう。しかし、ヘッドフォンやニアフィールドモニタでの受聴やコンテンツ制作のためのツールとしての利用を想定した場合、STRAIGHTの品質は不十分である。プロの要求に応えることのできる品質を実現するための検討が、幾つかの新しいアイデアに基づいて進められている。

STRAIGHTのアルゴリズムは、本来realtime処理を前提として開発されている[1]。幾つかの処理を省略することで、通常のPCの上でrealtimeで動作するSTRAIGHTは、既に実現されている[3]。省略部分の代替あるいは高速化アルゴリズムの開発により、実時間で動作し音楽デザイン転写に応用できる、機能制限の無いSTRAIGHTシステムを実現することが、本グループの重要な目標である。

STRAIGHTによる歌唱のモーフィング

モーフィングは、物理パラメータと心理的属性との対応関係についての明示的な知識に依存せずに、2つの試料の中間状態を容易に制作することを可能にする。また、異質の事例を組み合わせることにより、現実には存在し得ない効果を生み出すことを可能にする。これらは、モーフィングをコンテンツデザインのためのユニークなツールとする特長である。STRAIGHTは、聴覚メディアの領域において、高品質なモーフィングの利用を可能とした[4,5]。明示的な知識に依存せずに高い自然性を有する刺激連続体を作成できるというモーフィングの特長は、非分節的な構造と(議論はあるものの)非範疇的な性格を有する音楽デザインの研究と実現の手段としてまさに必要とされていたものである。

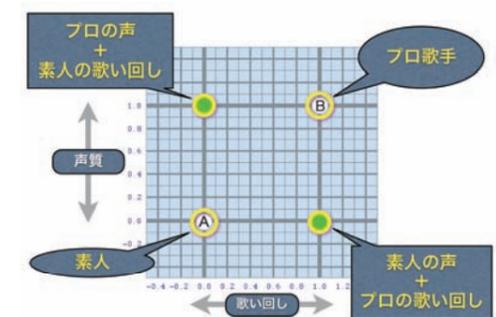


図2 歌唱モーフィングのインタフェース

モーフィングにおいては、まず、事例間の対応関係を指定する。時間周波数表現上での対応点を指定することにより、異なる事例を表すパラメータを共通の時間周波数座標の上で表現する。次いで、この共通の座標の各点において事例間のパラメータの補間・補外を行い、目標とする表現のパラメータの値を求める。こうして求められた値を、目標とする時間周波数軸上に戻し目標とする音声を合成する。対応点の指定は、共通の時間周波数座標での操作を可能にするための一つの手段として用いられている。ここで紹介する歌唱モーフィングの研究では、時間方向の対応点の生成にスケールフィルタを用いて、半自動的操作インタフェースを構築した[6]。なお、ラベル情報と個人毎の母音テンプレートを用いて、周波数方向の対応点を自動的に生成する方法についての検討も並行して進められている。

現在のSTRAIGHTによるモーフィングには、5つの独立に操作可能なパラメータがある。2つの事例のモーフィングは、一つの事例を超立方体の頂点(0,0,0,0,0)に配置し、もう一つの事例を頂点(1,1,1,1,1)に配置し、両者を等間隔にマークされた直線で結ぶことに相当する。デザイン転写の観点からは、与えられた事例間を補間・補外できるだけでは不十分である。事例を参照点として、意図した方向に演奏を変化させる(しかも、その変換によって品質を大きく劣化させない)ことが必要となる。STRAIGHTが用いている音源情報とスペクトル包絡という情報表現は、声質のような比較的低レベルの聴覚的印象とは良い対応を示す。ここではスペクトルの周波数軸方向の操作と各点での値の操作が声質の操作であると想定し、時間軸方向の操作と基本周波数およびエネルギーの操作が歌い回しの操作であると想定して、図2に示す操作空間の知覚的印象を調べた結果を紹介する。インタラクティブセッションにおいて、実際にそれらがどのような印象を生み出すか、また、実際にプロ歌手が「声真似」「歌真似」を試みた場合の表現とどう違うのかを、ぜひ体験して頂きたい。

新しい音楽の愉しみに向けて

「この曲を、あの歌手に、この人のスタイルを取り入れて、こんな雰囲気でも歌って欲しい。」という願いをかなえるシステムを実現するための道筋が見えはじめています。音源情報の表現、モーフィングの構造化、ユーザインタフェースの概念化など、実現への課題を、本プロジェクトを通して解決したい。

参考文献

- [1] Kawahara, H., Masuda-Katsuse, I., and Cheveigné, de A.: Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction, *Speech Communication.*, 27(3-4), pp.187-207 (1999).
- [2] Kawahara, H., Banno, H., Morise, M., and Hirachi, Y.: A cappella synthesis demonstrations using RWC music database, In Proc. NIME04, pp.130-131 (2004).
- [3] Banno, H., Hata, H., Morise, M., Takahashi, T., Irino, T., and Kawahara, H.: Implementation of realtime STRAIGHT speech manipulation system, [submitted].
- [4] Kawahara, H., and Matsui, H.: Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation, In Proc. ICASSP 2003, Vol.I, pp.256-259 (2003).
- [5] Yonezawa, T., Suzuki, N., Mase, K., and Kogure, K.: Gradually changing expression of singing voice based on Morph-ing, In Proc. Interspeech2005, pp.541-544 (2005).
- [6] 豊田健一, 片寄晴弘, 河原英紀: STRAIGHTによる歌声モーフィングの初期的検討, 情報処理学会研究報告, Vol.2006, No.19, 2006-MUS-64, (2006).

数理的アプローチに基づく音楽デザイン転写

嵯峨山グループ 嵯峨山 茂樹
(東京大学情報理工学系研究科 教授)

本グループの目的

音楽のデザインは、大きく分けて作曲デザインと演奏デザインから成り立っていると考えられよう。当グループの最終目的は、これらを数理的な観点から定式化し、学習理論を適用して既存の音楽からデザイン成分を抽出し、それにより音楽を解析・生成する技術を創成することにある。

その第一歩として、実際に演奏された音楽を解析する手法を確立する。CDに記録されたりインターネットを通じて配信されたりしている音楽音響信号には、曲を特徴付ける調、和音、メロディーのような作曲デザイン(楽譜情報)、演奏に用いられた楽器情報やそれらの音色情報、演奏者による強弱やテンポ、リズム変動といった演奏デザイン(表現情報)など、様々な情報が含まれている。例えば、ある曲の演奏デザイン(表現情報)を他の曲に“転写”するためには、1)演奏された音響信号を楽譜と比較することによりどのように強弱やテンポが変動しているのかを検出する、2)音高やリズムパターンと演奏者の表現がどのように関連しているのかをモデル化する、というような過程が必要であり、音響信号に混在している音楽情報を分離したり、それらを相互に変換したりする技術が必要となる。このように多彩な音楽情報を、楽譜情報、MIDIデータ(音高情報)、音響信号の3つに整理し現在の技術を鑑みると(図1)、与えられた楽譜に対してMIDIデータ、音響信号を得る過程はそれぞれ、自動演奏、シーケンサという形ですで実現されているのに対し、音響信号からMIDIデータ、MIDIデータから楽譜を得る過程はいわばその逆問題であり、未だ研究途上の段階にある。本グループでは、この音楽情報処理における逆問題に対し、数理的なアプローチ、具体的には確率モデル、統計学習、パターン認識等の手法を用いることによって合理的な解法を導出し、音楽デザイン転写に対し実用的な基礎技術を確立することを目的としている。

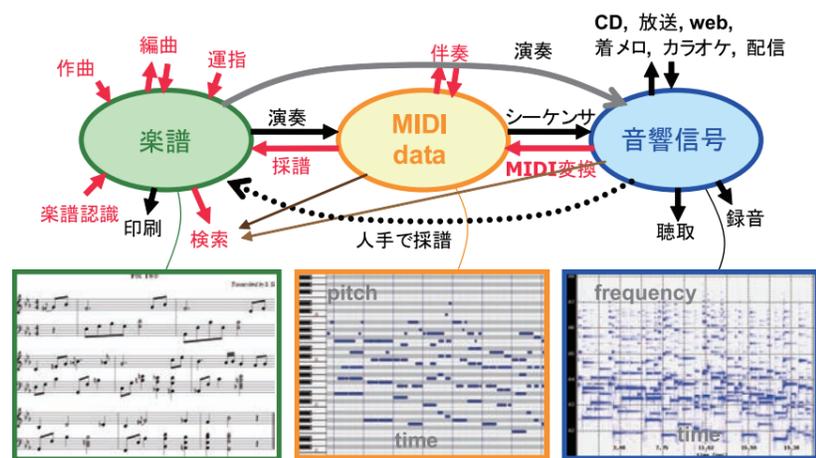


図1 音楽情報間の変換概念

1) 音楽音響信号からMIDIデータへの変換

音楽音響信号からMIDIデータへの変換において解くべき問題の中心は、音高分布の抽出である。時間的に変化する音高(基本周波数)情報の抽出には、スペクトログラム(時間周波数)表現が有効と考えられるが、音楽音響信号は楽器の音色を構成する豊富な倍音成分を含むため、得られたスペクトログラムの各成分を基本周波数と倍音成分に分離することは容易ではない。また、これらの楽器の倍音構造(調波成分のパワー比)は通常は未知で、分析時に用いることができないことや、対象とされる楽曲のほとんどが多旋律であり、音楽的な調和性のためにそれらの倍音成分が重なり合うことが多いことなども、この問題をさらに難しくしている。

我々はこの問題に対し、複数のアプローチで取り組んでいる。1つは音声認識などに用いられている混合正規分布の学習法の利用である。具体的には、観測されたスペクトルを調波構造の拘束をつけたGMMによりフィッティングするものであり、我々はこれを“Harmonic Clustering”と呼んでいる[2]。この手法はEMアルゴリズムにより効率的に解くことができる他、確率的な枠組みであることを利用して情報量規準による音源数判定[3]などが可能になる利点がある。また、このアプローチは近年さらに進展し、音楽音響信号の時間周波数領域上の2次元エネルギー分布に対して、周波数方向の調波構造と時間方向の滑らかなエンベロープを同時に表現する拘束付GMMをフィッティングすることにより、高精度な多重ピッチ解析を実現する“Harmonic-Temporal Clustering”という手法に発展している[4]。もう1つは、調波構造が音高に依らないという仮定の下では、対数周波数軸上のパワースペクトルが基本周波数分布と倍音構造の畳み込みでモデル化できることに着目し、対数周波数軸上の逆畳み込みにより倍音成分を抑制する手法であり、“Specmurt”と呼んでいる[5]。この手法では未知の調波構造をどのように設定するかが重要となるが、事後確率を最大化することによりブラインドでこれを推定する技術も整備されつつある[6]。

2) MIDIデータから楽譜への変換

音高情報が正確に得られても、これを楽譜に復元するためには、調性、音長、テンポ等を復元しなければならない。例えば音長は、演奏に表情を付けるためのテンポやリズムの変動などを反映して連続的に伸縮するため、量子化のような単純な対応関係だけでは、演奏された音の長さから音符の音価(記述された音符の長さ)を復元することはできない。我々は、音価の組み合わせ、すなわちリズムパターンには出現頻度の偏りがあること(例えばシンプルなパターンは楽譜に現れやすいなど)に着目し、これを音声認識と同型の問題ととらえた新しいアプローチを試みている。すなわち、連続音声認識における音響モデルに対応して隠れ状態を音価、状態出力を音長とする隠れマルコフモデル(HMM)を適用し、言語モデルに対応して音価の連鎖のリズム語彙あるいは音価のn-gram連鎖モデルを適用する手法であり[7]、市販のシーケンサソフトの量子化機能を上回る性能を得ている。この手法は現在、テンポも未知の場合に拡張され、尤度関数を反復的に最大化することによりリズム・テンポの両方を同時推定する手法へと発展している[8]。一方、調認識の問題に対しても、和声進行をHMMによりモデル化することにより、高精度で、転調の検出などへも拡張可能な枠組みを試みている[9]。

この技術の別の側面として、MIDI楽器演奏に実時間で合わせて伴奏譜を演奏する自動伴奏の技術へ応用でき、弾き誤りや部分飛ばしや繰り返しなどを確率現象として許容できる。

参考文献

- [1] 嵯峨山 茂樹, 武田 晴登, 亀岡 弘和, 西本 卓也, “音声認識技術を用いた音楽情報処理,” 音響学会誌, Vol. 61, No. 8, pp. 454-460, Aug. 2005.
- [2] Hirokazu Kameoka, Takuya Nishimoto, and Shigeki Sagayama, “Extraction of Multiple Fundamental Frequencies from Polyphonic Music Using Harmonic Clustering,” *Proc. International Congress on Acoustics (ICA) (Kyoto, Japan)*, Apr. 2004
- [3] 亀岡 弘和, 西本 卓也, 嵯峨山 茂樹, “拘束つき混合正規分布の最尤推定とAICによる同時発話複数音声のF0推定,” 情報処理学会研究報告, 2003-SLP-49, pp.229-234, Dec. 2003.
- [4] Hirokazu Kameoka, Takuya Nishimoto, Shigeki Sagayama, “Harmonic-temporal structured clustering via deterministic annealing EM algorithm for audio feature extraction,” in *Proc. International Conference on Music Information Retrieval (ISMIR2005)*, pp. 115-122, 2005.
- [5] Shigeki Sagayama, Hirokazu Kameoka, Takuya Nishimoto, “Specmurt Analysis: A Piano-Roll-Visualization of Polyphonic Music Signal by Deconvolution of Log-Frequency Spectrum,” *Proc. 2004 ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Processing (SAPA2004)*, (2 October 2004 - 3 October 2004, Jeju, Korea), Oct. 2004.
- [6] 齊藤 翔一郎, 亀岡 弘和, 小野 順貴, 嵯峨山 茂樹, “事後確率最大化Specmurt分析による多重ピッチの反復推定アルゴリズム,” 情報処理学会研究報告, 2006-MUS-66, pp.85-92, Aug. 2006.
- [7] 武田 晴登, 西本 卓也, 嵯峨山 茂樹, “確率モデルによる多声音楽演奏のMIDI信号のリズム認識,” 情報処理学会論文誌, Vol.45, No.3, pp.670-679, Mar. 2004.
- [8] 武田 晴登, 西本 卓也, 嵯峨山 茂樹, “HMMを用いたリズムとテンポの反復推定による多声MIDI演奏のリズム認識,” 日本音響学会2006年春季研究発表会 講演論文集, 3-2-3, pp.721-722 (in CD-ROM), Mar. 2006.
- [9] 齊藤 翔一郎, 武田 晴登, 西本 卓也, 嵯峨山 茂樹, “specmurt分析とchroma vectorを用いたHMMによる音楽音響信号の調認識,” 情報処理学会研究報告(MUS), 2005-MUS-61, pp. 85-90, Aug. 2005.

AIアプローチに基づく音楽デザイン転写

奥乃グループ 奥乃 博
(京都大学大学院情報学研究科 教授)

音環境理解 (Computational Auditory Scene Analysis) からのアプローチ

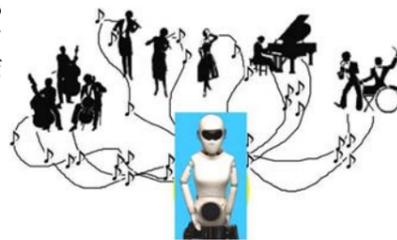
2006年の50年前の夏に米国ダートマスで人工知能 (Artificial Intelligence, AI) という用語が誕生した。当時AIとしてさまざまな研究が取り込まれ、文字認識などのように既に実用化されているものもある一方で、音楽の認識や生成は部分的な成功にとどまっている。本CrestMuseプロジェクトにおいて、我々は音を通じて音響事象の認識理解を行う音環境理解[1]という人工知能の立場から研究を行っている。音響事象に認識の課題は混合音の取り扱いである。音楽音響信号は複数の楽器音や音声から構成されており、音環境理解の主要な研究課題となっている。我々は、入力された混合音中に含まれる楽器音や歌声を分離し、その認識を通して、記号レベル表現を生成し、得られた記号表現を使用して、音楽的な高次予測である音楽デザイン転写へと展開する予定である。

研究の第1ステップは、楽器音、リズム、和声、メロディなどの基本音楽要素を認識し、MusicXMLに基づいた記号表現を生成する。最初は音響信号に基づいたボトムアップ処理を追究する。次に、音楽的な高次構造に基づく予測というトップダウン処理を追究し、最終的に両者を融合する。このようなアプローチを通して、音楽音響信号から記号表現への変換において、全データを参照することなく、逐次的に処理するオンラインアルゴリズムを開発する。このアルゴリズムは本研究課題である elaboration の実時間認識の基礎技術として展開していく。

第2ステップは、得られた記号表現を用いたシステムへの応用である。特に我々は音環境理解の応用としてロボット聴覚の研究も進めており、第1ステップで得られた技術をロボットに応用し、高度なロボットの耳と脳を開発する。

本グループの具体的な研究実施項目は以下の通りである：

- ・音楽基本要素の認識—楽器音、リズム、和声、メロディ
- ・MusicXMLによる音楽情報アノテーションシステムの開発
- ・音楽音響信号の擬音語・擬態語による認識
- ・音楽音響信号から記号変換へのオンラインアルゴリズムの開発と実時間処理の実現
- ・ヒューマノイドロボットの音楽音響信号に合わせた予測に基づく挙動生成
- ・音楽の分かるロボットの脳的设计とプロトタイプの開発



これまでの取り組みと成果

- (1) 楽器音表現法Instrogramの開発[2,10,12,14]：楽器音認識に対して、従来のような単音形成を陽に行わない楽器表現法Introgramを開発した。さらに、Instrogramを類似楽曲検索に応用し、その有効性の検証を進めている。
- (2) ドラム音認識手法 Adamastの開発とそのドラム音イコライザー Drummix への応用[3,8,19]：リズム認識に対して、市販CDレベルの複雑な音響信号中のドラム音認識技術AdaMastを開発した。さらに、AdaMastに基づく新しい音楽再生インタフェースDrummixを開発した。Drummixではユーザは、ドラムスの音量や音色、ドラムパターンなどをリアルタイムにコントロールすることができ、この結果能動的音楽鑑賞が可能となる基礎技術を確立した。
- (3) 歌声区間推定技術とそれに基づく歌手・話者認識技術の開発[9,11,15]：歌声区間推定技術とそれに基づく歌手・話者認識技術にも取り組んだ。
- (4) 標題音楽アノテーション法の開発[16]：標題音楽のシーンと音楽との記述の統合を行った。
- (5) 混合音中の歌声と歌詞との時間的対応付け手法の開発[20]
- (6) 混合音テンプレートによる混合音中の楽器音同定手法の開発[4,5].
- (7) 自動特徴重み付けによる混合音中の特定パートの抽出法の開発[7,17]
- (8) 環境音の擬音語に音高・持続時間を追加したXMLにより表現法の確立[18].
- (9) 音楽音響信号処理に基づいた楽曲推薦システムの構想と開発[6,13]

本シンポジウムでのデモンストレーション

上記の成果について、本シンポジウムでは次のようなデモンストレーションを行う予定である。

- ・「Drumix:ポピュラー音楽CDのドラムパート編集機能付きオーディオプレイヤー」体験デモ。吉井 和佳, 後藤 真孝 (産総研), 奥乃 博
- ・「Instrogramによる類似楽曲検索」実演デモ。北原 鉄朗, 後藤 真孝 (産総研) ・奥乃 博
- ・「混合音中の歌声と歌詞との時間的対応付け」実演デモ。藤原 弘将, 後藤 真孝 (産総研) 奥乃 博
- ・「混合音中のパートの音量を操作可能なオーディオプレイヤー」実演デモ。糸山 克寿, 吉井 和佳, 奥乃 博

発表文献一覧

- [1] David Rosenthal, and Hiroshi G. Okuno (eds): Computational Auditory Scene Analysis, Lawrence Erlbaum Associates, Mahwah, NJ, U.S.A., 1998.
- [2] Tetsuro Kitahara, Masataka Goto, Hiroshi G. Okuno: Pitch-dependent identification of musical instrument sounds, Applied Intelligence, Vol.23, No.3, pp.267-275 (Dec. 2005) Springer-Verlag
- [3] Kazuyoshi Yoshii, Masataka Goto, Hiroshi G. Okuno: Drum Sound Recognition for Polyphonic Audio Signals by Adaptation and Matching of Spectral Templates after Suppressing Harmonic Structure, IEEE Transactions on Speech and Audio Processing, in print, (Jan. 2007)
- [4] 北原鉄朗, 後藤真孝, 駒谷和範, 尾形哲也, 奥乃博: 多重奏を対象とした音源同定: 混合音テンプレートを用いた音の重なり頑健な特徴量への重みづけおよび音楽的文脈の利用, 電子情報通信学会論文誌, 採録決定, July 2006, 電子情報通信学会.
- [5] Tetsuro Kitahara, Masataka Goto, Kazunori Komatani, Tetsuya Ogata, Hiroshi G. Okuno: Instrument Identification in Polyphonic Music: Feature Weighting to Minimize Influence of Sound Overlaps, EURASIP Journal on Applied Signal Processing, Special issue on Music Information Retrieval Based on Signal Processing, accepted, Aug. 2006.
- [6] Kazuyoshi Yoshii, Masataka Goto, Kazunori Komatani, Tetsuya Ogata, Hiroshi G. Okuno: Hybrid Collaborative and Content-based Music Recommendation Using Probabilistic Model with Latent User Preferences, Proceedings of 7th International Conference on Musical Information Retrieval (ISMIR-2006), accepted, CA, Sep. 2006.
- [7] Katsutoshi Itoyama, Tetsuro Kitahara, Kazunori Komatani, Tetsuya Ogata, Hiroshi G. Okuno: Automatic Feature Weighting in Automatic Transcription of Specified Part in Polyphonic Music, Proceedings of 7th International Conference on Musical Information Retrieval (ISMIR-2006), accepted for presentation, Vancouver, CA, Sep. 2006.
- [8] Kazuyoshi Yoshii, Masataka Goto, Kazunori Komatani, Tetsuya Ogata, Hiroshi G. Okuno: An Error Correction Framework based on Drum Pattern Periodicity for Improving Drum Sound Detection, Proceedings of 2006 International Conference on Acoustics, Speech and Signal Processing (ICASSP'2006), Vol.V, pp.237-240, 2006
- [9] Hiromasa Fujihara, Tetsuro Kitahara, Masataka Goto, Kazunori Komatani, Tetsuya Ogata, Hiroshi G. Okuno: F0 Estimation Method for Singing Voice in Polyphonic Audio Signal based on Statistical Vocal Model and Viterbi Search, Proceedings of ICASSP'2006, Vol.V, pp.253-256, Toulouse, May 2006.
- [10] Tetsuro Kitahara, Masataka Goto, Kazunori Komatani, Tetsuya Ogata, Hiroshi G. Okuno: Instrogram: A New Musical Instrument Recognition Technique Without Using Onset Detection Nor F0 Estimation, Proceedings of ICASSP'2006, Vol.V, pp.229-232, Toulouse, May 2006.
- [11] 藤原 弘将, 北原鉄朗, 後藤 真孝, 駒谷 和範, 尾形 哲也, 奥乃 博: 伴奏音抑制と高信頼度フレーム選択に基づく楽曲の歌手名同定手法, 情報処理学会論文誌, Vol.47, No.6 (June 2006) pp.1831-1843, 情報処理学会.
- [12] 北原 鉄朗, 後藤 真孝, 駒谷 和範, 尾形 哲也, 奥乃 博: Instrogram: 発音時刻検出とF0推定の不要な楽器音認識手法, 音楽情報処理研究会, 2006-MUS-66 情報処理学会, Aug. 2006.
- [13] 吉井 和佳, 後藤 真孝, 駒谷 和範, 尾形 哲也, 奥乃 博: ユーザの評価と音響的特徴との確率的統合に基づくハイブリッド型楽曲推薦システム, 音楽情報処理研究会, 2006-MUS-66, ベストプレゼンテーション賞.
- [14] 北原 鉄朗, 後藤 真孝, 駒谷 和範, 尾形 哲也, 奥乃 博: Introgram: 楽器存在確率に基づく音楽視覚表現法, 音響学会春季講演会, 2-2-13, 715-716, Mar. 2006.
- [15] 藤原 弘将, 北原 鉄朗, 後藤 真孝, 駒谷 和範, 尾形 哲也, 奥乃 博: 調波構造抽出と高信頼度フレーム選択を用いた雑音下での話者識別, 音響学会春季講演会, 1-11-17, 55-56, Mar. 2006.
- [16] 西山 正紘, 北原 鉄朗, 駒谷 和範, 尾形 哲也, 奥乃 博: 標題音楽アノテーションのための階層的物語タグの設計, 情報処理学会第68回全国大会, 3L-6, Mar. 2006.
- [17] 糸山 克寿, 北原 鉄朗, 駒谷 和範, 尾形 哲也, 奥乃 博: 多重奏中特定パートの自動採譜における複数特徴量の自動重み付け, 情報処理学会第68回全国大会, 2L-4, Mar. 2006..
- [18] 田口 明裕, 北原 鉄朗, 石原 一志, 駒谷 和範, 尾形 哲也, 奥乃 博: 擬音語表現を利用した環境音のためのXMLタグの設計と自動付与, 情報処理学会第68回全国大会, 3L-7 (Mar. 2006) 学生奨励賞.
- [19] 吉井 和佳, 後藤 真孝, 駒谷 和範, 尾形 哲也, 奥乃 博: Drumix: ドラムパートのリアルタイム編集機能付きオーディオプレイヤー, インタラクティブ2006, 207-208, 情報処理学会. インタラクティブ発表賞.
- [20] 藤原 弘将, 後藤 真孝, 緒方 淳, 駒谷 和範, 尾形 哲也, 奥乃 博: 音楽音響信号と歌詞の時間的対応付け手法: 歌声の分離と母音のViterbiアラインメント, 音楽情報処理研究会, 2006-MUS-66, 情報処理学会, Aug. 2006.



Interactive Session

インタラクティブ セッション

体験デモ

MusicRainbow: A New User Interface to Discover Artists

Elias Pampalk and Masataka Goto

(Goto's Group: National Institute of Advanced Industrial Science and Technology (AIST))

Abstract MusicRainbow is a simple user interface for discovering artists. Similar artists are mapped close to each other on a circular rainbow using a traveling salesman algorithm. The similarity of artists is computed using a new audio-based approach. The colors of the rainbow encode different types of music. Furthermore, a new approach is used to combine this audio-based information with information from the web. In particular, the contents are summarized and labeled with words extracted from web pages related to the artists. We use different vocabularies for different hierarchical levels and heuristics to select the most descriptive labels.

Keywords: user interface, artist similarity, music collection visualization and summarization.

1. Introduction

MusicRainbow is a new music interface to discover artists. Artists are arranged on a circular rainbow (with no beginning or end, see Figure 1). Similar artists are located close to each other on the rainbow. The similarity is computed by analyzing the songs (audio signals) of each artist. The rainbow is labeled automatically with words occurring on web pages related to the artists. The web pages are found using Google and filtered using specific vocabularies.

Our primary objective is to support users in discovering new artists. First of all, given a large music collection with many different styles of music, the user is presented an overview of the contents. We do not require the user to start the search with a specific artist in mind. The secondary objective is to keep the interface as simple as possible. Usage of the interface should be intuitive. The interface should be implementable on mobile devices with limited screen sizes including phones and music players.

MusicRainbow was originally presented in [6]. An approach with similar objectives was presented in [5]. An approach mapping music to a wheel was presented in [8]. Generally, related user interfaces include, for example, [2, 7, 9].

2. User Interface

Figure 1 shows the MusicRainbow interface. Inside the circular rainbow are high-level genre terms which describe the different sections (e.g. Rap, Jazz, Soul). More specific terms (e.g. Female, Guitar, Beats) are located on the outside. Each of the 8 concentric color rings of the rainbow corresponds to one high-level term. In the screenshot, for example, the yellow-green color ring corresponds to Rap and red corresponds to Rock. The colors range from purple (inside) to red (outside). In segments of the rainbow where the re-

spective terms are frequently used to describe the artists, the colors are brighter.

In the center of the right side is the name of the currently selected artist. In the screenshot (Figure 1, right side), Outkast is selected. Artists in the neighborhood include, for example, Eminem, Jay-Z, Busta Rhymes, and Warren G. At the bottom right, the selected artist is described with some words. For Outkast the following (automatically generated) description is displayed: "rap, singing, southern, duo, beats". (Note that Outkast is a duo from Atlanta.)

Furthermore, below the description, 8 color bars (corresponding to the colors of the rainbow) are displayed. These color bars help the user understand which colors correspond to which type of music. For example, for Outkast the yellow-green color has the largest bar. This indicates that other segments of the rainbow where this color is highlighted contain similar artists.

All inputs from the user can be given using, for example, a Griffin Powermate knob (Figure 1, upper left). There are basically 3 input actions. (1) By turning the knob quickly the user can jump to different regions on the rainbow. (2) By turning the knob slowly the user stays in the same region and can explore similar artists. (3) By pushing the knob the user can select an artist to listen to. The currently playing song is displayed in the upper right. By turning the knob during playback, the user can read the summaries of other artists. From each song, only a segment of about 20 seconds length is played. To identify the segments that summarize the song as well as possible, we use a chorus-section detection method (RefraiD [1]).

3. Technique

The techniques are described in more detail in [6]. We use a computational model of music similarity to compute the similarities between pieces of music [4]. Based on these, we compute the artist similarity. We use

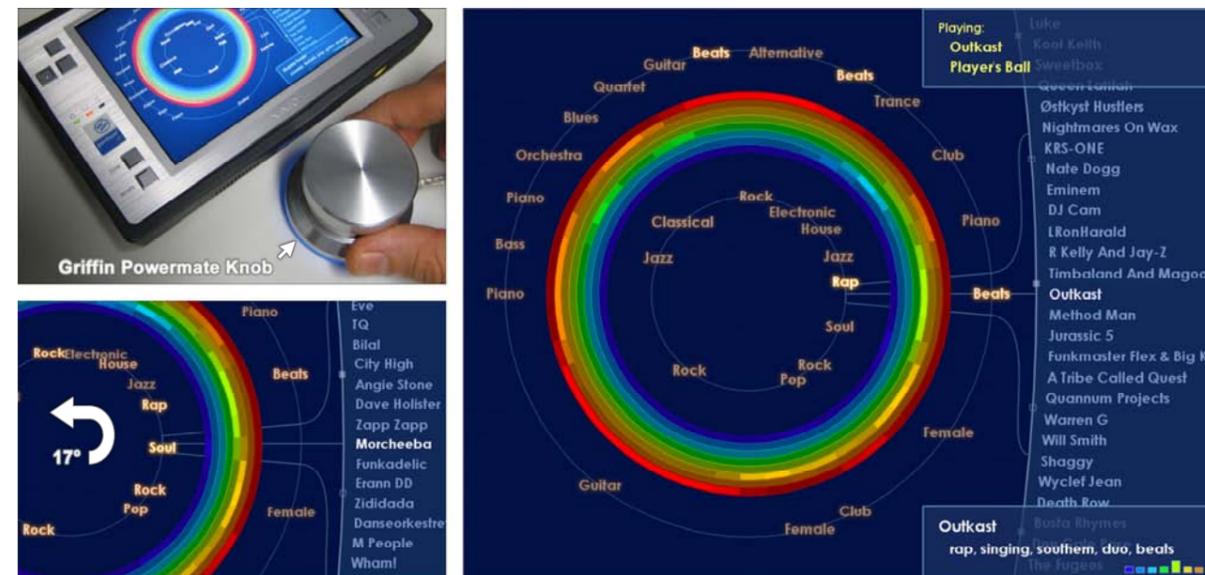


Figure 1: The MusicRainbow interface. The input device (Griffin Powermate knob) is shown on the upper left. On the right side shows a screenshot. The picture on the lower left shows how colors and labels change when the rainbow (shown on the right) is rotated counter-clockwise by 17 degrees.

a traveling salesman algorithm to find the shortest path which connects the artists and project them onto a circle (i.e. the circular rainbow) [8].

To extract words, we query Google with the artists' names and retrieve related pages [10]. Using three different vocabularies, we parse these pages. The first vocabulary is used to find words for inside the label, and the distribution of words of this vocabulary also defines the colors of the rainbow. The second vocabulary is used to find words which are placed along the rainbow on the outside. The third vocabulary is used to summarize each artist. Using an adaptation of the approach presented in [3], we select the labels for summarization of the contents.

4. Conclusions

MusicRainbow is a new interface to explore music collections at the artist level. The colors of the rainbow encode different styles of music. Similar artists are located close to each other on the rainbow. We use a new approach to compute artist similarities and a new approach to combine audio-based similarity and information extracted from web pages (using different vocabularies for different hierarchical levels). We believe MusicRainbow can help users make interesting discoveries amidst the rapidly growing number of artists on the market.

References

[1] Goto, M.: A Chorus-Section Detection Method for Musical Audio Signals and Its Application to a Music

Listening Station, IEEE Transactions on Audio, Speech and Language Processing, Vol.14, No.5, pp.1783-1794 (2006).

- [2] Goto, M. and Goto, T.: Musicream: New Music Playback Interface for Streaming, Sticking, Sorting, and Recalling Musical Pieces, In Proc. of ISMIR (2005).
- [3] Lagus, K. and Kaski, S.: A Keyword Selection Method for Characterizing Text Document Maps, In Proc. of ICANN (1999).
- [4] Pampalk, E.: Computational Models of Music Similarity and their Application in Music Information Retrieval, Doctoral Thesis, Vienna University of Technology, Austria (2006).
- [5] Pampalk, E., Flexer, A., and Widmer G.: Hierarchical Organization and Description of Music Collections at the Artist Level, In Proc. ECDL (2005).
- [6] Pampalk, E. and Goto, M.: MusicRainbow: A New User Interface to Discover Artists Using Audio-based Similarity and Web-based Labeling, In Proc. ISMIR (2006).
- [7] Pampalk, E., Rauber, A., and Merkl D.: Content-based Organization and Visualization of Music Archives, In Proc. ACM MM (2002).
- [8] Pohle, T., Pampalk E., and Widmer, G.: Generating Similarity-Based Playlists Using Traveling Salesman Algorithms, In Proc. of DAFX (2005).
- [9] van Gulik, R., Vignoli, F., and van de Wetering, H.: Mapping Music in the Palm of Your Hand, In Proc. ISMIR (2004).
- [10] Whitman, B. and Lawrence, S.: Interfering Descriptions and Similarity for Music from Community Metadata, In Proc. ICMC (2002).

実演デモ

Musicream: 楽曲を流してくっつけて並べて 思い出すことのできる音楽再生インターフェース

後藤 真孝 後藤 孝行
(後藤グループ：産業技術総合研究所)

概要 本研究では、従来の曲名等に基づく単なる楽曲検索を越えて、ユーザが漠然と抱く「何か聴きたい」という欲求を満たせる音楽検索・再生インターフェース Musicream を実現した。ユーザがある曲を選択して楽曲集合の中をドラッグすると、自動計算した楽曲類似度に基づいて似た曲がくっついて集まる機能や、それらの楽曲群をザッピング再生してプレイリストを作成する機能、操作履歴を自在に参照する機能等により、音楽との多様な出会い方を可能にした。

キーワード：音楽再生インターフェース、音楽情報検索、音楽類似検索、プレイリスト、流れ系インターフェース

1. はじめに

現在の音楽再生インターフェースでは、曲名やアーティスト名に基づいて「この曲を聴きたい」というユーザの欲求を満たせるが、「何か聴きたい」「思い通りに聴きたい」という欲求を満たす機能は不十分である。しかし、数十万曲の楽曲を無制限に聴くことができる定額制聴き放題の音楽配信サービスが普及しつつある今日、次世代の音楽再生環境としてこうした欲求は高まっていくと考えられる。例えば、何気なく手に取った1曲を気に入ったときに、それに似た雰囲気の曲を次々に聴いていきたい。机の上でCDを並び替えたりするのと同じぐらい高い自由度で、楽曲の再生順を検討したい。あの日聴いていた、あの時代に流行っていた曲を聴いてみたい。従来、こうした欲求を満たす機能は、不十分とはいえ個々の要素技術として検討されることがあった。しかし、それらすべてをシームレスで手軽に使えるような、総合的な音楽再生環境を実現するインターフェースは存在していなかった。

そこで我々は、こうした自由度の高い音楽の聴き方のすべてを可能にする音楽再生インターフェース「Musicream」(music + stream) を実現した[1, 2]。Musicreamでは以下の四つの機能を備えた、これまでに体験したことのない音楽再生環境を提供する。

- (1) 流しそめんや回転寿司のように、画面上を次々と楽曲群が流れる「流し楽曲」機能
- (2) 磁石で磁石をくっつけるように、手に持った楽曲で似た楽曲をくっつける「類似くつき」機能
- (3) 机の上にCDを並べるように、楽曲群を画面上に並べて曲順を検討する「メタプレイリスト」機能
- (4) タイムマシンに乗って音楽を聴いている過去の自分に戻るように、時間軸を自在に行き来する「タイムマシン」機能

2. Musicreamの機能

Musicreamが備える四つの機能を順に説明する。

2.1 「流し楽曲」機能

ユーザが利用できる音楽コレクションを、図1に示すように順々に提示する機能である。流れている任意の曲(Disc)を選択して聴けるので、「この曲を聴きたい」という明確な要求がないときに、多様な楽曲と何気なく出会う手段として有効である。Discとそれが流れ出る蛇口を、曲の雰囲気(色)を反映した色で着色することで、あるDiscを手にとって試聴して気に入ったら、その色を参考に他の雰囲気(色)の似た曲を選ぶことができる。

2.2 「類似くつき」機能

流れから取り出したDiscを持って、流れている他のDiscに接触させることにより、雰囲気の似たDiscだけを選択的に次々とくっつけて、図2のようにプレイリスト(再生曲順を示すDiscの連なり)を作ることができる機能である。一種の類似楽曲検索だが、従来のように類似曲をリストアップする検索ではなく、ユーザの意志で類似した曲を拾い集めることができる点が大きく異なる。ここで、楽曲の雰囲気が似ているほどDisc同士がくつきやすいように設計することで、手に取ったDiscを流れの中でどう漂わせるか(どう接触させるか)によって、追加選択する曲の類似度の範囲を容易に調整できる。

2.3 「メタプレイリスト」機能

図3のように2次元平面上に置いた複数のプレイリストを並べて、再生順序を指定できる機能である。これは、画面全体をさらにメタなプレイリストと捉え直し、全プレイリストを画面の上から下へ順番に再生することで実現する。図3の再生バーを画面上部に持ち上げると落下し、接触したプレイリストから再生される。このように、楽曲群を自在に入れ替えてメタなプレイリストを作るという、かつてなかった自由度の高いプレイリストの編集が可能になる。



図1 「流し楽曲」機能: 流しそめんや回転寿司のように、画面上を次々と楽曲群(円盤)が上から下へ流れ、ユーザは楽曲を何気なく手に取ることで、未知の楽曲に能動的に出会うことができる。

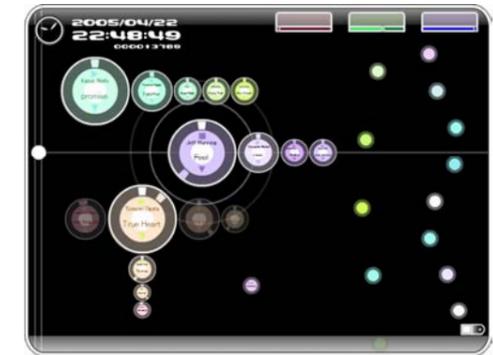


図3 「メタプレイリスト」機能: 全プレイリストを画面の上から下へ順番に再生することで、画面全体をさらにメタなプレイリストと捉え直し、プレイリスト(楽曲群)を画面上で自由に並べて曲順を検討できる。

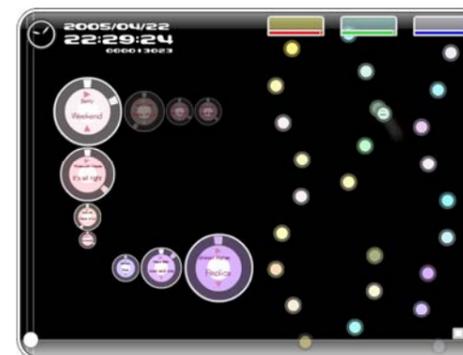


図2 「類似くつき」機能: 流れから取り出した楽曲を持って、流れている他の楽曲に接触させることにより、雰囲気の似た楽曲だけを選択的に次々とくっつけることができる(制御可能な類似楽曲検索に相当する)。くっつけて作った楽曲群(円盤の連なり)は、再生順を示すプレイリストになる。



図4 「タイムマシン」機能: Musicream上のすべての操作は記録され、タイムマシンに乗って音楽を聴いている過去の自分に戻るように、時間軸を自在に行き来できる。保存(save), 呼出し(load)が一切不要となる。

2.4. 「タイムマシン」機能

ユーザによる操作と画面の変化を記録し続けることで、図4のような操作画面で過去の記録を閲覧しながら、任意の時点に戻れる機能である。あたかもタイムマシンに乗って過去に戻ったかのように、その時点での画面を再現し、操作を続けることが可能になる。操作を間違っても、後悔することなくいつでも復帰できるので、気軽に試行錯誤できる。

3. まとめ

本稿では、上記の四つの機能をシームレスに操作できる総合的な音楽再生インターフェース Musicreamを紹介した。Musicreamでは、ユーザの「何か聴きたい」という欲求を最初の二つの機能によって満たし、「思い通りに聴きたい」という欲求を最後の二つの機能によって満たした。RWC研究用音楽データベース[3]の全315曲を利用して Musicreamを運用した結果、四つの機能を組み合わせながら、これまでに体験したことのない自由度の高い聴き方ができることを確認した。

参考文献

- [1] Masataka Goto and Takayuki Goto: Musicream: New Music Playback Interface for Streaming, Sticking, Sorting, and Recalling Musical Pieces, Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005), pp.404-411, September 2005.
- [2] 後藤 孝行, 後藤 真孝: "Musicream: 楽曲を流してくっつけて並べることのできる新たな音楽再生インターフェース", 日本ソフトウェア科学会 第12回インタラクティブシステムとソフトウェアに関するワークショップ (WISS 2004) 論文集, pp.53-58, December 2004.
- [3] 後藤 真孝, 橋口 博樹, 西村 拓一, 岡 隆一: "RWC研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース", 情報処理学会論文誌, Vol.45, No.3, pp.728-738, March 2004.

体験デモ

歌真似と声真似, STRAIGHTのモーフィングは、 どこまで人間に迫るか？

河原 英紀 森勢 将雅 畑 宏明 田原 佳代子 生駒 太一
豊田 健一[†] 片寄 晴弘[†]

(河原グループ：和歌山大学システム工学部聴覚メディア研究室)
[†] 関西学院大学工学部 (元)

概要 STRAIGHTに基づく高品質なモーフィングは事例に基づく音楽デザインの強力な研究並びに実装の手段を与える。ここでは、そのために必要となるSTRAIGHTに基づくモーフィングの拡張である部分モーフィングと再帰的モーフィングを紹介し、演奏デザインの転写への応用について説明する。また、実際の歌唱を例として、モーフィングにより声質の転写（声真似）と歌い回しの転写（歌真似）を行い主観評価した結果を紹介する。シンポジウムの会場では、STRAIGHTによる歌真似・声真似と、実際のプロ歌手による歌真似・声真似の比較のデモを体験して頂きたい。

キーワード：STRAIGHT, モーフィング, 声質, 歌い直し。

1. はじめに

モーフィングは、物理パラメータと心理的属性との対応関係についての明示的な知識に依存せず、2つの試料の中間状態の刺激を生成することを可能にする。また、異質の事例を組み合わせることにより、現実には存在し得ない効果を生み出すことを可能にする。これらは、モーフィングをコンテンツ制作におけるデザイン操作のためのユニークなツールとする特長である。STRAIGHT [1]は、聴覚メディアの領域において、高品質なモーフィングの利用を可能とした[2,3]。ここでは、STRAIGHTによるモーフィングの仕組みを紹介し、モーフィングを演奏デザインの再利用に応用する場合に必要な拡張について紹介する。また、それらに基づいて行った声質の転写（声真似）と歌い回しの転写（歌真似）について説明するとともにインタラクティブセッションで紹介するデモについて説明する。

2. STRAIGHTに基づくモーフィング

STRAIGHTは、音声を(1)基本周波数、(2)非周期成分から構成される音源情報と、滑らかに変化する(3)スペクトル包絡情報とに分解する。これらの情報は実数のパラメータとして表されており、複素数の情報である位相のような干渉を生じない。これら三種類の加工性の良いパラメータとして音声が表現されていることが、高品質のモーフィングを可能にしている。

2.1 モーフィング手順の概要

STRAIGHTに基づくモーフィングにおいては、まず、事例間の対応関係を指定する。時間周波数表現上での対応点を指定することにより、異なる事例を表すパラメータを共通の時間周波数座標の上で表現する。次いで、この共通の座標の各点において事例間のパラメータの補間・補外を行い、目標

とする表現のパラメータの値を求める。こうして求められた値を、目標とする時間周波数軸上に戻し目標とする音声と合成する[2]。

2.2 対応点を用いたインタフェース

対応点は、共通の時間周波数座標への写像関数を指定するためのインタフェースの一つの手段として用いられている。STRAIGHTによるモーフィングは、まず、音声の非言語およびパラ言語情報の研究のためのツールとして開発された。そのため、実験結果に大きく影響する共通の時間周波数座標への変換は、研究者の実験意図に応じて、手作業により注意深く設定されるべきであることを前提としていた。

しかし、コンテンツ制作に用いる場合には、手作業の低い生産性が問題となる。ここで紹介する歌唱モーフィングの研究では、時間方向の対応点の生成にスケールフィルタを用い、半自動の操作インタフェースを構築することで生産性の向上を図っている[4]。また、ラベル情報と個人毎の母音テンプレートを用いて、周波数方向の対応点を自動的に生成する方法についての検討も並行して進められている。

対応点をシステムとユーザのインタフェースとして用いる場合、ユーザが対応点の設定そのものに注意を奪われてしまう問題が生ずる。その結果、パラメータの補間を安定に行うことのできる共通の時間周波数座標への変換という本質が見失われ、品質の大きな低下を招く事態が多数生じた。この問題を避けるための、より適切なインタフェースにおけるメタファーの検討が必要である。

2.3 モーフィングの幾何学的意味

最初に挙げた三種類のSTRAIGHTのパラメータに加え、時間周波数表現を共通の座標によって表現することを可能にするための関数を指定するため

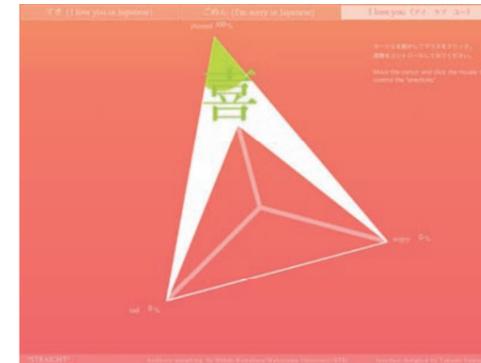


図1 再帰的モーフィングの応用

に、時刻の変換と、それぞれの時刻における周波数の変換が必要となる。これらは、モーフィングの最初の実装では、一つのスカラー量のパラメータにより制御されていた。しかし、これらの制御パラメータが共通である必要はない。結局、現在のSTRAIGHTによるモーフィングには、ベクトルの要素として表される5つの独立に操作可能なパラメータがあることになる。

この一般化の下では、一つのスカラー量による2つの事例のモーフィングは、一つの事例を超立方体の頂点(0,0,0,0,0)に配置し、もう一つの事例を頂点(1,1,1,1,1)に配置し、両者を等間隔にマークされた直線で結ぶことに相当する。デザイン転写の観点からは、与えられた事例間を補間・補外できるだけでは不十分である。事例を参照点として、意図した方向に演奏を変化させる（しかも、その変換によって品質を大きく劣化させない）ことが必要となる。

2.4 再帰的モーフィング

事例数の増加に伴い、モーフィングを二項関係として扱う場合の組み合わせは、事例の階乗で増加する。この組合せの爆発を避けるには、共通の時間周波数座標に変換するための手掛かりを、事例自体の属性として定義する必要がある。ここでは、そのような属性をSTRAIGHTによる分析結果に付与し構造体として表現したものをモーフィングオブジェクトと定義し、モーフィングを、複数のモーフィングオブジェクトからモーフィングオブジェクトを生成する演算として定義する。このようにモーフィングの概念を拡張することにより、2個以上の事例間のモーフィングを容易に実装することができる。図1は、三種類の感情を任意に混合したモーフィング音声と合成するためのインタフェースへの応用例である[5]。

3 歌唱のモーフィングと真似

STRAIGHTが用いている音源情報とスペクトル包絡という情報表現は、声質のような比較的低いレベルの聴覚的印象とは良い対応を示す。歌唱へのモーフィングの応用の第一段階として、スペク

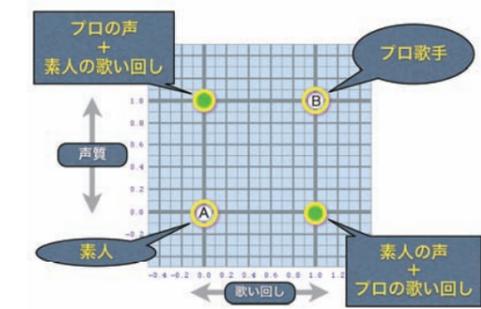


図2 歌唱モーフィングのインタフェース

トルの周波数軸方向の操作と各点での値の操作が声質の操作であると想定し、時間軸方向の操作と基本周波数およびエネルギーの操作が歌い回しの操作であると想定して、図2に示す操作空間の知覚的印象を調べた[4]。その結果、「歌手らしさ」の印象には声質が支配的であることが示された。この結果が一般に成立するものであるか、プロの歌手が他の歌手を演じる場合に、モーフィングと同様な戦略が用いられるのか、新たな素材を用いて検討を進めている。

4 おわりに

STRAIGHTに基づくモーフィングを用いた声真似と歌真似について紹介した。インタラクティブセッションにおいて、実際にそれらがどのような印象を生み出すか、また、実際にプロ歌手が「声真似」「歌真似」を試みた場合の表現とどう違うのかを、ぜひ体験して頂きたい。

参考文献

- [1] Kawahara, H., Masuda-Katsuse, I., and Cheveigné, de A.: Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction, *Speech Communication*, 27(3-4), pp.187-207 (1999).
- [2] Kawahara, H., and Matsui, H.: Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation, *In Proc. ICASSP 2003, Vol.1*, pp.256-259 (2003).
- [3] Yonezawa, T., Suzuki, N., Mase, K., and Kogure, K.: Gradually changing expression of singing voice based on Morphing, *In Proc. Interspeech2005*, pp.541-544 (2005).
- [4] 豊田健一, 片寄晴弘, 河原英紀: STRAIGHTによる歌声モーフィングの初期的検討, *情報処理学会研究報告*, Vol.2006, No.19, 2006-MUS-64, (2006).
- [5] 日本科学未来館特別企画展「恋愛物語展-どうして一人ではいられないの?」4/23-8/15(2005). (音声モーフィング: 河原英紀, インタフェースデザイン: 山口崇司)

STRAIGHTの原理とrealtime STRAIGHT,そして, 応用へ

河原 英紀

(河原グループ : 和歌山大学システム工学部聴覚メディア研究室)

概要 STRAIGHTは、聴覚における情報表現と機能的に等価な工学的表現に基づく音声・音響処理を実現することを目標として開発されたシステムである。ここでは、まず、この目標を実現するために導入された主要なアイデアを紹介する。次いで、STRAIGHTが実時間処理に適した構造であることを説明し、それを実証するシステムとしてrealtime STRAIGHTを紹介する。最後に、STRAIGHTの応用と解決すべき技術課題について説明する。

キーワード : 音声, 周期信号, 群遅延, vocoder.

1. はじめに

STRAIGHTは、聴覚における情報表現と機能的に等価な工学的表現に基づく音声・音響処理を実現することを目標として開発されたシステムである。STRAIGHTの中核部分は、生態学的に重要な意味を持ち聴覚においても特別扱いされている、周期性のある音の処理のために開発された。周期性のある音は、コミュニケーションにおける知覚と行動の循環を形成する言葉の鎖の重要な一つの輪である。この輪に介入することによる音声コミュニケーション変容の研究は、当初からSTRAIGHTの主要な応用として想定されていた。STRAIGHTで用いられているアルゴリズムが、realtime処理として無理無く実装できるように構成されているのは、そのためである。

2. STRAIGHTの原理

周期的な励振の利用は、様々な妨害音や伝播特性を有する環境中において、発音体の形状情報を確実に伝えるための有力な戦略である。生物のコミュニケーションにおいては、この戦略が広く用いられており、人間の聴覚においても、周期信号は特別扱いされている。聴覚系の解くべき課題は、このような形で提供される信号から、発音体の形状の動的变化や、励振の頻度に乗せられたメッセージを取り出すことであると考えられる。STRAIGHTでは、この課題に対して、拡張されたピッチ同期分析と相補的時間窓というアイデアを用いて、工学的な解を与えている。

2.1 相補的時間窓

ピッチ同期分析は、周期信号から安定に信号の特徴を抽出するためによく用いられる方法である。しかし、現実に観測される音声信号では、基本周期は必ずしも一定ではなく、繰り返し波形も完全には同一ではない。そのため、基本周期に同期した波形の切り出しを行っても、不完全な周期性に起因する不連続性が、得られるパラメタの大きな変動要因となる。

この問題は、まず、両端での不連続性が小さな

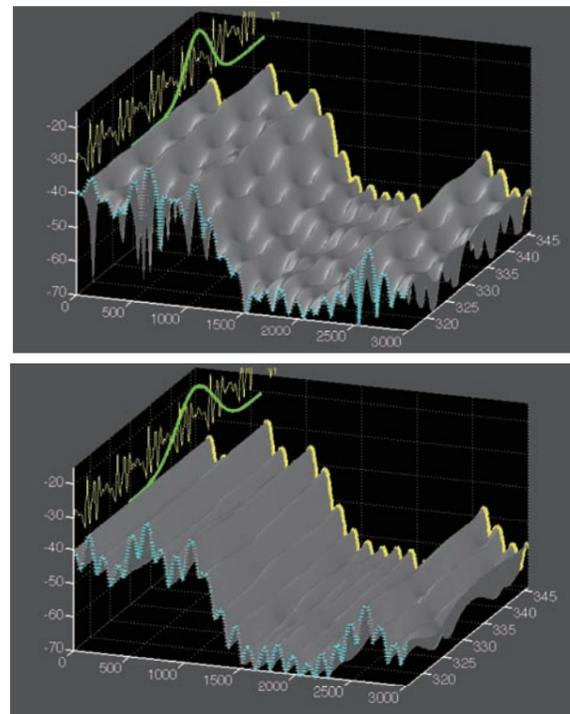


図1 等方的な窓によるスペクトル(上)と、相補的時間窓によるスペクトルを合成して時間変動を最小化したスペクトル(下)

窓をピッチ同期化することにより、各調波の位置においては、解決される。ピッチ同期化は、基本周期に同期した矩形窓やBartlett窓を、元の窓に畳込むことにより実現できる。

しかし、この対策を行っても調波間の干渉に起因するスペクトルの周期的な変動の問題が残る。この問題は、隣接する調波の位相差が基本周期毎に 2π 変化することによる。この変動は、単一の窓を用いる場合には避けることができない。

相補的時間窓は、元の窓に基本周期数の半分の周波数の正弦波を乗じて作られる窓である。この窓を用いると、調波間の干渉で元の窓による振幅スペクトルが穴を持つ位置が極大となるスペクトルが得られる。STRAIGHTでは、元の窓と相補的時間窓から求められる二種類のスペクトルの重み付き自乗和の時間変動が最小となるように、重みを

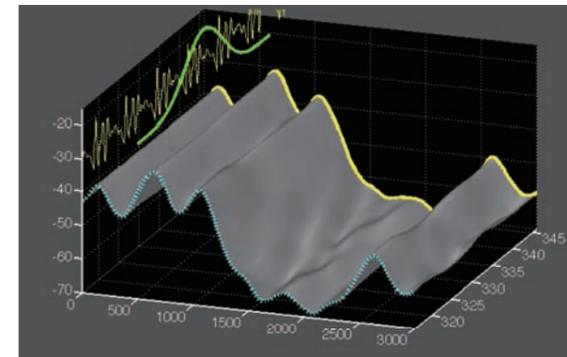


図2 調波空間での逆フィルタ処理により求められたSTRAIGHTスペクトログラム

決定する。その結果、分析窓の位置と波形の相対位相に影響されないピッチ同期分析が可能となる。

2.2 調波空間での逆フィルタ処理

このようにして求められたスペクトルには、周波数方向に基本周波数毎の周期的構造が残る。この構造が線スペクトルであれば、基本周波数の二倍を底辺とする二等辺三角形の関数を用いた周波数領域での平滑化は、区分的一次関数による補間と同じ結果を与える。実際には窓関数による周波数方向の広がりや補償するため、調波位置での値を保存する平滑化関数を用いる。この平滑化関数は、調波位置でのスペクトルの値が張る空間における窓関数のスペクトルの逆元として求められる。

2.3 基本周波数の抽出

これらの処理では基本周波数が重要な役割を果たす。基本周波数の抽出には、既存の方法を用いることもできる。しかし、自然音声と同等の高い品質の分析合成のためには、抽出誤りを既存の方法より一桁低いものとする必要がある。そのため、対数周波数軸上に等間隔に配置されたフィルタ群出力の瞬時周波数を用いる方法[1][3]、帯域毎の正規化自己相関関数と併用する方法[4]等、STRAIGHTに特化した処理が開発されている。

2.4 非周期性

信号の周期性は、それが生態学的に重要な信号であることを示すマーカーの役割を果たす。同時に、周期性からの逸脱は、豊かな情報のチャンネルを提供する。この逸脱を、基本周波数に適応した楕円フィルタを用いて、調波位置以外に表れるエネルギーの割合を表す非周期性指標として求めている。なお、基本周波数の軌跡の中に既に表現されている周波数変調成分が二重に計上されることを避けるため、瞬時周波数を利用した時間軸の正規化を利用している[3]。

2.5 群遅延操作

STRAIGHTによる分析合成は、波形を保存しない。スペクトルの概形と、群遅延のパラメタとして表される微細な時間構造の統計的特徴が保存さ

れる。合成時には、群遅延の操作が、標準化周期よりも高い時間分解能での基本周波数の制御と、最小位相応答の生成、分析合成音声特有のバズ音による品質劣化の軽減のために用いられている。

3. realtime STRAIGHT

STRAIGHTは、研究用のツールとしての柔軟性を重視して、Matlabを用いる一括処理のシステムとしての実装が先行している。しかし、STRAIGHTを構成する処理は、オフラインでの高品質変換を目的に開発された高精度の基本周波数抽出法以外の要素では、最適化のための繰り返し演算を含まず履歴に依存しないアルゴリズムに基づいている。そのため、マシンの速度が十分であれば、realtime化することに原理的な困難は無い[1,3]。Realtime STRAIGHTでは、この履歴依存性が無いという特徴を生かす逐次処理のアーキテクチャを用いて、Matlab版とは異なる実装が行われている[5]。

4. おわりに

VOCODER型の処理は、概念が明確であり高い加工の自由度を有する。この処理の唯一の欠点であった品質の問題を解決したSTRAIGHTは、様々な領域に応用され始めている。しかし、ヘッドフォンやニアフィールドモニタを用いて注意深く聴取すると、STRAIGHT特有の品質劣化に気付く。プロによるコンテンツの制作の厳しい要求に応えることのできるよう、頑健性と柔軟性を併せ持つシステムの実現に向けて、検討を進めている。

参考文献

- [1] Kawahara, H., Masuda-Katsuse, I., and Cheveigné, A.: Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction, *Speech Communication.*, 27(3-4), pp.187-207 (1999).
- [2] Kawahara, H., Banno, H., Morise, M., and Hirachi, Y.: A cappella synthesis demonstrations using RWC music database, In *Proc. NIME04*, pp.130-131 (2004).
- [3] Kawahara, H., Katayose, H., Cheveigné, A., and Patterson, R. D.: Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of F0 and periodicity, *EUROSPEECH'99*, Vol.6, pp.2781-2784 (1999).
- [4] Kawahara, H., Cheveigné, A., Banno, H., Takahashi, T., and Irino, T.: Nearly defect-free F0 trajectory extraction for expressive speech modifications based on STRAIGHT, *Interspeech'2005*, pp.537-540 (2005).
- [5] Banno, H., Hata, H., Morise, M., Takahashi, T., Irino, T., and Kawahara, H.: Implementation of realtime STRAIGHT speech manipulation system, [submitted].

実演デモ

音楽音響信号をMIDIデータに変換する

亀岡 弘和 齊藤 翔一郎 西本 卓也 小野 順貴 嵯峨山 茂樹
(嵯峨山グループ：東京大学大学院情報理工学系研究科 嵯峨山研究室)

概要 iPodをはじめとした大容量HDDを搭載したオーディオプレイヤの登場により、音楽は「大量保有、選択視聴」するものへと変わりつつある。そのような流れの中で、楽曲を効率よく手に入れられる音楽版Googleのようなシステムの出現が期待されるのは当然の帰結であるが、現在一般に流通している楽曲のデータ形式は楽曲の波形を表すもので、検索に適しているとはいえない。そこで我々は、楽曲の波形データを楽譜に近い情報量をもつMIDIデータに変換するためのシステムを開発した。このシステムは、学習理論に基づいた新しい多重音解析法であるHarmonic Clustering法を用いて高い精度でMIDI化を行うものである。また、多重音解析一般の問題の性質上、推定誤りや修正すべき箇所を容易に素早く発見でき、かつ負担なく修正できるような機能を備えたユーザ支援インタフェースも有している。

キーワード：MIDI変換, Harmonic Clustering, Specmurt.

1. はじめに

iPodをはじめとした大容量HDDを搭載したオーディオプレイヤの登場により、大量の楽曲を効率的に入手したいというリスナの欲求は強まり、インターネット音楽配信サービスは今後ますます普及してくると予想される。そのような流れの中で、人間の一生をかけても聴き切れないほど大量の楽曲の中から自分好みの楽曲だけを選択的に手に入れられるようなシステム(音楽版Google)の出現が期待されるのは当然の帰結と言えるだろう。例えば、同一ジャンルの楽曲をグループ化できたり、リスナの特定のお気に入り楽曲に類似した楽曲を検索できたり、著作権を侵害しない範囲で楽曲を自分好みに自在にアレンジができるようになれば、音楽産業のさらなる活性化が期待される。現在のところ一般的に流通している楽曲のデータ形式(wavやmp3形式など)は、音圧の物理的な振動パターンを単に記述しただけの波形データであるが、上記の目標を実現するには、楽曲それぞれに再生用の波形データだけでなく、検索用に楽譜に相当するような形態のデータも対応されているのが望ましい。録音技術がまだ存在しなかった時代には、作曲家達は楽譜を通して当時の作品を後世の人々に残すことができたわけであるが、どの楽器でどの時刻にどの音高で発音されるかという演奏内容を詳細に記述するデータ形式であるMIDI(Musical Instruments Digital Interface)はまさに現代における新しい形態の楽譜と言える。しかしながら、人手で世界中に配信されている楽曲すべてをMIDI形式に記述するのは、当然のことながら現実的とは言えない。そこで我々は、楽曲の波形データをMIDIデータに変換するためのシステムを開発することとした。

2. MIDI変換用ソフトウェア

波形データからMIDIデータに全自動で変換するには、複数混在する音源信号の基本周波数(音高に対応する物理量)やオンセット(発音開始時

刻)などを個別に求める必要があり、高度な信号処理技術を要する。筆者は修士課程在籍時に、学習理論に基づいた新しい多重音解析法(Harmonic Clustering法)を提案した。この手法は従来法に比べて極めて高い性能であることが確認されているが、多重音解析一般の問題の性質上、推定誤りは避けられない。従って、誤りを手修正する操作が容易に行えるような配慮は少なからず必要であると考えられる。そこで、Harmonic Clustering法による波形データからMIDIデータへの全自動変換機能と、そのステージにおいて生じた推定誤りや修正すべき箇所を容易に素早く発見でき、かつ負担なく修正できるような機能を備えたユーザ支援インタフェースの開発を目的とする。

2.1 音高推定

多重ピッチ解析は音響信号処理の分野ではいまだ未解決の懸案課題の1つである。MIDI自動変換ソフトウェアを実現するためには、音響信号から混在する各音の音高、オンセット、音長を正しく推定する方法論の開発が必要である。我々は、多重音パワースペクトルは、楕形の調波構造のスペクトル(有ピッチ音は一般に、知覚的に音高に関係すると言われる基本周波数と、その整数倍の周波数に強い成分をもつため、スペクトル領域で見ると楕形の構造をなす。)が複数重ね合わされたものであると考え、図1のような楕形構造をなす拘束つき混合ガウス分布モデルで1つの調波構造をモデル化し、その重みつき和で多重音スペクトルをモデル化することとした。このモデルのパラメータ

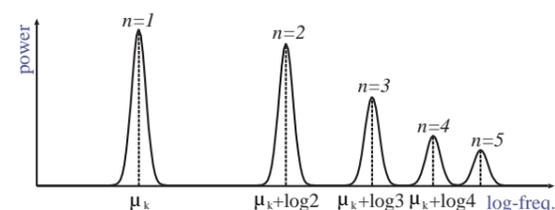


図1 調波構造モデル

は、EMアルゴリズムと呼ばれる効果的な反復アルゴリズムで局所最適推定することができ、複数音の基本周波数をはじめとする物理量を高い精度で同時測定できる。我々はこの原理をHarmonic Clusteringと呼んでいる。Harmonic Clusteringは基本的には短時間窓における多重ピッチ分析法であり、短時間窓をある一定周期でずらしながら解析していくと、推定されたピッチの点列を得ることができる。次に、このピッチ点列からオンセットと1音長を推定するのにHMM(隠れマルコフモデル)を応用した。詳細の原理についてはここでは割愛する。

2.2 音高可視化

変換MIDIデータの修正操作が技術的に不可欠であるとするならば、その操作が容易に行えるような配慮は少なからず必要であるが、容易に誤りや修正すべき箇所を視覚的に提示ができれば有用である。そこで、多重音信号の音高に関する視覚的手がかりを与える、音高可視化手法(Specmurt法)を簡単に紹介する。Specmurt法は、多重音スペクトルから基本周波数成分以外の成分を抑圧し、音高らしさを表す分布を出力するものである。まず図2のような多重音スペクトルモデルを考える。u(x)が基本周波数とその成分を表す分布のとき、h(x)のような楕形構造のパターンとu(x)の畳み込みにより、同一の調波構造パターンをもった構成音からなる多重音スペクトルv(x)が表現できる。これは、対数周波数領域では、調波構造の周波数の相対関係が基本周波数に依らず保たれ、全体が基本周波数に応じて単純にシフトするという性質に基づく。ここで、観測されたスペクトルをv(x)と考え、h(x)で逆畳み込みを行えば、u(x)に相当する分布が得られるであろうというのがSpecmurt法の考え方である。逆畳み込みは、フーリエ変換領域では除算になることを利用すれば、u(x)は容易に求められるわけだ。この処理は、観測スペクトルv(x)の中に調波構造があった場合に基本周波数成分以外の成分を抑圧する操作に相当し、音高らしさの分布が得ることができ

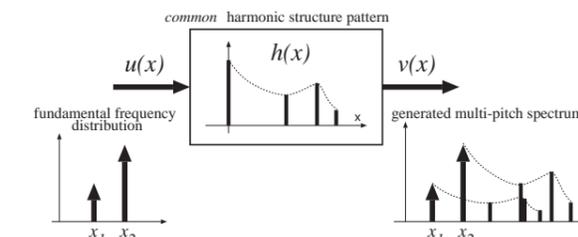


図2 基本周波数分布u(x)と楕形構造パターンh(x)の畳み込み表現による多重音スペクトルモデル

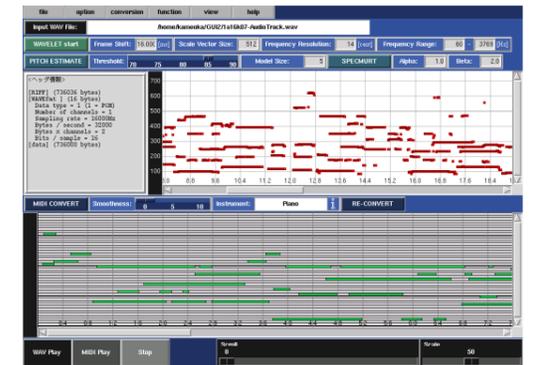


図3 作成したGUI. Harmonic Clusteringにより推定したピッチが赤点で表示される。



図4 「Specmurt」ボタンを押下すると、ポップアップで基本周波数分布(グレースケール濃淡表示)と変換MIDIデータ(緑色の矩形オブジェクト)を表示したウィンドウが現れる。緑色の矩形オブジェクトは、生成/消去/移動/伸縮でき、基本周波数分布の背景画を頼りにMIDIが編集できる。

3. 実装

Harmonic Clusteringにより全自動変換されたMIDIデータを、Specmurt法によって得られた音高可視化結果を背景画像にして矩形オブジェクトとして重ねて表示し、背景画像を参考に矩形オブジェクトの伸縮/移動によりMIDIデータが直接編集できるインタフェースを実装した。開発環境は以下の通りである。OSはLinux、内部の各モジュール(音高推定と音高可視化)はC言語、GUIはTcl/Tkを開発言語とした。GUIの概観を図3, 4に示す。

4. 終わりに

我々は、計算機のみによる全自動変換を行う音高推定モジュールと、容易に誤りや修正すべき箇所を視覚的に提示する音高可視化モジュールを両用することで、手早くお手軽に音楽の音響信号波形データからMIDIデータに変換するためのソフトウェアを開発した。

MIDIデータを楽譜に変換する

武田 晴登[†] 山本 遼 西本 卓也 小野 順貴 嵯峨山 茂樹

(東大グループ：東京大学大学院情報理工学系研究科嵯峨山研究室)

[†]2006.10からCRESTデジタルメディア領域片寄チーム研究員(関西学院大学)

概要 我々が開発した人間の器楽演奏のMIDI信号を記録したMIDIデータを楽譜に変換する技術を紹介する。MIDI信号には音高情報が含まれているので、演奏の音長から音価を推定できればMIDIデータを楽譜に変換できる。音の長さからリズムとテンポを求めるために、我々はこれまでにリズムとテンポの特性を確率的モデルで定式化し、確率的に最も尤もらしいリズムとテンポを求める手法を研究してきた。この手法によりMIDIデータを楽譜に変換できることを示す。

キーワード：自動採譜，隠れマルコフモデル，リズム認識，テンポ推定，事後確率最大化推定

1. はじめに

音楽をコンピュータで処理することは一般的になった今日、楽譜の情報もコンピュータ上で扱うことが多い。例えば、楽譜浄書などはパソコンで行われるのが一般的である。また、携帯電話の着信メロディーやカラオケの伴奏データ等のシンセサイザーを鳴らすためのデータの作成にも、楽譜に近いデータをコンピュータで処理している。さらに、音楽を検索するためにも、楽譜の情報が有用である。

このような重要な楽譜の情報は、今日でも人手で入力されることが一般的である。この入力作業は音楽の専門知識を要する人が「耳コピ」や楽譜を見ながら手作業で打ち込みを行う労力を要する仕事である。このため、楽譜や音楽コンテンツの制作の効率化や音楽検索技術の発展の為に、音楽演奏からの楽譜の自動変換(「自動採譜」と呼ばれる)技術が広く望まれている。

さて、ここでは、電子ピアノやMIDI(Musical Instrument Digital Interface)キーボードのように人間の演奏情報がMIDI信号で与えたときに、MIDI信号を楽譜に変換することを考える。楽譜の音価(time value; 時価ともいう)とは、四分音符の様に楽譜上の各音符が持つ論理的な音の長さを指す。実際に楽譜が演奏されると、音価はあるテンポのもとで演奏者によって演奏され、発音(note-on)と消音(note-off)のMIDIメッセージが生じた時刻を記録できる。各音符の音高(pitch)はすでにMIDIデータに含まれるので、逆に、MIDIデータを楽譜に変換するには、図1に示すように各音符の時刻情報から各音符の音価を推定すれば良い。本稿では、我々がこれまでに研究してきた実演奏からテンポとリズムの同時推定手法[1-3]とこの手法を用いた楽譜変換結果を紹介する。

2. リズムとテンポの同時推定

2.1 確率モデルを用いたリズム認識

観測された演奏された音長(note length)から、その音の記譜上の音符の長さである音価と、演奏

時のテンポの両者を推定する問題は、解を一意に決定できない優決定問題である。例えば、図2に示すように不自然な解釈でも成り立ちうるが、これに対して、人間はテンポを時間的に緩やかに変動するものとしつつ、演奏を典型的なリズムパターンに当てはめて聴くことで適切なリズムとテンポの解釈を得ているのであろう。我々は、このような音楽の特徴を次のようにモデル化した。

2.2 HMMによる音楽演奏のモデル化

音楽リズムのリズム語彙 リズムパターンの出現頻度には統計的な偏りがあるので、リズムパターンの表れ易さはリズムパターンの出現確率により評価できる。典型的なリズムの出現確率は高く、ほとんど現れないリズムの出現確率は低くなるように、既存の楽曲を用いて統計的に学習できる[1]。具体的には、図3に示すようにリズムの単位「リズム単語」からなる「リズム語彙」と、リズム単語のN-gramモデルによる文法を考え、リズムの状態遷移ネットワークで表現できる[1]。

音長の変動 実際に人間の音楽演奏では、演奏者の意図や無意識に起こす変動により、演奏される音の発音タイミングやテンポが変動する。この変動を確率的に扱うことにより、リズム単語の演奏は、図4のようにHMMでモデル化できる[1-3]。

テンポ曲線 演奏のテンポも一音毎に変動しているが、背後には時間に対して緩やかに変化するテンポ曲線があると仮定すれば、テンポ曲線からの確率的な変動としてモデル化できる[3]。

2.3 事後確率最大化推定によりリズムとテンポの同時推定

このように演奏を確率的にモデル化すると、リズム認識は与えられた演奏に対して最も尤もらしいリズムとテンポを求める問題として定式化できる。この事後確率最大化問題は、推定の一部は連続単語音声認識[4]と同型の問題として定式化されているので、音声認識で開発された時間同期Viterbi探索をリズム認識に応用し確率を効率的に計算できる[2]。

3. 実装

我々のリズム認識手法を用いて演奏のMIDIデータを楽譜情報に変換するプログラムを実装した。入力演奏を記録した標準MIDIファイルであり、出力はabc楽譜記述言語で記述されたファイルとした。図5は、実際に人間の演奏したテンポの変動の大きい演奏を楽譜変換した出力結果である。市販ソフトでは変動するテンポの推定を行わないので図6のように適切な音価を推定できないのに対し、提案手法で常識的な楽譜表現を出力している。この演奏は図7の楽譜を演奏したものであり、演奏者の意図したリズムを正しく推定したことになる。

4. まとめ

リズムとテンポの同時推定法を用いて、MIDIデータを楽譜に変換する手法を紹介した。今後は、強弱記号、表情記号、調号、楽器指定、声部の分離などを含んだ自動採譜システムを開発したい。

参考文献

- [1] 齋藤, 中井, 下平, 嵯峨山, “隠れマルコフモデルによる音楽演奏情報からの音符列推定,” 電気関係学会北陸支部連合大会講演論文集, F-62, pp.362, 1999.
- [2] 大槻, 齋藤, 中井, 下平, 嵯峨山, “隠れマルコフモデルによる音楽リズムの認識,” 情報処理学会論文誌, Vol.43, No.2, pp.245-255, 2002.
- [3] 武田, 西本, 嵯峨山, “HMMを用いたリズムとテンポの反復推定による多声MIDI演奏のリズム認識,” 日本音響学会2006年春季研究発表会 講演論文集, 3-2-3, pp.721-722 (in CD-ROM), Mar, 2006.
- [4] X. Huang, A. Acero, H-W. Hon: Spoken Language Processing: A Guide to Theory, Algorithm, and System Development, Prentice Hall, 2001.

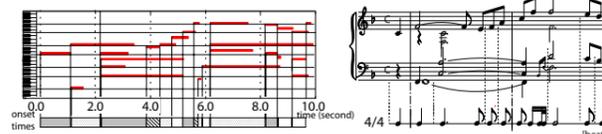


図1 MIDI演奏の自動採譜とリズム認識

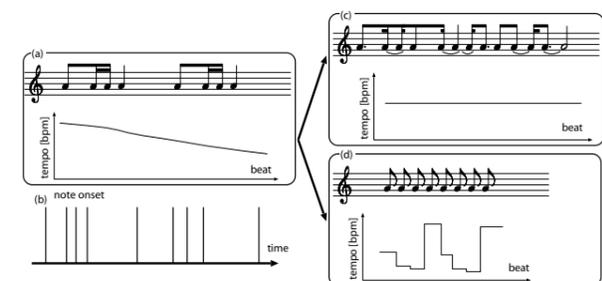


図2 音長からのリズムとテンポの解釈

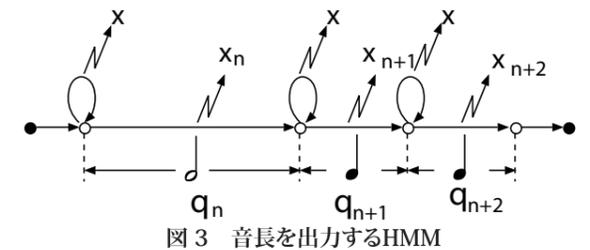


図3 音長を出力するHMM

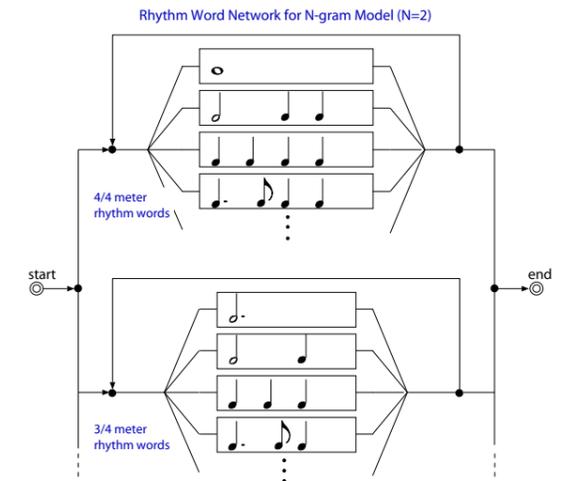


図4 リズム語彙: リズム単語のネットワークモデル



図5 システムの出力



図6 市販ソフトによる楽譜変換結果



図7 演奏曲のオリジナルの楽譜

体験デモ

自動伴奏システム

武田 晴登[†] 嵯峨山 茂樹

(東大グループ：東京大学大学院情報理工学系研究科嵯峨山研究室)

[†]2006.10からCRESTデジタルメディア領域片寄チーム研究員(関西学院大学)

概要 本稿では我々が開発した演奏者の演奏に合わせて伴奏を再生する自動伴奏システムを紹介する。本システムは、演奏者が演奏している楽譜上の位置をリアルタイムで追跡しつつ、演奏者のテンポに合わせて伴奏の再生を行う。演奏はMIDI楽器を使用し、再生にはMIDI音源を用いる。演奏者の演奏を確率的に扱い最も尤もらしい演奏位置を推定することで、従来は難しかった人間の実際の演奏では演奏誤りや弾き直しに対応した楽譜追跡を実現した。楽譜追跡手法と実装した自動伴奏システムの動作例を紹介する。

キーワード：自動伴奏, HMM (隠れマルコフモデル), Viterbi アルゴリズム, 楽譜追跡

1. はじめに

音楽は聴くだけでなく、演奏することでより深くその音楽作品を鑑賞できる。現在、MP3 (MPEG audio Layer-3)に代表される音声信号の圧縮技術の進展、また、iPodに代表される携帯音楽プレーヤの普及に伴い、ユーザは大量の音楽を聴くことができるようになり、音楽を鑑賞に関するIT技術の進化には目を見張るものがある。一方で、音楽を演奏して楽しむことへのIT技術の適用は、今でもカラオケや少数のゲームコンテンツに見られる程度に限られる。しかし、MIDI (Musical Instrument Digital Interface)規格や電子楽器が広く普及していることから、今後はこれらのハードを活かしたソフトウェアが増えていくと予想される。このような背景を踏まえ、我々は電子楽器、シンセサイザー、WEB上で入手が容易なMIDIファイルを用い、幅広い用途に使用できるアプリケーションとして自動伴奏システムを開発した。

我々が開発した自動伴奏システムは、楽譜をもとに電子楽器を演奏する演奏者に合わせて伴奏を再生するもので、いわば、演奏者の演奏を理解してテンポを合わせてくれる知能的なMIDI演奏用カラオケである。このシステムは、ミスタッチを含みテンポの変動する実際の人間の演奏にリアルタイムに追従し、演奏者のテンポを推定し、次の演奏者の拍を予想しつつ伴奏を再生する機能を有する。この楽譜追跡機能のより、ピアノの片手演奏のように初心者の練習から協奏曲などの本格的な練習とあらゆるレベルでの練習支援ツールとして使用できる。更に、任意のテンポで演奏を行う仮想的な指揮システムとしても使用できる。本稿では、この楽譜追跡手法と製作した自動伴奏システムを紹介する。

2. 多旋律MIDI演奏の楽譜追跡

2.1 楽譜追跡の従来研究

演奏者が楽譜をもとに演奏を行う場合、演奏者は楽譜どおりには弾かずにミスタッチや弾き直しが含まれることがある。このため、事前に演奏す

る曲の楽譜が与えられている場合でもシステムが演奏者の演奏位置をリアルタイムで特定することは自明な問題ではなく、音楽情報処理において重要な課題のひとつとなっている。現在の市販の電子楽器に組み込まれている自動伴奏機能は、演奏者の演奏誤りや弾き直しなどに十分対応していない。もし演奏が単旋律で弾き直しを含まないのであれば、Dannenbergが開発した音高の時系列をDP (Dynamic Programming, 動的計画法) マッチングする手法[1]により、多少のミスタッチが含まれていても楽譜追跡が可能である。しかし、演奏が和音を含み、かつ、弾き直しやスキップのように演奏箇所が途中で変わる演奏の楽譜追跡は、これまで限定されたタスクや条件のもとでの動作例が報告されているのみであった[2]。これに対して、我々はタスクに依存せずに使用できる汎用的な楽譜追跡手法を開発した[3]。

2.2 HMMを用いた楽譜追跡

以下、我々が開発した楽譜追跡手法[3]の概要を紹介する。我々は、人間の音楽演奏には、音の脱落・挿入・置換などのミスタッチと呼ばれる演奏誤りや、同時に発音することを意図した音の実際の発音時刻のずれや、あるいは、弾き直しなどの演奏箇所の遷移など、これら全ての演奏に見られる特徴を確率的に扱えると考えた。実際、これらの現象には、出現に偏りがあることが予想される。例えば、演奏のミスタッチは正しい音の鍵盤位置に対して近い音に間違えるであろう。これらの確率的な要素はHMM (Hidden Markov Model, 隠れマルコフモデル) としてモデル化でき、演奏に対して最適な楽譜とのマッチングを求める問題は、通常のHMMにおける最適状態系列を求める問題と同様にしてViterbi アルゴリズムで求まる[4]。これにより、高速に動作する楽譜追跡が実現できる。

3. 自動伴奏システムの実装と動作例

3.1 実装

システムは、楽譜追跡とテンポ推定のアルゴリズムはC言語で、実装し、GUIはCocoa Objective-C

を、MIDI信号の入出力にはCore audio を用いて、MAC OSX上で実装した。システムは、ノートブックパソコンMac Book Pro (CPU: Dual Core 2.2GHz)とMIDI音源内蔵の電子ピアノCASIO PX-300から構成される。

3.2 動作例

今回のインタラクティブセッションで用意したサンプル曲を紹介する。

ピアノ曲の片手練習

ピアノ練習においては、楽譜を初めて読むときに片手ずつ読んで練習することがある。演奏者が右手を演奏するときに、システムが左手のパートを伴奏として演奏することで、曲全体の感じをイメージしながら楽しく練習できる。

今回は、サンプルとして、初心者が練習するソナチネアルバム中の曲、難易度の高い曲としてショパン作曲の「幻想即興曲」「革命のエチュード」等を用意した。

メロディーの片手演奏への伴奏付け

本格的なピアノ演奏は技術的に難しいが、楽譜は読めて簡単なメロディーならば鍵盤楽器機で演奏できるという音楽愛好家が多い。演奏者の演奏したメロディーに合わせて本システムが伴奏することで、演奏をより楽しいものにできる。サンプル曲に、瀧廉太郎作曲の「花」を用意した。

ピアノを含むアンサンブル音楽の演奏

ピアノ協奏曲や室内楽のようにピアノを含むアンサンブル曲には、多くの人にとって演奏してみたいと思う名曲が多く、上級者のレッスンで取り上げられる曲もある。これらの曲を本格的に演奏するには一緒に演奏するメンバーや団体を見つけ練習場所を確保する必要があり、現実的にこれら全てを調整するのは難しく、練習や演奏の機会を十分に得られないことが多い。

本システムでは、演奏者がピアノパートを演奏し、オーケストラや弦楽等の他のパートをシステムが演奏させることで、室内楽や協奏曲の演奏を楽しむことができる。サンプルとして、ドヴォルザーク作曲のピアノ5重奏曲、モーツァルトのピアノ協奏曲第23番からの抜粋を用意した。

オーケストラの演奏

人間がオーケストラのあるパートを演奏し、それに合わせて伴奏を演奏させることで、演奏者の意図したテンポでオーケストラを演奏できる。今回、サンプルとしてチャイコフスキー作曲の「弦楽セレナーデ」を用意した。演奏者が第1ヴァイオリンを演奏し、それに合わせて弦楽オーケストラのパートが再生できる。

4. まとめ

我々が開発した自動伴奏システムを紹介した。今後は、演奏者に追従するのみでなく、伴奏としても特定のスタイルで演奏できるよう拡張を行いたい。

謝辞

自動伴奏のソフトウェアは、武田が戦略ソフトウェア創造人材養成教育コースに受講時に作ったものが原型となっています。このコースにおいて、平木敬教授、稲葉真里特任助教授、鈴木隆文講師には有益な議論をして頂きソフトウェア開発に必要な環境を提供して頂きました。また、金子勇元助手、土村展之助手には実装について多くの適切な御指導を頂きました。本ソフトウェア制作を支えて頂いた皆様に深く感謝致します。

参考文献

- [1] R. B. Dannenberg: An On-line Algorithm for Real-Time Accompaniment, Proc. of International Computer Music Conference, pp. 193-198, 1984.
- [2] 大島, 西本, 鈴木, “家庭における子どもの練習意欲を高めるピアノ連弾支援システムの提案,” 情報処理学会論文誌, Vol. 46, No. 1, pp. 157-171, 2005.
- [3] 武田, 西本, 嵯峨山, "HMMによるMIDI演奏の楽譜追跡と自動伴奏," 情報処理学会研究報告(MUS), pp. 109-116, Aug, 2006.
- [4] Rabiner, L. R.: A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proc. IEEE, vol. 77, no. 2, pp.257-286, 1989.

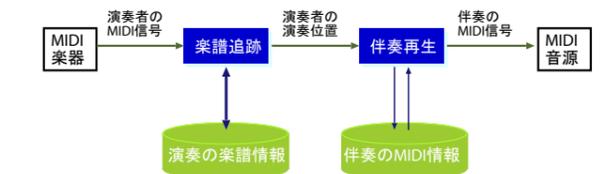


図1 自動伴奏の構成



図2 誤りのない演奏(左)と誤りのある演奏(右)の状態遷移

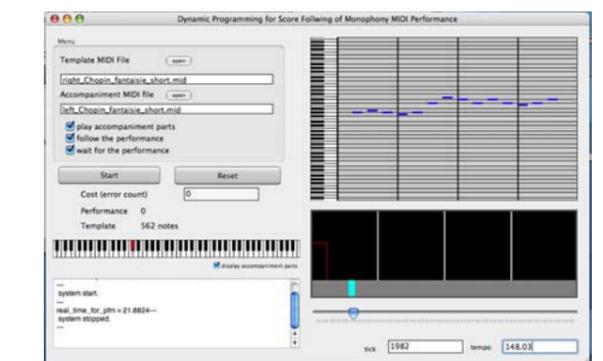


図3 ソフトウェアのGUI

体験デモ

Drumix: ポピュラー音楽CDのドラムパート 編集機能付きオーディオプレイヤー

吉井 和佳 後藤 真孝[†] 奥乃 博

(奥乃グループ: 京都大学情報学研究科奥乃研究室)

[†]産業技術総合研究所(AIST)

概要 オーディオプレイヤーDrumixは、市販CDに録音されているようなポピュラー音楽を再生する際に、ドラムスの音量や音色を自由に変更し、特別な音楽知識を必要とせずにドラムパターンを簡単に編集できるインターフェースを備えている。このインターフェース上での操作結果は再生される信号にリアルタイムに反映される。したがって、Drumixを用いれば、リスナは自分の好みに合うように音楽を作り変えるという能動的な音楽鑑賞を行える。このとき、リスナはエンドユーザとして音楽を聴く楽しみだけでなく、音楽を制作する楽しみも同時に味わうことができる。

キーワード: 能動的音楽鑑賞, ドラム音認識, ビートトラッキング

1. 能動的な音楽鑑賞

従来の音楽鑑賞は、聴きたい楽曲をそのまま再生して楽しむだけの受動的な体験であった。そこで後藤らは、楽曲中の聴きたい箇所を容易に頭出しできるような、楽曲構造(サビの位置など)を表示できるサビ出し機能付き音楽試聴機SmartMusicKIOSK [1] を考案した。しかし、歌声に集中するため他の楽器パートの音量を下げる、ドラムスの演奏パターン(ドラムパターン)を自分の好みに合うように編集しながら音楽鑑賞をするといった、より能動的な音楽鑑賞を実現する技術はこれまで存在しなかった。

我々は、市販CDレベルの複雑な音響信号中のドラム音認識技術に基づく新しい音楽再生インターフェースを開発した。これを用いれば、ドラムスの音量や音色、ドラムパターンなどをリアルタイムにコントロールできる。例えば、楽曲のビート感を強調したければ、ドラムスの音量を上げればよい。ドラムスの音色が気に入らなければ、好みの音色に差し替えてもよい。また、ドラムパターンは楽曲のグルーブ感(ノリ)と密接に関係しているため、それを変更することで、楽曲の印象を変えながら音楽を楽しむことができる。

今回開発したオーディオプレイヤーDrumixでは、誰でも直感的にドラムパート編集が行える。音量を調節するにはスライダを移動させればよく、音色を差し替えるには、リストから好きな音色を選択すればよい。また、視覚化されたドラムパターン集合から適当なパターンを選択するだけでドラムパターンを変更できる。

2. ドラムパートのリアルタイム編集

オーディオプレイヤーDrumixは通常のプレイヤー機能(再生・停止など)の他に、3種類のドラムパート編集機能(音量調節機能・音色置換機能・ドラムパターン編集機能)を備える。これら3つの機能は組み合わせることができる。インターフェースは音楽経験に関係なく直感的に操作が可



図1 音量調節・音色置換インターフェース

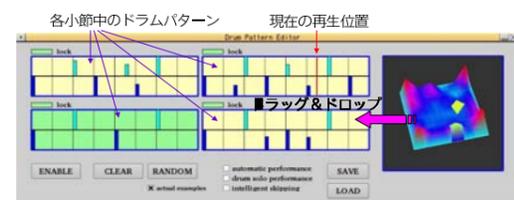


図2 ドラムパターン編集インターフェース

能なものとした。操作結果はリアルタイムに反映されるため、ユーザは自分の想像した操作結果と聴き比べながら音楽鑑賞することができる。

このような操作により、エンドユーザ(楽曲購入者)としてだけでなく、楽曲制作サイドの楽しみが味わえる。通常の楽曲制作では、作曲されたメロディーに対し、編曲者が適切なドラムスの音色を吟味し、ドラムパターンを付与する。その後、ミキシングエンジニアがドラムスの音量バランスを考慮しながら1つの曲へと仕上げる(ミックスダウン)。Drumixを使えば、ドラムパートの編集とミックスダウンを手軽に体験できる。以下に Drumix の使い方を機能ごとに説明する。

2.1 ドラムス音量調節機能

ユーザはミキシングエンジニアになった感覚が味わえる。音量スライダ(図1参照)を左右に移動させることで、バスドラムとスネアドラムの音量を個別に調節できる。スライダの左端が最小音量(ミュート)に、右端が最大音量に対応する。

2.2 ドラムス音色置換機能

ユーザは編曲者と同様に、ドラムスの音色を吟味しながら音楽鑑賞ができる。複数の音色がリストアップされたドロップダウンリスト(図1参照)から好きな音色を選択することで、楽曲中のドラムスの音色を差し替えて再生できる。音色はあらかじめいくつか登録されているが、ユーザ自身があとで追加登録することもできる。

2.3 ドラムパターン編集機能

ユーザは編曲者の仕事の一つであるドラムパターン作成を体験できる。どの時刻にどの音量でドラムスが発音するのか一目瞭然に把握可能なグラフィカルな譜面表示上(図2参照)でドラムスの譜面を編集できる。4つの長方形内がそれぞれ1小節内(4/4拍子を仮定)のドラムパターンを表す。それぞれ下段はバスドラム、上段はスネアドラムに対応し、横軸が時間、縦軸が音量を表す。さまざまな高さのバーはドラムスの発音に対応し、その高さが音量を表す。もとの楽曲中のドラムパートはこの4小節の繰り返しに差し換えられて再生される。このとき、現在の再生位置を表す縦線が小節内を随時移動していく(右下端に到達すると左上端に戻る)。ドラムスの発音バーはクリックするだけで追加と消去ができる。また、バーの上端をドラッグすれば、バーの高さ、つまり音量を変化させることができる。作成したドラムパターンはセーブとロードが可能である。

2.4 音楽経験の少ないユーザのサポート

何も参考にせずドラムパターンを作成するにはある程度の音楽経験が必要であるので、音楽経験が少ないユーザに対する支援が必要である。そのため、すでに完成されたドラムパターンを雛形にして編集できるような工夫を施した。

図2中右側の3D表示は、あらかじめ大量の楽曲から集めたドラムパターンをクラスタリングし、36個(6x6)のセルに対応付けたものである。山の高いセルはそこに対応付けられたドラムパターン数が多い(出現頻度が高い)ことを示す。ユーザは適当なセルからドラッグを開始し、いずれかの小節内にドロップすることで、そのセルに対応付けられた類似したドラムパターンのうちいずれかをランダムで雛形にすることができる。この操作によって意外性のある体験ができ、好みのドラムパターンを探す楽しみが生まれる。

2.5 好みに合った聴き方のサポート

煩わしい編集作業を一切することなく楽しみたい人のため、小節ごとにランダムでドラムパターンを変化させながら再生する機能(ランダム演奏機能)も追加した。再生中に気に入ったドラムパターンが現れれば、クリック一つでその小節のドラムパターンを固定できる。一方、ドラムパターンの編集作業に集中したい人のため、他の楽器音をミュートして、ドラムパートのみを再生する機能(ソロ演奏機能)も追加した。

3. システムアーキテクチャ

これまで、音響信号をMIDIのように楽器単位で処理することは技術的に困難であり、実現できていなかった。本章では、Drumixの各機能を実現する上での問題点と解決方法を述べる。詳細は文献[2]に記述されている。

3.1 ドラムスの音量調節・音色置換機能

ドラムスの音量調節と音色置換機能を実現するには、(1)市販CDレベルの複雑な音響信号中のドラム音のスペクトログラムを推定し、(2)発音時刻を検出する必要がある。この問題を解決するため、我々が以前考案したドラムスの発音時刻検出手法AdaMast [3]を利用した。AdaMastは、ドラム音のスペクトログラムをテンプレートとするテンプレート適応部とテンプレートマッチング部から構成され、前者が問題(1)を後者が問題(2)をそれぞれ解決する。ドラムスの音量を調節するには、推定されたドラム音のスペクトログラムを変化量に合わせて伸縮させた後、各発音時刻においてもとの楽曲のスペクトログラムに足したり、そこから引いたりすればよい。音色を置換するには、もとの楽曲中のドラム音をキャンセルした後、差し替えたい音色のドラム音信号を加算する。

3.2 ドラムパターン編集機能

ドラムパターンの編集機能を実現するには、複雑な音響信号中の拍時刻(1~4拍目の時刻)を推定する必要がある。この問題を解決するため、後藤らの提案したマルチエージェントに基づくビートトラッキング手法[4]を利用した。ビート解析を行った後、もとの楽曲中のドラム音をキャンセルし、編集した譜面に従って差し替えたい音色のドラム音を加算する。また、ドラムパターンをクラスタリングして可視化するアイデアはPampalkらの“Islands of Music”[5]を参考にした。

参考文献

- [1] 後藤真孝: SmartMusicKIOSK: サビ出し機能付き音楽試聴機, 情報処理学会論文誌, vol. 44, no. 11, pp. 2737-2747 (2003).
- [2] 吉井和佳, 後藤真孝, 奥乃博: 実世界の音楽音響信号に対するドラムスの音源同定手法を利用したドラムイコライズシステムINTER:Dの開発, FIT2004, LG-005.
- [3] K. Yoshii, M. Goto, and H.G. Okuno: AdaMast: A drum sound recognizer based on adaptation and matching of spectrogram templates, MIREX2005.
- [4] M. Goto: An audio-based real-time beat tracking system for music with or without drum-sounds, New Music Research, vol. 30, no. 2, pp. 159-171, (2001).
- [5] E. Pampalk, S. Dixon, and G. Widmer: Exploring music collections by browsing different views, ISMIR2003, pp. 201-208.

実演デモ

Instrogramによる類似楽曲検索

北原 鉄朗 後藤 真孝[†] 奥乃 博

(奥乃グループ: 京都大学大学院情報学研究所奥乃研究室)
[†]産業技術総合研究所(AIST)

概要 楽器構成の類似度に基づく音楽検索システムを紹介する。どのような楽器で演奏されているかは楽曲の雰囲気や大きな影響を与えるため、楽器構成に基づいた類似楽曲検索は様々な用途において有用と考えられる。我々は、楽器構成に基づく類似楽曲検索をInstrogramと呼ばれる楽器存在確率の時間・周波数表現に基づいて実現した。この方法では、各楽器存在確率を全時刻・全周波数に渡ってしらみつぶ的に求めていくため、従来の楽器音認識手法の多くで不可欠であった各単音の発音時刻や音高の推定をせずに求めることができる。

キーワード: 楽器音認識, 楽器存在確率, 音楽情報検索, 音楽情景分析, 聴覚的情景分析

1. はじめに —なぜ楽器音認識か—

本研究では、楽器構成に基づいて楽曲を検索する技術の確立を目指している。「ピアノソナタ」「弦楽四重奏」などのような分類が用いられることから分かるように、「どのような楽器で演奏されているか」は、その楽曲を特徴づける重要な要素といえる。楽器構成に基づく楽曲検索にはいくつかの方式が考えられるが、ここでは、いわゆるQuery-by-Example方式を取り上げる。この方式では、ユーザがある楽曲を指定すると、システムはその楽曲と楽器構成の近い楽曲を探し出す。これは、たとえばBGMのプレイリスト自動生成などに有用な技術である。

このような検索を実現する上でキーとなる技術は、音響信号からの楽器の認識である。楽器音認識は、従来単一音を対象としたものがメインであり、実際の楽曲へは応用できなかった。1990年代に入り混合音を対象とした研究が少しずつ増えてきたが、その多くは、各単音(1つの音符に対応する一単位の音)の発音時刻や音高(F0)をあらかじめ推定する必要があった。

本研究では、発音時刻や音高をあらかじめ推定しなくても楽器音の認識が可能な技術開発に取り組んでいる。この技術は、Instrogramと呼ばれる楽器存在確率の時間・周波数表現に基づいている。従来の楽器音認識が、まず時間・周波数平面的にどこに単音があるか(いつ、どの音の高さで単音が鳴るか)を探し出してから、その成分を解析するのに対し、本技術では、各楽器が当該時刻・当該周波数において発音している確率(楽器存在確率)を全時刻・全周波数に渡って網羅的に計算する。すなわち、楽器音解析処理をしらみつぶしに行うことで、前処理としての発音時刻・音高推定を省略できるようにしたと考えることができる。

2. Instrogram —楽器構成の確率表現—

Instrogramは、楽器構成を表す視覚表現である。対象楽器の各々に1枚ずつ画像が存在し、各

画像がその楽器が存在する確率を表す。具体的には、各画像の横軸が時刻、縦軸がF0を表し、各点 (t, f) の色が強さによって、時刻 t において周波数 f をF0とする当該楽器音が存在する確率を表す。図1に、「蛍の光」(三重奏)をピアノ、バイオリン、フルートで演奏した音響信号に対して、ピアノ、バイオリン、クラリネット、フルートを対象に求めたInstrogramを示す。図より、この楽曲は高音部はフルート、中音部はバイオリン、低音部はピアノによる演奏であることがわかる。

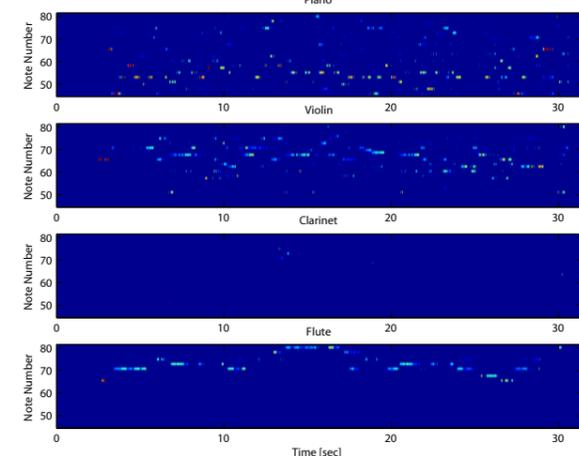


図1 Instrogramの例(ピアノ、バイオリン、フルートによる「蛍の光」(三重奏))

3. Instrogramに基づく楽器音認識

上で述べたようにInstrogramはどんな楽器による演奏かを確率的に表現するものだから、本研究ではInstrogramを求めることこそが楽器音認識であると考えられる。Instrogramを自動的に求める手法の詳細は文献[1,2]に譲り、基本的な考え方のみ述べる。ここで言うべきことは、対象楽器ごとに、各時刻・各F0においてその楽器の音が存在する確率(楽器存在確率)、つまり、対象楽器を $\Omega = \{\omega_1, \dots, \omega_m\}$ とすると $i = 1, \dots, m$ に対して、 $p(\omega_i; t, f)$ を求めることである。これは、いくつか

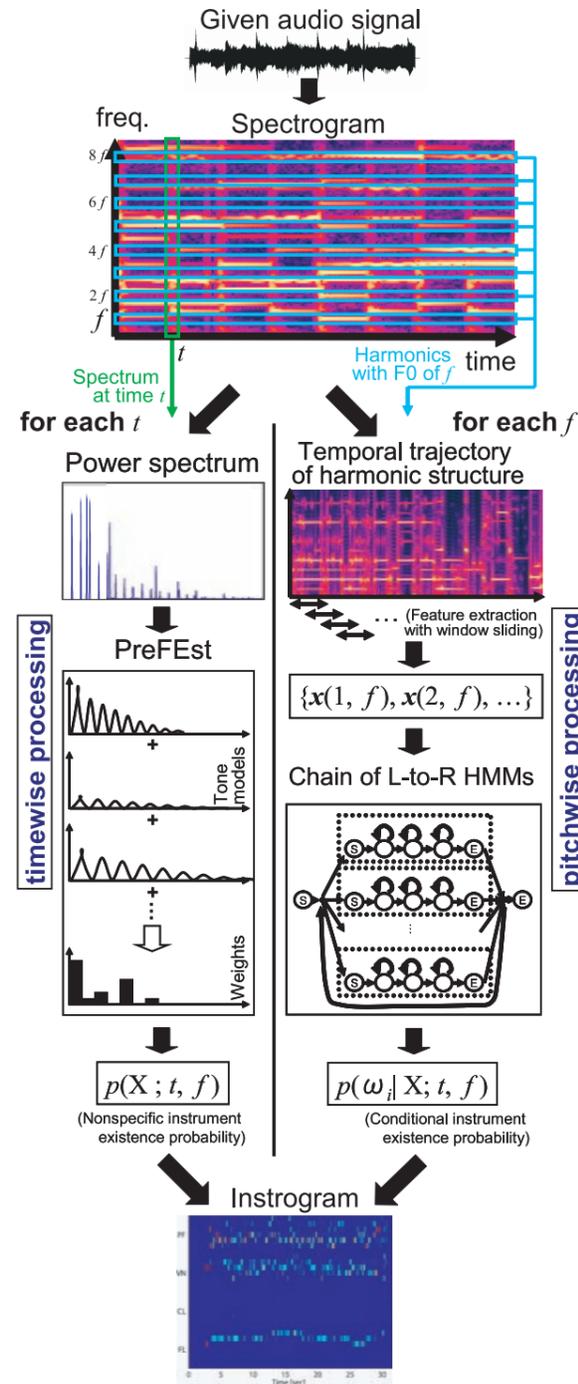


図2 Instrogram作成手法の処理の流れ

の仮定(文献[1,2]参照)をおくことで、次の2つの確率の積で表すことができる。

$$p(\omega_i; t, f) = p(X; t, f) p(\omega_i | X; t, f)$$

ここで、 $p(X; t, f)$ は不特定楽器存在確率といい、「時刻 t において f をF0とする何らかの音が存在する確率」を表し、 $p(\omega_i | X; t, f)$ は条件つき楽器存在確率といい、「時刻 t において f をF0とする何らかの音が存在するとすると、その楽器が ω_i である確率」を表す。入力された音響信号から不特

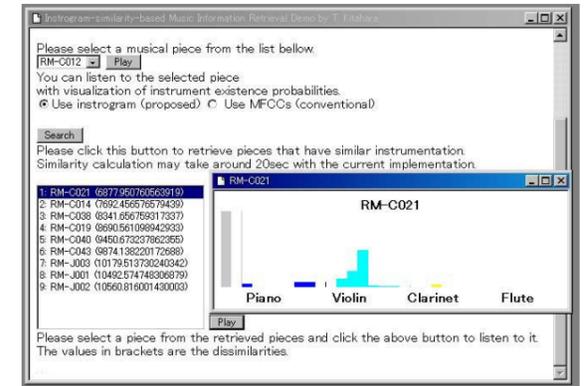


図3 Instrogram間の類似度に基づく音楽検索

定楽器存在確率と条件つき楽器存在確率を計算する処理の流れを図2に示す。不特定楽器存在確率は、時刻ごとにPreFEst [3]という手法を使って求める。条件つき楽器存在確率は、F0ごとに隠れマルコフモデル(HMM)に基づく音声認識の考え方を応用して求める。

4. Instrogramを用いた類似楽曲検索

Instrogram間の類似度を計算することで、楽器構成の類似度に基づいた楽曲検索を実現した(図3)。この検索では、ユーザがある楽曲を指定すると、システムはその楽曲とデータベース中の各曲との間の楽器構成の類似度を計算し、類似度が高い順に楽曲を並べて表示する。ユーザがそのなかから1曲選ぶと、システムは楽器存在確率をリアルタイムに表示しながら指定された楽曲を再生する。この表示は、各楽器の存在確率が棒グラフとして表され、音楽の再生にあわせて時々刻々と変化するもので、よく音楽プレーヤーに搭載されているディスプレイ上のスペクトル表示を楽器存在確率に置き換えたものである。このサンプルを<http://winnie.kuis.kyoto-u.ac.jp/~kitahara/instrogram/>に掲載する。類似度計算法の詳細は文献[2]に記す。

参考文献

- [1] T. Kitahara, M. Goto, K. Komatani, T. Ogata, and H. G. Okuno: Instrogram: A New Musical Instrument Recognition Technique Without Onset Detection Nor F0 Estimation, Proc. IEEE Int' l Conf. Acoust., Speech, Signal Proc. (ICASSP 2006), Vol.V, pp.292-232 (2006).
- [2] 北原 鉄朗, 後藤 真孝, 駒谷 和範, 尾形 哲也, 奥乃 博: Instrogram: 発音時刻検出と F0 推定の不要な楽器音認識手法, 情報処理学会研究報告 2006-MUS-66, pp.69-76 (2006).
- [3] M. Goto: A Real-time Music-scene-description System: Predominant-F0 Estimation for Detecting Melody and Bass Lines in Real-world Audio Signals, Speech Communication, Vol.43, No.4, pp.311-329 (2004).

実演デモ

混合音中の歌声と歌詞との時間的対応付け

藤原 弘将 後藤 真孝[†] 奥乃 博

(奥乃グループ: 京都大学大学院情報学研究所奥乃研究室)
[†]産業技術総合研究所(AIST)

概要 本研究では、伴奏音を含む音楽音響信号と対応する歌詞の時間的対応付け手法について述べる。クリーンな音声信号とその発話内容の時間的対応付けを推定するViterbiアラインメント手法はこれまで存在したが、歌声と同時に演奏される伴奏音の悪影響で市販CD中の歌声には適用できなかった。本研究では、この問題を解決するため、歌声の調波構造を抽出・再合成することで混合音中の歌声を分離する手法、歌声・非歌声状態を行き来する隠れマルコフモデル(HMM)を用いた歌声区間検出手法、音響モデルを分離歌声に適応させることでViterbiアラインメントを適用する手法を提案する。また、本手法の出力結果を用いて、楽曲の再生と同期して歌詞を表示する音楽再生インタフェースを実装した。

キーワード: アラインメント, 時間的対応付け, 歌詞

1. はじめに

歌声とその歌詞は、楽曲の主題やストーリーを表現し、楽曲を特徴づける要素の一つであるため、ポピュラー音楽を始めとする多くのジャンルの音楽で重要な役割を果たしている。実際、人が楽曲を聴く際には、歌声のメロディを聴き、その歌詞を耳で追うことが多い。そのため、音楽ビデオやテレビの音楽番組の中には、演奏の映像と同期して歌詞を表示することで、視聴者の楽曲鑑賞の手助けをするものもある。

本研究では、伴奏音を含む音楽とその歌詞の時間的対応付け手法について述べる。つまり、与えられた音楽音響信号と対応する歌詞をアラインメントすることで、歌詞の各フレーズの開始時刻と終了時刻を推定する。対象データは、市販CDなどの音楽音響信号であり、歌声に加えて他の楽器音を含んでいる。また、推定された結果を用いて、楽曲の再生と同期して歌詞を表示する音楽再生インタフェースについても述べる。(図1)

Wangら[1]は、我々と同様の問題に取り組んでいた。しかし、手法は、音韻的特徴を考慮せず、歌詞中の各音素の発声長の情報のみを用いていた。各音素の発声長は、楽曲中での登場位置によって大きく異なるため、この手法は不十分であった。

2. 本研究の課題

本研究では、歌声と歌詞の対応付けには、歌声の音韻的な特徴を考慮することが必要であると考え、音声認識で使われるViterbiアラインメント(強制アラインメント)を適用するアプローチをとる。しかし、このとき、歌声と共に演奏される伴奏音や、歌が歌われない間奏部の存在が問題となる。現行の音声認識で用いられるアラインメント手法はクリーンな音声を対象とするため、伴奏音を含む音楽とその歌詞の時間的対応付けに用いることは出来なかった。

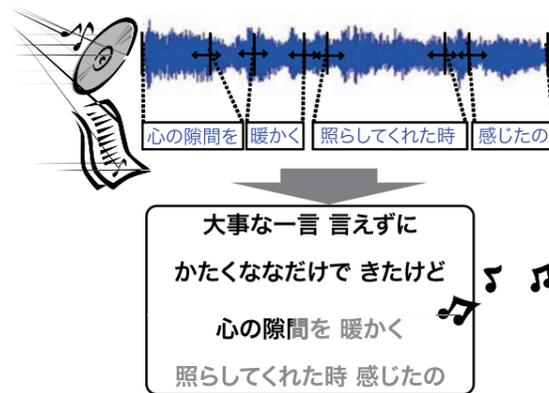


図1 音楽と歌詞の時間的対応付け

3. 歌声と歌詞の時間的対応付け手法

この問題に対処するため、本研究ではまず我々が以前提案した、伴奏音抑制を適用する。この手法では、メロディの調波構造を抽出・再合成することで、歌声を含むメロディのみが分離された音響信号を得る。次に、歌声・非歌声状態を行き来する隠れマルコフモデル(HMM)に基づく歌声区間検出を用いて、分離されたメロディから実際に歌声が存在する区間を検出する。最後に、Viterbiアラインメントを用いて、分離歌声と歌詞のアラインメントをする。また、ここでは、音響モデルを特定歌手の分離歌声に適応させる手法についても述べる。以下に各手法の概要について述べる。詳細については、文献[2]を参照。

3.1 伴奏音抑制

混合音からメロディのみの音響信号を得るため、メロディの調波構造を抽出し、再合成する。これらの処理を、以下のように行う。

まず、後藤のPreFEst[3]を用いて、メロディのF0を推定する。PreFEstは、制限された周波数帯域において最も優勢な調波構造のF0をEM法を用いて推定する手法である。

次に、推定されたF0に基づき、歌声の調波構造

の各倍音成分のパワーを抽出する。それぞれの周波数成分の抽出の際には、前後20centずつの誤差を許容し、その範囲で最もパワーの大きなピークを抽出する。

最後に、抽出された調波構造を正弦波重量モデルに基づき再合成する。このとき、各フレーム間の周波数が線形に変化するよう、位相の変化を2次関数で近似する。また、各フレーム間の振幅の変化は1次関数で近似する。

3.2 歌声区間検出

伴奏音抑制手法によって得られたメロディの音響信号は、間奏部などでは歌声以外の楽器音を含んでいる。そこで、それらの非歌声区間を歌声状態と非歌声状態を行き来するHMMに基づく歌声区間検出手法を用いて除去する。

具体的には、入力音響信号から抽出された特徴ベクトル列に対して、上述のHMMから出力されたと仮定した場合の歌声・非歌声状態の最尤経路を求める。各状態の出力確率は混合ガウス分布の確率密度関数で近似する。

特徴量は、LPCメルケプストラム(LPMCC)と $\Delta F0$ の二種類を用いる。LPMCCはLPCスペクトルから計算されたメルケプストラム係数であり、歌声の音韻的特徴を表現する。 $\Delta F0$ はF0の微分係数であり歌声の動的な性質を表現する。

3.3 Viterbiアラインメント

与えられた歌詞を元に作られたアラインメント用の文法と分離歌声から抽出された特徴ベクトルを用いて、Viterbi(強制)アラインメントを行う。また、アラインメントに用いる音響モデルを、入力音響信号中の特定歌手に適応させる手法についても述べる。

まず、与えられた入力音響信号に対応する歌詞を用いて、アラインメントに用いる文法を作成する。本研究では、アラインメントの際に、調波構造が安定して抽出出来る母音のみを用いる。具体的には、まず歌詞を音素列に変換し、その後、以下の三つの規則を用いて文法に変換する。

- 撥音を表す音素以外の子音を削除する。
- 歌詞中の文やフレーズの境界を複数回のショートポーズに変換する。
- 単語境界を一回のショートポーズに変換する。

アラインメントに用いる音響モデルを、入力楽曲中の特定歌手に適応させる。音響モデルとしては、大量の歌声のデータから学習されたモデルを使用することが理想的であるが、現段階ではそのようなデータベースは構築されていない。そこで、本研究では初期音響モデルとして、話し声用の音素HMMを利用する。

適応手法は、以下の3段階からなる。

- (1) 話し声用の音響モデルを単独歌唱の歌声に適応させる。
- (2) 単独歌唱用の音響モデルを伴奏音抑制手法によって抽出された分離歌声に適応させる。

(3) 分離歌声用の音響モデルを入力楽曲中の特定楽曲に適応させる。

(1)と(2)は教師あり適応で、事前に行われる。一方、(3)は教師なし適応で、認識時に行われる。ここで、教師情報とは、各音素の時間情報(音素の始端時間、終端時間)を指している。つまり、教師あり適応には、手動で付与した時間情報により正確にセグメンテーションされた音素データを用いる。適応時のパラメータ推定には、MLLRとMAPを組み合わせた手法を用いた。

最後に、歌詞を元に生成された文法、分離歌声の信号から抽出された特徴量と特定歌手に適応された音響モデルを用いて、Viterbiアラインメントを行う。特徴量は、MFCC, $\Delta MFCC$, Δ パワーを用いた。

4. 評価実験と再生インタフェース

本研究で述べた全ての手法を実装し、歌声と歌詞との時間的対応付けシステムを開発した。RWC音楽データベースから選ばれた10曲に対して評価実験を行うと、10曲中8曲に対して90%以上の精度で対応付けすることが出来た。ただし、評価はフレーズ単位で行われ、楽曲長に対する正しくフレーズがラベル付けされた区間の割合である。その他、実装・評価条件の詳細は文献[2]に順ずる。

また、このシステムの出力結果を用いて、再生と同期した歌詞のリアルタイム表示機能と、歌詞を用いて楽曲中のユーザの望む地点の頭出しが出来機能を持った音楽再生インタフェースを開発した。楽曲の再生中に、再生中の歌詞がリアルタイムに表示されるため、ユーザはより楽曲の鑑賞を楽しむことが出来る。また、ユーザが、表示される歌詞の聞きたい箇所をクリックすると、その場所から再生が開始されることで、自分好みのやり方で音楽鑑賞することが出来る。

参考文献

- [1] Wang, Y., Kan, M.-Y., Nwe, t. L., Shenoy, A. and Yin, J.: LyricAlly: Automatic Synchronization of Acoustic Musical Signals and Textual Lyrics, In. *Proc. of the ACM Multimedia*, pp.212-219 (2004).
- [2] 藤原弘将, 後藤真孝, 緒方淳, 駒谷和範, 尾形哲也, 奥乃博: 音楽音響信号と歌詞の時間的対応付け手法: 歌声の分離と母音のViterbiアラインメント, 情報処理学会研究会報告, 2006-MUS-66, pp.37-44 (2006).
- [3] Goto, M.: A real-time music-scene-description system: predominant-F0 estimation for detecting melody and bass lines in real-world audio signals, *Speech Communication*, Vol.43, No.4, pp.311-329 (2004).
- [4] 後藤真孝, 橋口博樹, 西村拓一, 岡隆一: RWC研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース, 情報処理学会論文誌, Vol.45, No.3, pp.728-738 (2004).

実演デモ

混合音中のパートの音量を操作可能なオーディオプレイヤー

糸山 克寿 吉井 和佳 奥乃 博

(奥乃グループ：京都大学大学院情報学研究科奥乃研究室)

概要 楽譜をもとに混合音の音響信号からそれぞれのパートの音響信号を分離し、それらを再び合成して再生することのできるオーディオプレイヤーです。ユーザは各パートの音量を自由に操作して、聴きたいパートの音量を大きくしたり、コンポーザのように楽曲のミキシングを行ったりすることで、能動的な音楽鑑賞を楽しむことができます。

キーワード：音源分離，モデル適応，多重音解析

1. 能動的音楽鑑賞とは

音楽を聴くためのツールやインタフェースは、これまでに様々なものが提案・開発されてきた。特に、グラフィックイコライザを持ったオーディオプレイヤーやエフェクト機能を持ったサウンドカードは音楽の聴こえ方を変化させることによって、新鮮な気分で音楽鑑賞を楽しみたいというユーザの要求に応えるようとしている。これらのツールでは、簡単な信号処理が用いられており、楽曲に用いられている全楽器パートに対して一様に処理が行われる。しかし、このような操作は必ずしもユーザの要求に答えるものではない。

ユーザの欲しい機能は周波数帯域に対する操作ではなく、「ベースパートを抑制して、ギターパートを強調したい」「ボーカルだけにエフェクトをかけたい」といった、楽器パートごとに対することが多いのではないだろうか。このようなユーザの要求に応えられる技術として、吉井らはDrumixを開発している[1]。ユーザはDrumixを使ってドラムスの音色を置換え、また、ドラムパターンを編集できるので、能動的な音楽鑑賞という新たな鑑賞法が可能となった。ただ、Drumixは楽曲中のドラムスだけを対象としていた。

本研究では、Drumixの目指した能動的な音楽鑑賞をドラムスだけでなく、ピアノやギターなどの一般的な楽器にまで対象を拡張することが目的である。すなわち、楽曲に含まれるそれぞれの楽器音が構成するパートに対して、ユーザが音量などを自由に操作できるオーディオプレイヤーを実現することである。

2. インタフェースの設計

オーディオプレイヤーとして設計したインタフェースを図(1)に示す。このインタフェースは、3つの部分で構成されている。上部はプレイヤー状態の表示および制御部分で、再生ボタン・停止ボタン・再生位置スライダーからなっている。中央部はそれぞれのトラックの楽器名を表示

し、その楽器パートのボリュームを操作スライダーが配置されている。このスライダーを操作することで、ユーザは楽曲作成におけるミキシングのように、能動的な音楽鑑賞を行うことができる。また、「sync」を複数のトラックでチェックし、チェックされたいずれか一つトラックのボリュームを操作すると、他の「sync」がチェックされたトラックのボリュームも連動して変化している。これは、複数の楽器が同じパートを演奏しているとき、その複数楽器のボリュームをまとめて操作したいときに有効だと考えられる。下部はボリュームを一括して操作する部分で、全てのトラックのボリュームを最小・最大・逆転・ランダムに変化させることができる。

このインタフェースを使い、各パートの音量を操作して楽曲を聴くデモンストレーションを行う予定である。

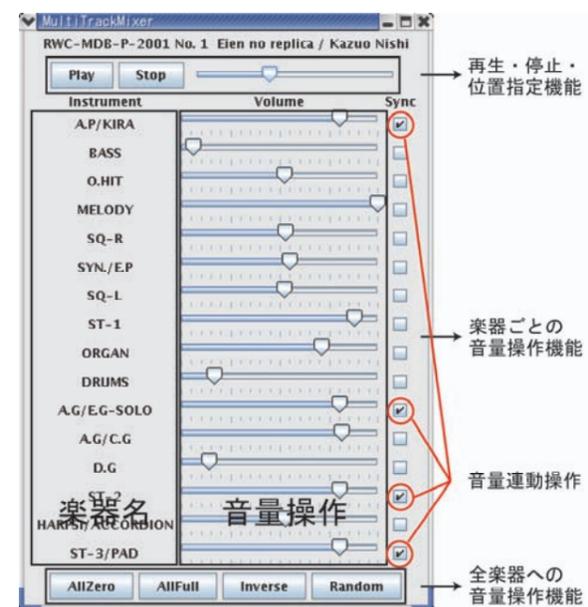


図1 プレイヤーのインタフェース

3. 音源分離と楽器音モデル適応

混合音からなる楽曲のそれぞれの楽器パートごとに対して操作を行うためには、まず混合音からそれぞれの楽器パートを分離して、楽器パートごとの音響信号を作る必要がある。Drumixでは、ドラム音認識に基づいたドラムパート分離を行っていたのに対して、混合音中の一般楽器音認識システムの精度は高くないので、本システムでは楽譜を用いることとした。

3.1 楽譜を用いた音源分離

各パートを分離するために、楽曲に対応したMIDIファイルを用いる。MIDIファイルを用いた音源分離は、以下のような手順で行われる。

1. MIDIファイルから、音符単位でMIDIメッセージを抽出し、MIDI音源でそれぞれの音符をテンプレート音として録音する。
2. 対象楽曲をテンプレート音の加重混合とみなし、対象楽曲中から分離音を音符単位で推定する。
3. 分離音をパートごとにまとめる。

楽譜を使用しているからといって、楽器音認識がやさしくなるわけではない。通常、MIDI音源から録音された楽器音の音色は、実際に楽曲中で使われている楽器音の音色と一致しない。このことは、音源分離において正しく分離ができなくなる原因の一つである。また、同じ楽器が同じ音高、音長で演奏していても、わずかにビブラートが異なっていたりするなど、それぞれの音ごとに微妙な違いが存在する。

3.2 楽器音モデル適応

上記の問題を解決するため、楽器音テンプレートのモデル適応を行った。具体的には、まず楽器音テンプレートからそれぞれの音のモデルを作る。次に、同じ楽器に由来するモデルが同じように変化するような制約をかけながら、モデルを楽曲中の音色に近づけていく。このときのデータの処理の流れをデータの流れを図(2)に示す。処理の手順は以下のとおりである。

1. テンプレートから音符ごとのモデルを作る。
2. 対象楽曲をモデルの加重混合とみなし、対象楽曲中から分離音を音符単位で推定する。
3. 分離音とテンプレートから、音符ごとのモデルを更新する。
4. 2~3を、モデルが十分に分離音に近づくまで繰り返す。

モデル更新では、最初はテンプレート音をより重視して更新を行い、以降徐々に分離音を重視する割合を増やす。このような適合により、モデルを対象楽曲で使われている楽器音に近づけるようにしている。

3.3 楽器音モデルの設計

それぞれの楽器音をどのように表現するかは、本研究において重要な要素である。特に、ポピュラー音楽のような楽曲では、ドラムスのような調波構造を持たない音、シンセサイザでの合成音のような純粋に調波構造だけを持つ音、ピアノのような調波構造と非調波構造を併せ持つ音など、多様な楽器音が含まれている。

本研究では、楽器音を調波構成成分(パラメトリックに表現)と非調波成分(ノンパラメトリックに表現)を足し合わせたもので表現する。ドラムスなどの楽器音は、調波構成成分を持たないものとすればよい。このようなモデルを用いることの最大の利点は、楽曲中に含まれるあらゆる音と同じ枠組みから捉えることができる点である。

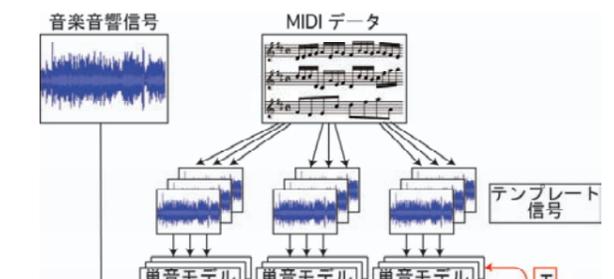


図2 モデル適応

4. おわりに

本研究では、混合音からなる楽曲の各楽器パートの音量を操作可能なオーディオプレイヤーを実現するためのテンプレートを用いた音源分離手法、さらにモデル適応手法を開発し、そのための楽器音モデルを作成した。また、プレイヤーのインタフェースを実装した。今後は、音源分離手法の定量的評価、プレイヤーへの機能の追加、さらにインタフェースの評価の被験者実験を行っていく予定である。

参考文献

- [1] 吉井和佳, 後藤真孝, 駒谷和範, 尾形哲也, 奥乃博: Drumix: ドラムパートのリアルタイム編集機能付きオーディオプレイヤー, インタラクシオン2006論文集, pp.207-208, (2006).

体験デモ

名演奏追体験インタフェース iFP

福岡 浩行, 酒造 祐介, 片寄 晴弘 奥平 啓太[†], 豊田 健一[†], 野池 賢二[‡]

(片寄グループ: 関西学院大学理工学部片寄研究室)

[†]元関西学院大学 [‡]元PRESTO

概要 我々は、表情のついた演奏をテンプレートとして利用し、拍打や指揮動作によって演奏を実施するインタフェース iFPの開発を行ってきた。iFPでは、1) 演奏表現の理解の手がかりとなる情報の可視化、2) 二つの演奏事例のモーフィング等の機能をサポートしている。iFPのユーザは名演奏家を指揮したり、名ピアニストの手を使って演奏をするような感覚を味わうことができる他、音楽教育、音楽解釈研究など、幅広い使用法が想定される。今回のデモンストレーションでは、演奏位置やテンポ情報などのユーザ提示を強化したiFP2の展示を行う。

キーワード: 演奏システム、能動的芸術鑑賞、音楽エンタテインメント、音楽教育

1. はじめに

「Conducting Program (指揮プログラム)を用いることで、演奏者は指揮びを気にすることなく、演奏表現でもっとも重要なフレージングのテンポ表現、音量表現に集中できる。」これは、ICMAの Video Review (vol.1)[1]の中でMax Mathewsが Radio Baton の説明をする際に語った言葉である。1983年にMax MathewsがConducting Programを発表して以来、多くの指揮システムや拍打型のプログラムが開発されてきたが、最近では、楽譜(機械的演奏)ではなく名演奏家の名演の指揮操作を目的としてシステムの開発が進められている[2,3]。

iFPでは、名演奏を拍以上のテンポ・音量表現と拍以下の微細な演奏表現を分離し、それぞれのレイヤにおいて演奏モデルとプレイヤーの演奏意図を合成するという考えに基づいたデザインでインタフェースを構成している[3]。今回、演奏位置やテンポ情報などのユーザ提示を強化したiFP2の展示を行う。

2. iFPシステム概要

2.1 表情テンプレート

iFP で使用する演奏表情テンプレートの記述例を図1に示す。図中太字で示されるものは、譜面上の基本的なデータであり、機械的演奏に相当する。左端の数字は各音符の発音時刻を、括弧で囲まれた範囲はひとつの音符を表し、「音名(ノートネーム)、持続時間」の組として記述される。

細字で示されるものが、演奏表情に関わるデータである。括弧で囲まれた範囲内のデータは、各音符に与えられた拍内表情で、左から順に「発音時刻の deviation, 当該音符の音量, 持続時間の deviation」をあらわす。テンポに関わる情報は、「BPM」が含まれる行として表される。左から順に「時刻, "BPM", テンポ値, 単位となる音符」として記述する。ここで使用するテンプレートは、実演奏の解析から[4]、もしくは、表情付けシ

テムの演奏出力[5]として得られる。同一楽曲の2つの演奏データを用いた演奏表情のモーフィングも可能である。

```

.....
2.00 BPM 126.2 4
2.00 (0.00 E3 78 3.00 -0.11)
=2
1.00 TACTUS 2 4
1.00 BPM 128.1 4
1.00 (0.00 C#4 76 0.75 -0.09) (0.04 E1 60 1.00 -0.13)
1.75 (0.10 D4 77 0.25 -0.14)
2.00 BPM 130.0 4
2.00 (0.00 B3 75 1.00 -0.03) (0.00 G#3 56 1.00 0.03)
3.00 BPM 127.7 4
3.00 (0.00 B3 72 1.00 0.00) (0.09 G#3 56 1.00 -0.12) (0.14 D3 57 1.00 -0.21)
=3
1.00 TACTUS 1 4
1.00 BPM 127.6 4
1.00 (0.00 B3 77 2.00 -0.05) (0.00 G#3 47 2.00 -0.05) (-0.06 D4 57 2.00 -0.32)
.....
    
```

図1 演奏表情テンプレート

2.2 スケジューラ

iFP は、手振りや打鍵といった簡単な操作で演奏表現を行うシステムである。音符の配置(音量やテンポの計算)は、予測型スケジューラによって実施される。予測スケジューラでは、拍打の検出とスケジューリングの処理が独立するため、拍打の省略や倍打ちなどが可能となる。つまり、ユーザは、任意の場所での演奏への介入が可能となる。

iFPでは、テンポ、拍全体でのダイナミクス、拍内の微細なニュアンスのそれぞれに対して、テンプレート、および、ユーザ意図をどの程度反映させるかのパラメータを与えるようにしている。これにより、自身の好みの応答を設定することができる(図2)。

予測スケジューラを利用した場合、拍打ポイントとユーザが与える拍打ポイントがずれてしまうという問題が発生する。iFP では、使用者の拍打が予測より早く検出された際、どの程度、システムが追いついて演奏するかを設定するパラメータも用意している。

3. デモンストレーション内容

iFP2のユーザ提示画面(イメージ)を図3に示す。今回のデモンストレーションでは、操作イン

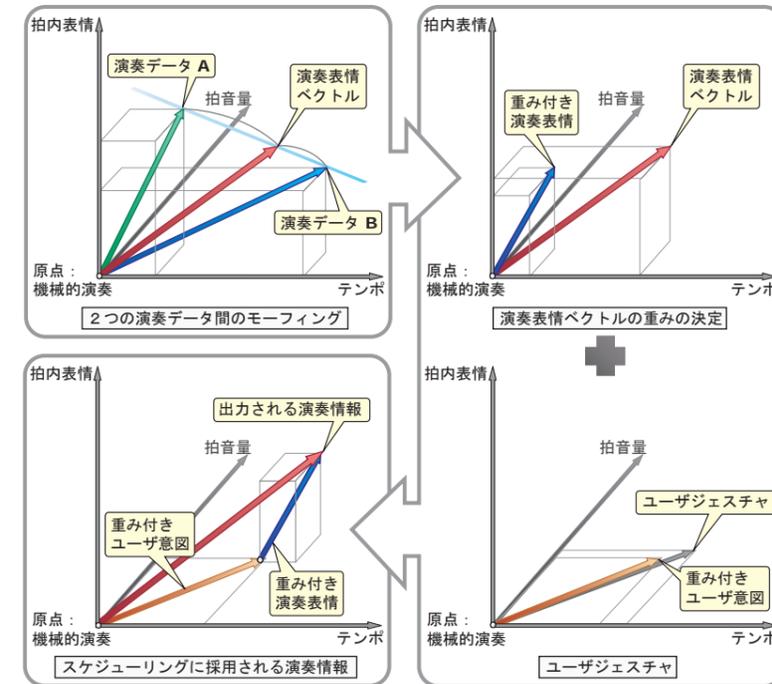


図2 演奏データの計算の様子
演奏表現の特徴をテンポ、拍音量、拍内表情を3次元座標として表現している

タフェースとして、図4に示すような静電容量センサを利用した指揮風のジェスチャインタフェースを用意している。

演奏データとしては、ディズニーアニメの「星に願いを」「美女と野獣」、コルトー版の「別れの曲」、ラグタイムの名曲「エンターテナー」、ショパンワルツ Op.64, No.2 (人間の演奏2編, 表情付けシステムの生成演奏)等、全部で20点のサンプルを用意している。お好みの演奏データを用いた、演奏表現を楽しんでいただきたい。

参考文献

- [1] ICMA Video Review, Vol. 1, 1991 (90 minutes) Review of research in Computers and Music.
- [2] C. Raphael Music Plus One: A System for Flexible and Expressive Musical Accompaniment Proc. Int. Computer Music Conf. Havana, Cuba, 2001.
- [3] 片寄晴弘, 奥平啓太, 橋田光代: 演奏表情テンプレートを利用したピアノ演奏システム:sfp, 情報処理学会論文誌, Vol.44, No.11, pp.2728-2736, (2003.11).
- [4] Ken'ichi Toyoda, Kenzi Noike, and Haruhiro Katayose: Utility System for Constructing Database of Performance Deviations, Proc. ISMIR2004, pp.373-380 (2004.10).
- [5] Hashida, M.; Nagata, N. & Katayose, H. Pop-E: A performance rendering system for the ensemble music that considered group expression. Proceedings of International Conference on Music Perception and Cognition, pp.526-534, (2006.8).

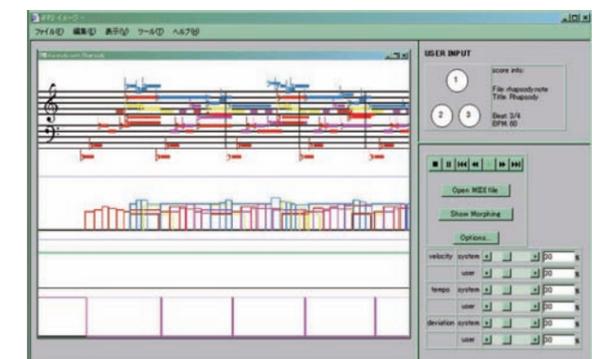


図3 ユーザ提示画面(イメージ)



図4 指揮風のジェスチャインタフェース

音楽を体験している脳をはかる

風井 浩志 藤澤 隆史 森 悠太 山岡 晶 八田原 慎悟
須田 一哉[†] 高橋 理宇真 長田 典子 片寄 晴弘

(片寄グループ：関西学院大学理工学研究科/ヒューマンメディア研究センター)

[†] 東京大学情報理工学研究科

概要 音楽を体験しているときの認知の仕組みを解明するための研究例として、①和音認知の心理物理モデル研究、および和音聴取時の脳活動計測 (fMRI) 研究、②色聴者の音楽聴取時の脳活動計測 (fMRI) 研究、③太鼓演奏時の脳活動計測 (fNIRS) 研究、の3つの研究を紹介する。

キーワード：音楽体験、音楽聴取、音楽演奏、心理物理モデル、fMRI (functional magnetic resonance imaging), fNIRS (functional near infra-red spectroscopy)

1. はじめに

音楽を聴取している時、あるいは自分自身が音楽を演奏している時に、我々はどうのような仕組みでその音楽を認知しているのだろうか。我々のグループでは、脳機能イメージング (fMRI と fNIRS) と心理物理計測とを併用しながら、この問題に取り組んでいる。音楽の認知の仕組みを理解することは、音楽を創るための手がかりのひとつになるだろう。

本稿では、①和音認知に関する心理物理モデルと和音聴取時の脳活動計測 (fMRI) 研究、②色聴者の音楽聴取時の脳活動計測 (fMRI) 研究、③太鼓演奏時の脳活動計測 (fNIRS) 研究、の3つの研究を紹介する。

2.1 和音の認知メカニズム

2.1.1 和音認知の心理物理モデル

長調の音楽は明るく楽しい感じで聴こえ、短調の音楽は暗く悲しい感じで聴こえる。本研究は音楽において重要な構成要素である「和音」に注目し、その音響情報と主観的印象との対応関係を明らかにすることを目的としている。和音的印象を定量的に定義することで、感性情報処理や音楽情報処理への適用が可能となり、また認知科学や脳科学で得られた知見やデータとの比較が容易となる。

本研究のモデルにおいて和音的印象は、①協和性 (澄んだ-濁った)、②緊張性 (緊張した-落ちついた)、③モダリティ (楽しい-悲しい) の3要素から構成されると仮定する。緊張性とモダリティは三和音に特有の知覚現象で、それは「音程の等間隔性」を基盤としており、例えば緊張性の場合、和音構成音の音程が等間隔であるならば緊張的であり、そうでないならば落ちついているとモデル化される。モデルから予測される和音性のそれぞれの値は、経験的にも、心理実験で得られた印象評価データとも、よく符合するものであった [1][2]。

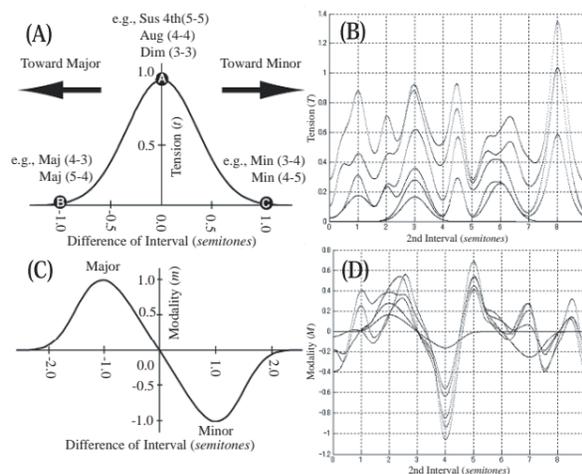


図1 緊張性とモダリティに関する理論曲線 (A), (C)は緊張度およびモダリティそれぞれに関する基本モデルの曲線、(B), (D)は基本モデルに倍音の影響を含めたもの。

2.1.2 和音聴取時の脳活動：fMRI計測

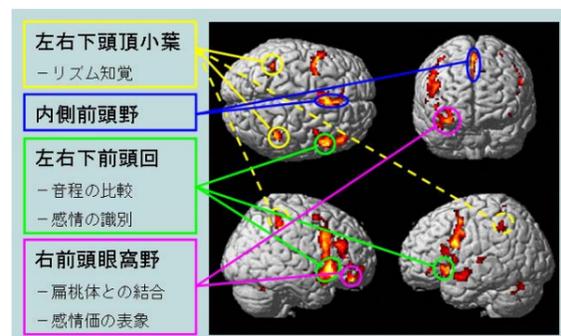


図2 和音聴取時の脳活動

和音性の認知メカニズムを明らかにするために、種々の和音を人間が聴取した場合における脳活動について fMRI を用いて検討してきた。fMRI とは、核磁気共鳴を利用して血中の酸素レベルの変化を測定することにより、脳のどの部分が特異

的に賦活しているかを明らかにする手法である。実験では、様々なタイプの和音を聴取させ、その脳活動部位の特定化を行った。その結果、まず全体的には、左半球よりも右半球において大きな賦活が見られた。また領域別では、リズム知覚の役割を担う下頭頂小葉や、音程比較を行っている下前頭回のほかに、右半球において特異的に前頭眼窩野の賦活が確認された。前頭眼窩野は、情動経験司る扁桃核と結合しており、表情や声の抑揚による感情評価に関連する部位であるとされている。本研究の結果は、和音を聴取することも同様に、人の情動的経験と深い関わりがあることを示している。

2.2 色聴者の音楽聴取時の脳活動

「音を聴くと、色が見える」という現象は、色聴と呼ばれており、共感覚の一種である。こうした現象は、人間の知覚特性や、感覚モダリティ間の関係を明らかにする上で重要な手がかりになると期待される。本研究では、色聴保持者が音楽視聴時において、実際に色知覚に関与している脳内領域で活動が生じているかを、fMRI を用いて計測した。実験では、目を閉じた協力者に対して、様々な調性の音楽や音階を呈示した。分析の結果、色調保持者固有の賦活領域として紡錘状回および上前頭回が抽出された。紡錘状回では色知覚に関わる V4/V8 付近で有意な活動が見られた。このことは、色聴が聴覚系と視覚系の直接的な相互作用により生じていることを示唆している。また左紡錘状回が右に比べて広く賦活していることより、色聴は左紡錘状回が相対的に重要な役割を担っていると考えられる [3]。

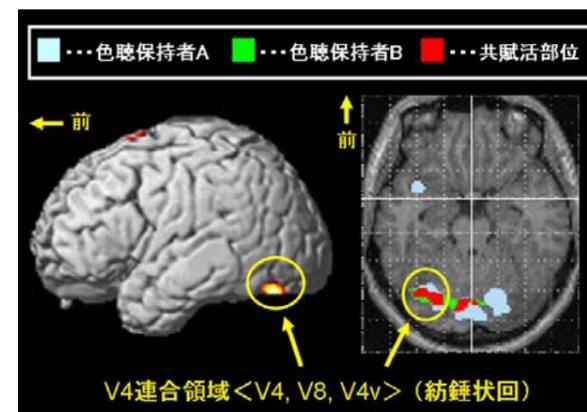


図3 色調保持者の音楽聴取時の活動部位

2.3 太鼓演奏時の脳活動：fNIRS研究

実際に楽器を演奏している時と演奏していない時との前頭前野の脳活動の違いを検討するために、太鼓を演奏している時とドラムパッドを叩いている時の脳活動を比較した。また、演奏するリズムによる脳活動の変化を検討するために、四拍子を演奏している時と三三七拍子を演奏している

時とを比較した。

脳活動計測の手法としては、fNIRSを用いた。fNIRSとは、近赤外光の反射率の変化から血中の酸素レベルの変化を測定することにより、脳の賦活レベルを調べる手法である。fNIRSは、空間分解能ではfMRIに劣るが、計測中の拘束が少ないという利点を持つ。つまり、楽器演奏などの身体運動を伴う状況でも脳活動計測が可能である。

実験の結果として、太鼓を演奏している時の方がドラムパッドを叩いている時よりも脳の活動が高いことが示された。また、太鼓を演奏する場合でも三三七拍子の方が四拍子を演奏しているよりも活動が高いことが示された。太鼓の演奏とドラムパッドの拍打とは身体運動の要因はほぼ等価であると考えられることから、前者の結果が得られた理由として、演奏時の音の大きさ・音の周波数成分の相違が考えられる。後者の結果からは、単調なリズムの演奏よりも複雑なリズムの演奏の方が前頭前野の活動が高まること示唆される [4]。

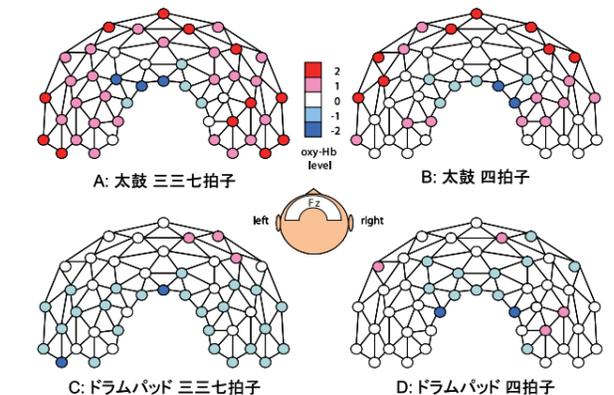


図4 太鼓演奏時における脳の活動部位

参考文献

- [1] Cook, N. D. and Fujisawa, T. X.: The Psychophysics of Harmony Perception: Harmony is a Three-Tone Phenomenon, *Empirical Musicology Review*, vol.1, No.2, pp.106-126 (2006).
- [2] 藤澤隆史, Cook, N. D., 長田典子, 片寄晴弘: 和音認知に関する心理物理モデル, 情報処理学会研究報告 2006-MUS-66, pp.99-104 (2006).
- [3] 高橋理宇真, 藤澤隆史, 長田典子, 杉尾武志, 井口征士: fMRI による共感覚の計測—色聴者の音楽聴取時の脳活動—, 情報処理学会研究報告 2006-MUS-66, pp.105-108 (2006).
- [4] Katayose, H., Nagata, N. and Kazai, K.: An investigation of brain activation while listening to and playing music using fNIRS, *Proceedings of 9th International Conference on Music Perception and Cognition (ICMPC2006)*, pp.107-114(2006).

演奏表情付けシステム Pop-E

橋田 光代 長田 典子 片寄 晴弘

(片寄グループ：関西学院大学理工学研究科/ヒューマンメディア研究センター)

概要 本稿では、複数の声部を伴う音楽演奏に対し自然な演奏表現を実施するための演奏表現モデル Pop-E について紹介する。従来の演奏表情付けシステムにおいては、主旋律に対する演奏表現の質は向上したものの、音楽構造分析の自動化や、合奏やピアノ曲など複数の声部を伴う楽曲を対象とした演奏表現のモデル化が行われていなかった。Pop-E では、音楽構造分析におけるユーザの介入を前提として、声部別に演奏表現(ルール)を付与しつつ、必要に応じて各声部間の発音タイミングを同期させる手法を導入している。当日のセッション会場では、演奏表現の大きさ(パラメータ)をユーザが調整し、オンラインで演奏を生成できる Java インタフェースによるデモンストレーションを行う。

キーワード：演奏表情付け, 演奏デザイン支援, グループ演奏表現, アテンションパート, タメの表現

1. はじめに

演奏表情付け(Performance rendering)は、楽譜を入力として、あたかも人間が弾いているような情緒豊かな演奏を計算機上で実現させることを目指す研究である。傍目には、単に「良い演奏」を追求しているようであるが、実現にあたっては、音楽演奏に主眼を置いた人間の音楽認知過程のモデル化を目指していると言える。

表情付け研究における根本的な課題は演奏表現手法の定式化にあるが、現時点での重要課題としては、(1) グループ構造解析の自動化、(2) 複数旋律音楽に対する自然な演奏の生成、(3) 演奏デザイン支援における効率性・生産性の確保、の三点があげられる。特に(1)については、ひとつの楽曲に対して複数の解釈が存在するため、グループ構造を一意に決定することが困難である。現状では、ユーザによる手作業を交えた上で演奏表情付けシステムの開発が進められている。

一方で、生成された演奏に対する音楽表現能力には顕著な向上がみられるようになった。最近では特に、複数旋律の自然な表現を目指したシステムが開発されるようになり、成果を挙げ始めている[1-3]。しかしながら、従来システムにおける複数旋律の表現は楽曲固有のテンプレートとして利用されるのみで、統合的な演奏表現モデルについては検討されていない。また、ユーザの介入を前提としたシステムでは、演奏の作り込み段階において、音楽構造や制御パラメータを何度も修正し、生成を繰り返す必要がある。

筆者らは、複数旋律音楽のための演奏表現モデル Pop-E[3] を提案した上で、ユーザが嗜好に合わせてその場で音楽構造や演奏パラメータを調整できる演奏生成ツールの実装を進めている。以下、Pop-E の概要と、演奏制御パラメータの調整機能を備えた Pop-E のインタフェース設計について述べる。

2. 複数旋律音楽の演奏表現モデル Pop-E

Pop-E(Poly-phrase Ensemble) は、(1) 複数旋律音楽に対する自然な演奏の生成、(2) 演奏デザイン支

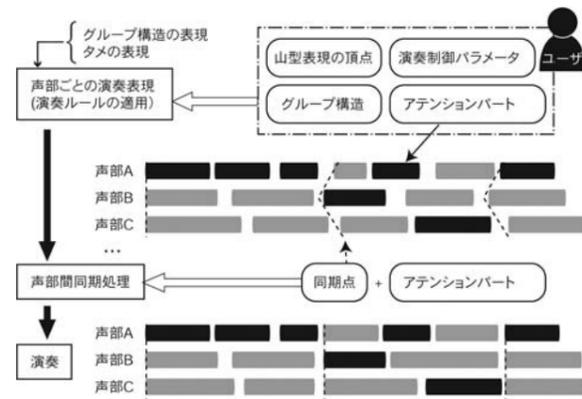


図1 Pop-Eによる複数旋律音楽の生成手順

援における効率性・生産性の確保を目標とした、ルール適用型の複数旋律音楽の演奏表現モデルである。

2.1 概略

図1に演奏生成過程の概略を示す。楽曲の入力情報として、各音の音高や音価の他、あらかじめ各声部のグループ構造、グループ表現のための頂点音、アテンションパート(全声部を通じて楽曲を余すことなく口ずさむことで得られる、ある時刻上もっとも優位な音符列)が与えられている。

複数旋律音楽の演奏においては、各声部は、フレーズ表現として音量およびテンポを個別に増減させる一方で、楽曲として成り立つ程度に声部間の発音時刻を同期させている。そこで Pop-E では、まず声部ごとに固有の制御パラメータを持つ演奏ルールを適用させる。この処理により、声部によって演奏に必要な占有時間がばらつく。次に求められる処理は、必要に応じて声部間を同期させることであるが、このとき、どの時刻でどの声部の占有時間に揃えるかが問題となる。ここでは、楽譜上において声部間のグループの開始音または終了音の発音時刻が一致する箇所を同期点として算出した上で、アテンションパートの占有時間を基準に、残りの声部の占有時間を伸縮させる(図2)。

2.2 声部ごとに適用する演奏表現

Pop-E では、音量(MIDIにおける各音の velocity)とテンポ(同一声部内の隣接した二音の

実演デモ

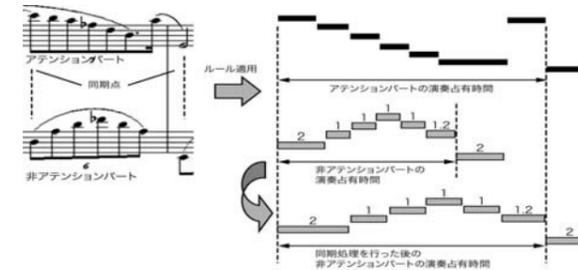


図2 声部間同期処理

発音時刻間の時間長)を演奏の制御対象とし、入力情報の条件がマッチした各音に対して逐次的にルールを適用し、各ルールに与えられた音量とテンポの変化率をそれぞれ乗算することによって、表情のついた演奏を生成する。

Pop-E は、ルールベース型での実装を想定しているが、演奏デザイン支援という観点からは、大規模なルール群のパラメータを操作するよりも、少数に絞り込んだ主要な演奏ルールのパラメータのみを操作の方が望ましい。ここでは以下に述べるグループの表現、タメの表現のみを取り扱い、ユーザが制御すべきパラメータ数の抑制を図る。

グループ構造に対する演奏表現は音楽演奏表現において根幹的なものである。Pop-E では、これまでの演奏システムが採用してきた代表的な表現手法として、(a) グループ先頭音へのアクセント付け、および(b) 山型表現(グループ中のある一音を頂点として、その頂点音が強調されるように、グループ開始音から頂点までに音量と演奏速度を増加させ、頂点を過ぎればそれらを減少させる手法)の二つの表現を取り扱う(図3)。

タメの表現については、該当する音符の音長を伸ばすことについて扱う。対象とする音符は、経験則をもとに、装飾音符や連符音、大きな跳躍音程を伴う隣接した二音の先行音、アテンションパートが他の声部へ移行する際の両端の音符とした。

3. Pop-E に基づく演奏支援インタフェース

演奏制御パラメータの調整を通じた表情付けの作り込みの支援を目的とした演奏生成ツールを実装した(図4)。以下、各機能について述べる。

プリセットされた演奏制御パラメータの読込

音量および音長それぞれにおける初期値に対する比率を演奏制御パラメータとして、6種類16パラメータが声部数分用意されている。これらは外部ファイルで記述され、楽曲や演奏様式ごとに一括して保存できる。ここでは、サンプル楽曲それぞれに対応する標準パラメータセットを用意した。

演奏制御パラメータの個別調整

読み込んだパラメータセットを叩き台として、各パラメータを個別にカスタマイズできる。「演奏生成」ボタンを押すことで、そのパラメータを反映した演奏が生成される。

生成された演奏の再生、ピアノロール表示

パラメータ調整を通じて生成された演奏はその

●フレーズ先頭音に対するアクセント付け ●フレーズの山型表現



図3 フレーズに付与する演奏表現ルール

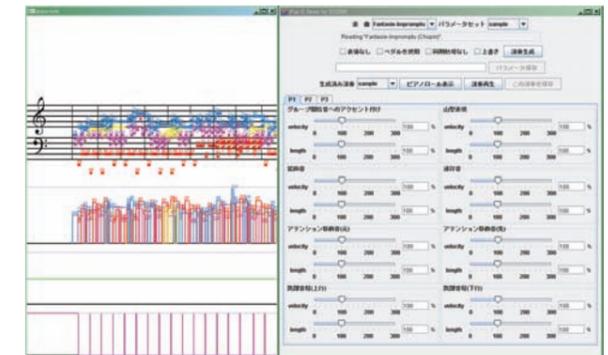


図4 JavaインタフェースのGUI(イメージ)

場で試聴できる。また、各音の音量および音長を示したピアノロール形式の楽譜も表示されるので、ユーザは生成演奏を視覚的に確認しながら繰り返しパラメータ調整と生成を実行できる。その他、生成オプションとして、簡易なペダリングの有無を選択できるようにした。

なお、今回はサンプル曲として、ショパン作曲「幻想即興曲」「別れの曲」、モーツァルト作曲「メヌエット KV.1(1e)」、ラフマニノフ作曲「パガニーニの主題による狂詩曲(第18変奏)」、チャイコフスキー作曲「白鳥の湖」より第2幕の「アダージョ」(いずれも抜粋)の5曲が用意されている。

4. まとめ

本稿では、複数旋律音楽のための演奏表情付け支援インタフェースのプロトタイプを示した。今後は、Pop-Eを用いた演奏の作り込みと評価実験を実施するとともに、音楽構造情報の入力支援として、楽譜から解釈可能な構造の候補を推定、提示する機能を追加していく予定である。今回のデモンストレーションでは、Pop-Eにおけるパラメータ調整機能を用いた複数の演奏生成を行う。今回用意した楽曲のサンプル演奏ならびにPop-Eに関する最新情報は <http://www.m-use.net/research/PopE/> に掲載するので参照されたい。

参考文献

- [1] Raphael, C.: Orchestra in a Box: A System for Real-Time Musical Accompaniment, In IJCAI2003 workshop program, APR-5, pp.5-10 (2003).
- [2] 片寄晴弘, 奥平啓太, 橋田光代: 演奏表情テンプレートを利用したピアノ演奏システム:sfp, 情報処理学会論文誌, Vol.44, No.11, pp.2728-2736, (2003).
- [3] Hashida, M., Nagata, N. and Katayose, H.: Pop-E: a performance rendering system for the ensemble music that considered group expression, Proc. of ICMP2006, pp. 526-534 (2006).



片寄グループ〈関西学院大学〉

片寄 晴弘 教授

1991年大阪大学大学院基礎工学研究科博士課程修了。工学博士。イメージ情報科学研究科、和歌山大学を経て、現在、関西学院大学工学部教授。ヒューマンメディア研究センターセンター長。音楽情報処理、感性情報処理、HCIの研究に従事。科学技術振興機構 さきがけ「協調と制御」領域研究者。CrestMuseプロジェクト代表研究者。ハイテク、ローテク双方のアプローチから「新しい音楽の愉しみ」について取り組んでいきたいと考えています。



藤澤 隆史 博士研究員(私立大学学術研究高度化推進事業)

1975年大阪生まれ。2004年3月関西学院大学大学院総合情報学研究所博士課程修了。博士(情報学)。関西大学ネットワーク研究センター博士研究員を経て、現在、関西学院大学工学部研究科・ヒューマンメディア研究センター博士研究員。主に、感性情報学、認知心理学、および脳イメージングの研究に従事。なぜ人間だけが音楽を楽しめるのかに興味があります。そのために和音認知に関するモデル構築や、心理実験・脳イメージング研究に取り組んできました。



長田 典子 助教授

1983年京都大学理学部数学系卒業。同年、三菱電機(株)に入社。産業システム研究所に在籍しながら、1995年に大阪大学大学院基礎工学研究科博士課程へ進み、翌年に修了。博士(工学)。2003年より関西学院大学工学部情報科学科助教授。メディア工学、感性情報学、脳科学(共感覚)の研究に従事。ヒトが音楽を楽しむメカニズムを明らかにして、感性豊かな音楽創りに役立てていきたいと思っています。



武田 晴登 専門技術員(CREST片寄チーム研究員)

2001年慶應義塾大学工学部卒。2003年東京大学大学院情報理工学系研究科修士課程を修了後、同専攻博士課程に進学。現在 Crest デジタルメディア領域 片寄チーム研究員。東京大学嵯峨山研究室で音楽情報処理の研究を始め、リズム認識の研究に従事。戦略ソフトウェア人材養成コースに参加し、演奏者の演奏に追従する自動伴奏を制作。多くの人が楽しんで世の中の役に立つ音楽情報処理のアプリケーションを作りたいと思います。

風井 浩志 専門技術員(私立大学学術研究高度化推進事業)

1993年関西学院大学文学部心理学科卒業。博士(心理学)。関西学院大学大学院文学研究科博士研究員を経て、現在、関西学院大学大学院理工学研究科専門技術員。実験心理学・精神生理学の立場から視知覚の研究に従事。主として低次な知覚の研究に携わってきましたが、今後はより高次の認知レベルの研究に取り組んでいくつもりです。



後藤グループ〈産業技術総合研究所〉



橋田 光代 博士研究員(私立大学学術研究高度化推進事業)

国立音楽大学音楽デザイン学科、同大学院を経て、2006年和歌山大学システム工学研究科博士後期課程修了。修士(音楽)、博士(工学)。現在、関西学院大学工学部研究科／ヒューマンメディア研究センター博士研究員。SIGGRAPH 98 Television, ICMC'99等にて作品入選。作曲、音楽理論、デジタルコンテンツ制作、音楽情報科学の研究や、楽器演奏、バレエの経験を通じて、音楽表現に至るまでの人間の認知・行動過程に興味を持つ。



後藤 真孝 主任研究員

1998年早稲田大学大学院理工学研究科博士後期課程修了。博士(工学)。同年、電子技術総合研究所(2001年に産業技術総合研究所に改組)に入所し、現在、主任研究員。2000年から2003年まで科学技術振興事業団 さきがけ研究21「情報と知」領域研究員、2005年から筑波大学大学院助教授(連携大学院)を兼任。WISS2000論文賞・発表賞、情報処理学会論文賞等18件受賞。CrestMuseプロジェクトでは、音楽音響信号理解研究とその応用に取り組んでいる。

Elias Pampalk 特別研究員

Elias Pampalk is currently working as a post-doctoral research scientist at AIST within the CrestMuse project on new user interfaces to music collections supervised by Masataka Goto. He completed his PhD studies in March 2006 at the Vienna University of Technology in Austria. His doctoral dissertation deals with computational models of music similarity and their application in music information retrieval.



河原グループ〈和歌山大学〉



河原 英紀 教授

1977年北海道大学大学院工学研究科博士課程修了。工学博士。NTT基礎研究所、ATR人間情報通信研究所等を経て、現在、和歌山大学システム工学部教授。聴覚メディア処理、音声分析合成、音響・音声知覚の研究に従事。科学技術振興機構CREST聴覚脳プロジェクト(1997-2002)研究代表者。人間の歌声の豊かさ美しさを理解し自在に操作できる技術の開発を通じて「新しい歌の楽しみ」を追求して行きたいと考えています。

森勢 将雅 (日本学術振興会特別研究員) 博士後期課程

2006年和歌山大学大学院システム工学研究科修士課程修了。現在は、同大学院博士課程に在籍中。日本学術振興会特別研究員(D C 1)。音響信号処理に関する研究に従事。システムから再生される音を人間がどのように知覚しているかを解明し、ライブ感のある合成歌唱の実現に貢献して行きたいと考えています。



西本 卓也 助手

1995年早稲田大学大学院理工学研究科修士課程修了。京都工芸繊維大学工学部電子情報工学科助手を経て、現在、東京大学大学院情報理工学系研究科助手。パターン情報処理、ヒューマン・マシン・インタフェース、音声対話システム、福祉情報工学、擬人化音声対話エージェントの研究に従事。1999年人工知能学会研究会奨励賞、2003年日本音響学会粟屋潔学術奨励賞。音声対話技術コンソーシアム(ISTC)実行委員として音声対話ツールキットの開発に従事。音楽情報処理から生まれた理論や手法が音声対話や画像処理など他の分野にも影響を与えていくことを目標に、研究に取り組みたいと思います。



酒向 慎司 特任助手

2004年名古屋工業大学大学院電子情報工学専攻博士後期過程修了。現在東京大学大学院情報理工学系研究科システム情報学専攻特任助手。博士(工学)。音声合成の研究に従事し、音楽分野では歌唱音声の合成に取り組んできた。歌唱音声に限らず、音楽の生成系の幅広い問題として、特定の個性やスタイルをもった音楽情報・信号の生成に取り組んでいきたいと思っています。



亀岡 弘和 博士後期課程

2002年東京大学工学部計数工学科卒。2004年東京大学大学院情報理工学系研究科システム情報学専攻修士課程修了。現在、同専攻博士課程在学中。音楽情報処理や音声情報処理を含む音響信号処理の研究に興味をもつ。2005年第20回電気通信普及財団テレコムシステム技術学生賞、IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2005) Best Student Paper Award Finalist、情報処理学会平成17年度山下記念研究賞などを受賞。音楽情報処理という大変魅力的な研究分野を、数理的アプローチを優雅に駆使しながら開拓してゆくことが私の目標であり、夢でもあります。



嵯峨山グループ〈東京大学〉

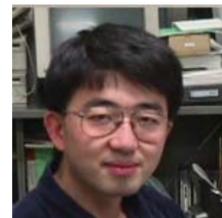


嵯峨山 茂樹 教授

1974年東京大学大学院工学系研究科計数工学専攻修士課程修了。日本電信電話公社武蔵野電気通信研究所、ATR自動翻訳電話研究所、NTTヒューマンインタフェース研究所、北陸先端科学技術大学院大学情報科学研究科を経て、現在、東京大学大学院情報理工学系研究科システム情報学専攻教授。博士(工学)。音声・音楽・音響信号情報処理の研究に従事。1990年発明協会発明賞、1996年科学技術庁長官賞、電子情報通信学会論文賞などを受賞。音声認識技術に長年携わった経験を生かして、数理的な方法論による音楽信号処理と情報処理を極めて行きたいと思っています。

小野 順貴 講師

2001年東京大学大学院工学系研究科計数工学専攻博士課程修了。同大学院情報理工学系研究科システム情報学専攻助手を経て、現在同専攻講師。博士(工学)。音響センサ・音響信号処理の研究に従事。2002年日本音響学会佐藤論文賞、2004年「センサ・マイクロマシンと応用システム」シンポジウム五十嵐賞などを受賞。大学時代の多少の音楽の経験と、これまでの研究につながる数理的な興味から、音楽信号処理の問題に取り組んでいきたいと思っています。



山本 遼 博士後期課程

2004年東京大学工学部計数工学科卒。2006年同大学院情報理工学系研究科システム情報学専攻修士課程修了。現在同博士課程在籍中。パターンの背後にある構造を利用したパターン認識問題に特に興味を持ち、オンライン手書き数式認識、音楽リズム認識等の研究を行う。歌が趣味。音楽を上手に記述できる確率モデルを追求したいと思っています。



米林 裕一郎 博士後期課程

1997年東京大学大学院数理科学研究科数理科学専攻修士課程修了。(株)ジャストシステム、(株)日本システム開発、(株)アスケイドの三つのソフトウェア/システム開発企業を経て、2006年10月東京大学大学院情報理工学系研究科博士課程に進学予定。修士時代に培った数学の素養、企業でのソフトウェア開発経験のほか、幼少から趣味として続けているピアノ演奏の経験を活かした自分らしい音楽情報処理の研究を行いたいと思います。音楽情報の本質を数理モデルを用いてエレガントに解明していくことが私の目標です。



奥乃グループ〈京都大学〉

奥乃博 教授

1972年東京大学教養学部基礎科学科卒業。博士(工学)。NTT基礎研究所, 科学技術振興事業団ERATO北野プロジェクト, 東京理科大学を経て, 2001年より京都大学情報学研究科教授。音環境理解, 人工知能, 音楽情報処理, ロボット聴覚などの研究に従事。人工知能学会論文賞, IEA/AIE-2001, IEA/AIE-2005 Best Paper Award, 第2回船井情報科学振興賞, IEEE & RSJ, Nakamura Award for IROS-2001 Best Paper Nomination Finalist, などを受賞。さまざまな音の混合音から音を「聞き分ける」研究を通じて, 音の認知, 知覚機構を明らかにし, 音のストリームの創発を音楽へ適用していきたいと考えております。



北原 鉄朗 博士後期課程学生

2002年東京理科大学理工学部卒業。2004年京都大学大学院情報学研究科修士課程修了。現在, 同大学大学院博士後期課程在学中。2005年より日本学術振興会特別研究員(DC2)。音楽情報処理, 特に楽器音認識およびその音楽情報検索への応用の研究に従事。電気通信普及財団テレコム技術学生賞, 情報処理学会 第67回全国大会 大会奨励賞などを受賞。「音楽を聴くコンピュータ」の実現を通して音楽理解のからくりを迫ってみたいと考えています。

吉井 和佳 博士後期課程学生

2003年京都大学工学部情報学科卒業。現在, 同大学大学院情報学研究科博士後期課程在学中。2005年より日本学術振興会特別研究員(DC1)。音楽情報検索, 音楽鑑賞支援インタフェースなどに興味を持つ。2004年FIT2004論文賞, 2005年MIREX2005 Best in Class Award, 2006年情報処理学会 山下記念研究賞各受賞。情報処理学会, 電子情報通信学会, IEEE各会員。



片寄グループ〈関西学院大学〉

片寄 晴弘	関西学院大学	工学部	教授
長田 典子	関西学院大学	工学部	助教授
風井 浩志	関西学院大学	理工学研究科	専門技術員(私立大学学術研究高度化推進事業)
橋田 光代	関西学院大学	理工学研究科	博士研究員(私立大学学術研究高度化推進事業)
藤澤 隆史	関西学院大学	理工学研究科	博士研究員(私立大学学術研究高度化推進事業)
武田 晴登	関西学院大学	理工学研究科	専門技術員(CREST片寄チーム研究員)(2006.10から)
池淵 隆		サポートスタッフ	
伊藤 洋介	関西学院大学	理工学研究科	博士前期課程学生
小岩 亮太	関西学院大学	理工学研究科	博士前期課程学生
白井 克明	関西学院大学	理工学研究科	博士前期課程学生
土橋 佑亮	関西学院大学	理工学研究科	博士前期課程学生
酒造 祐介	関西学院大学	理工学研究科	博士前期課程学生
森 悠太	関西学院大学	理工学研究科	博士前期課程学生
横井 真一	関西学院大学	理工学研究科	博士前期課程学生
上田 哲司	関西学院大学	理工学研究科	博士前期課程学生
勝占 真規子	関西学院大学	理工学研究科	博士前期課程学生
高島 杏菜	関西学院大学	理工学研究科	博士前期課程学生
宮本 直幸	関西学院大学	理工学研究科	博士前期課程学生

後藤グループ〈産業技術総合研究所〉

後藤 真孝	産業技術総合研究所	情報技術研究部門	主任研究員
Elias Pampalk	産業技術総合研究所	情報技術研究部門	特別研究員

河原グループ〈和歌山大学〉

河原 英紀	和歌山大学	システム工学部	教授
森勢 将雅	和歌山大学	システム工学研究科	博士後期課程学生
田原 佳代子	和歌山大学	システム工学研究科	博士前期課程学生
畑 宏明	和歌山大学	システム工学研究科	博士前期課程学生
生駒 太一	和歌山大学	システム工学研究科	博士前期課程学生

嵯峨山グループ〈東京大学〉

嵯峨山 茂樹	東京大学	情報理工系研究科	教授
小野 順貴	東京大学	情報理工系研究科	講師
西本 卓也	東京大学	情報理工系研究科	助手
酒向 慎司	東京大学	情報理工系研究科	特任助手
武田 晴登	東京大学	情報理工系研究科	博士後期課程(2006.9まで)
亀岡 弘和	東京大学	情報理工系研究科	博士後期課程
山本 遼	東京大学	情報理工系研究科	博士後期課程
米林 裕一郎	東京大学	情報理工系研究科	博士後期課程(2006.10から)
齊藤 翔一郎	東京大学	情報理工系研究科	博士前期課程
陳 映融	東京大学	情報理工系研究科	博士前期課程
松本 恭輔	東京大学	情報理工系研究科	博士前期課程
宮本 賢一	東京大学	情報理工系研究科	博士前期課程
諸岡 孟	東京大学	情報理工系研究科	博士前期課程
Stanislaw Raczynsky	東京大学	情報理工系研究科	博士前期課程(2006.10から)

奥乃グループ〈京都大学〉

奥乃 博	京都大学	情報学研究科	教授
北原 鉄朗	京都大学	情報学研究科	博士後期課程学生
吉井 和佳	京都大学	情報学研究科	博士後期課程学生
藤原 弘将	京都大学	情報学研究科	博士前期課程学生



戦略的創造研究推進事業
CREST
Core Research for Evolutional Science and Technology

CrestMuse Symposium

発行者・編集者・主催:

独立行政法人 科学技術振興機構CREST(戦略的創造研究推進事業)

「デジタルメディア作品の制作を支援する基盤技術」領域

片寄チーム(**CrestMuse** プロジェクト)

プロジェクトメンバー:

片寄 晴弘・後藤 真孝・河原 英紀・嵯峨山 茂樹・奥乃 博

協力: 関西学院大学

装丁: 橋田 光代・長澤 育子

発行日: 2006年10月26日

連絡先: 〒669-1337 兵庫県三田市学園2-1

Tel & Fax (079)565-7841

E-Mail: nagasawa@ksc.kwansei.ac.jp (長澤 育子)

<http://www.crestmuse.jp/>