

デモンストレーション：音楽情報処理の研究紹介 XIII

平井 重行	京都産業大学 コンピュータ理工学部
池宮 由楽	京都大学大学院 情報学研究科
石先 広海	株式会社 KDDI 研究所
奥村 健太	名古屋工業大学
菅野 幸夫	岡山理科大学 総合情報学部
木下 尚洋	日本大学 文理学部 情報システム解析学科
鈴木 潤一	日本大学 文理学部 情報システム解析学科
鳥本 拓志	京都工芸繊維大学大学院
藤原 和弘	関西大学
前澤 陽	ヤマハ株式会社
山本 和彦	ヤマハ株式会社 / 東京大学

あらまし 本セッションでは、デモンストレーション形式で、音楽情報処理の研究を紹介する。

デモセッション実施に寄せて

平井 重行

本デモンストレーションは、音楽情報処理の研究分野における様々な研究システムを研究会参加者が体験するための発表セッションとして毎年行われている。2004年に始めた当初は若手研究者の研究の発展に向けたものであったが、実際は特に若手に限らず「自称若手」や「永遠の若手」も含む多彩な年代、顔ぶれによるデモンストレーションが行われている。また、デモ発表会場に空きスペースがある場合は、飛び入りデモも次々と行われており、研究報告資料の原稿やポスターなどの説明資料もない状態で突然デモを実施することも可能なカジュアルな様相でもある（図1参照）。

さて、このセッションでは、新しい楽器や楽器インタフェースの研究など実際に触れたり演奏しないとわからないものから、従来楽器を用いて演奏を伴うもの、開発中のもの（未完成のもの含む）、以前に発表したものがそれ以降デモンストレーションの機会に恵まれていないシステム、デモセッションと同回に一般発表されるもの、などがデモンストレーションされる。果ては、飛び入りデモとして、一般発表では発表しにくい思いつきで作ってしまったものや、お蔵入りになった研究システムまでが披露されることもある。

このようなデモセッションを実施してきたなか、音楽情報科学研究会では、研究分野の社会への認知度向上や

研究内容の公表を目指し、2012年度から各研究発表会の発表の様子を Ustream とニコニコ生放送を通じてインターネット動画中継を始めた。ただし、このデモセッションに関しては、発表者が短時間でデモ内容を紹介するコーナー以外は動画中継されないため、実際に発表会場へ足を運んだ参加者のみが体験・議論できる。このような体験型の研究が大事なものは、時系列メディアとしての重要性が非常に高く、楽器演奏などインタラクティブな要素が多分にある研究テーマが数多く存在するからと言える。今後も継続してデモンストレーションの場を作り、各研究の活発な議論や新たなアイデアが生まれる機会を増やすと共に、音楽情報処理研究分野の拡大や活性化に繋げていきたい。



図 1: MUS96(2012 年) でのデモ会場の様子

混合音中の歌声 F0 軌跡に対する 歌唱表現転写システム

池宮 由楽, 糸山 克寿, 吉井 和佳, 奥乃 博

市販楽曲に含まれる歌声の基本周波数 (F0) 軌跡に対してビブラートやこぶしなどの歌唱表現を転写することを可能とするシステム (図 1) を提案する。歌声の F0 軌跡の変動は、歌い回しの知覚に大きく影響する。これを自由に加工することができれば、ユーザはプロ歌手の歌い方を自分の好みによって鑑賞を楽しむことができる。

(1) 歌唱表現データベース

事前にプロ歌手の歌唱表現をデータベース化しておくことで、ユーザはそのデータベースから参照しながら転写を行うことができる。歌唱表現データベースは、市販楽曲から 3 種の歌唱表現 (ビブラート・こぶし・グリッサンド) を抽出し蓄積することで構築する。

(2) 歌声 F0 軌跡に対する歌唱表現転写

混合音中から歌声の F0 のみをピッチシフトする必要がある。そこで、入力信号を時間周波数領域へ変換し、歌声のスペクトルのみを変化させる。ユーザの入力を援用し正確な F0 軌跡を得た後、歌声のスペクトルを選別して対数周波数軸でシフトすることでピッチを変える。この際、音韻性や声質を保つため、音色の補正を行う。位相を復元し、時間領域へ再構成することで、歌唱表現の転写された音響信号を得る。

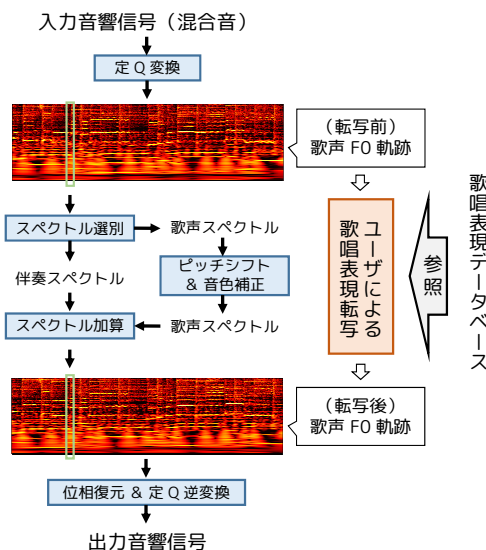


図 1: システム概要図

参考文献

- [1] 池宮 由楽, 糸山 克寿, 吉井 和佳, 奥乃 博: “混合音中の歌声 F0 軌跡に対する歌唱表現転写システム”, 音楽情報科学研究会 (SIGMUS 104), 2014.
- [2] Yukara Ikemiya, Katsutoshi Itoyama, Hiroshi G. Okuno: “Transcribing Vocal Expression from Polyphonic Music”, Proc. ICASSP, 2014.

シーン理解に基づくウェアラブルデバイス連携 BGM プレイヤーの試作

石先 広海, 服部 元, 滝嶋 康弘

近年のスマートフォンやウェアラブルデバイスの普及に伴い、ユーザが日常的に写真を撮影し、PC やクラウド上に大量の写真やライフログを蓄積することが可能となっている。蓄積した写真やライフログをもとに、ユーザは旅行などのイベントにおける体験を振り返ることが出来る。本研究では、このような振り返り体験をリッチ化し、写真やライフログを閲覧するだけでなく、体験時の感情や雰囲気までも再体験できる体験再現メディアの創造を目標としている。本稿では、ユーザの体験をリッチ化するための一要素として音楽に着目し、イベント体験時に場を彩る音楽を自動選択・再生し、写真とともに記録するシステムについて検討する。

(1) 試作システム

試作システム (図 1) は、ユーザがイベント時にスマートグラスなどのウェアラブルデバイスを装着していることを前提とし、まず事前準備としてシステムで利用する全ての楽曲に対して、歌詞から全体印象 [1] を推定し、メタ情報として楽曲 DB に格納しておく。システム利用時には、スマートグラスを用いて定期的に状況を自動撮影する。自動撮影した画像に対し、オブジェクト検出による自動タグ付与処理 [2] を適用し、付与した画像タグに基づいてユーザの状況を判定する。状況判定方法は、画像タグと楽曲に付与した全体印象との変換ルールを設定することで行う。最終的に、画像と楽曲間の全体印象のマッチングにより、ユーザの状況に応じた楽曲を選定・再生する。

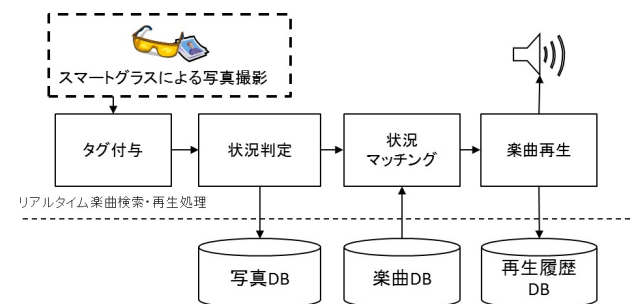


図 1: システム概要図

参考文献

- [1] 石先広海, 他: “歌詞情報に基づく Web 画像検索を利用した歌詞連動スライドショー生成システム”, 情報処理学会論文誌 Vol.54-4, pp. 1263-1274, 2013.
- [2] 石先広海, 他: “二言語間のタグ変換を用いた画像タグ付与システム”, FIT2010 RI-009.

カラオケのための Wii タンバリンの試作

木下 尚洋, 栗原 拓也, 山口 竜之介, 北原 鉄朗

カラオケが娯楽として根付いている今日, 複数人でカラオケを訪れた際に歌手以外の人々が一体感がなく楽しめない事が少なからずあると思われる. 本研究ではカラオケ店で一般的に置いてある楽器, タンバリンを用いてあらゆる人がカラオケを盛り上げられる事を目的とする. 曲のリズムに合わせてタンバリンを叩くタイミングを自動生成, 表示 (以下, これをタンバリン譜と定義する) し, リズムを取ることが苦手な人にもタンバリンを叩きつけかけを与える.

(1) システムに用いる器具

演奏の入力には一般的に MIDI ギターなどの MIDI 楽器が用いられるが, タンバリンには MIDI 楽器がない. そこで演奏の入力を Wii リモコン [1] を装着したタンバリンを用いる事によって行う.

(2) タンバリン譜の生成

タンバリン譜の生成は MIDI の中で最もリズムを取る事に適しているスネアドラムパートを用いる. だが, これだけでは場を盛り上げる為に重要な部分への対応が不十分である. そのため, 以下の条件に合わせて生成方法を変える.

1. 曲中のサビ前に多く見られる印象に残りやすいキメの部分の生成
2. スネアドラムパートが無音でそれ以外のパートが演奏されている部分に頭打ちの生成

また, タンバリン譜では「叩く」と「振る」の動作がある. タンバリン譜内で連続で三音以上のものがあつた場合は「振る」, それ以外は「叩く」と定義し演奏する.

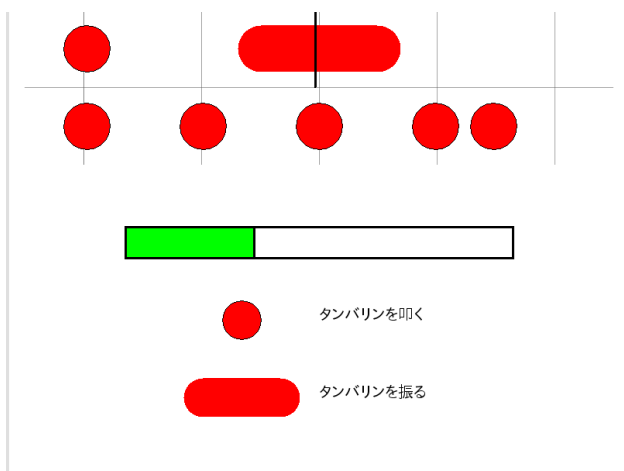


図 1: タンバリン譜

参考文献

- [1] <http://www.nintendo.co.jp/wii/controllers/>

Bluetooth を用いてグループ化されたミュージックプレイヤー

鈴木 潤一, 末次 尚之, 北原 鉄朗

未知の楽曲との出会いを複数ユーザー間で実際に聴かせ合うというアプローチから提供することを目的としたシステムを紹介する. 新規楽曲の発見を目的とした研究 [1] は成されてきたが, ユーザー間で楽曲を聴かせ合うことに焦点を当てた議論は十分にされていなかった. 我々はこの「実際に耳にするという行為」に着目し, ユーザーの持つ楽曲を自動選曲し, 同一スピーカーから再生することでユーザー間で楽曲を聴かせ合うことができるシステムを構築した.

(1) システムの概念図

図 1 に本システムの概念図を示す. デバイスを親機, 子機に割り振り. Bluetooth を用いてネットワークを構築する. これを以後グループ化と呼ぶものとし, 楽曲情報の共有や再生信号の送受信などを行う.

(2) 楽曲再生の流れ

楽曲の再生は, 次の流れで行われる.

1. 複数のデバイスを用意し, 1 台を親機, それ以外を子機として Bluetooth を用いてグループ化を行う.
2. 親機は子機あるいは自身に対し, 楽曲の再生信号を送る.
3. 再生信号を受けたデバイスはスピーカーとの通信を開始し, 楽曲の再生を実現する.

(3) 今後の課題

現段階のシステムではグループ化によってデバイスとの再生信号と楽曲信号のやりとりを行うのみで, 選曲の考慮はしていない. 楽曲情報から「複数ユーザーの持つ楽曲から 1 つのプレイリストを生成する」といった機能を拡張できる余地があるといえる.

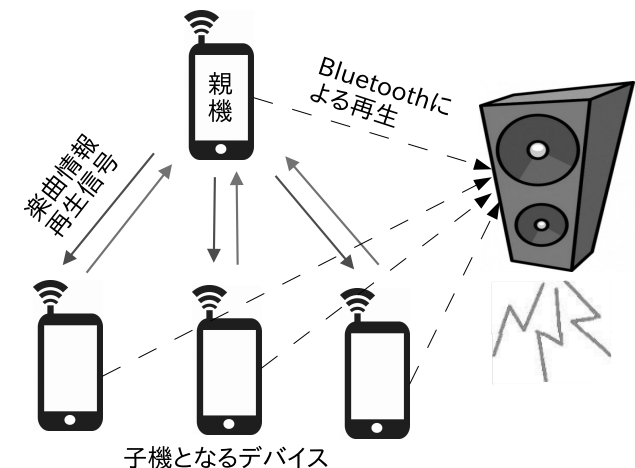


図 1: システムの概念図

参考文献

- [1] Arthur Flexer, Dominik Schnitzer, Martin Gassar, Garhard Widmer: Playlist Generation Using Start and End Songs, ISMIR, pp.173-178, 2008.

演奏できる掃除用具 - SOJIC

鳥本拓志, 浅野絵里, 植木淳也, 小島惇, 小中貴晃,
津田井美香

(1) 主軸

音楽の演奏体験は習熟を必要とする。日常で耳にする音楽の殆どはプロの演奏によって成り立つものであり、セッションやジャム、即興演奏で得られるような演奏体験の快感は多くの人に共有し得ず、そのハードルはステージのように高く難しい。その快感を素人がいかにして得られるか、またはそのハードルを如何に低くできるかということが本研究の軸である。

(2) 機能及び設計

簡単な日常的所作と演奏体験を結びつけるために分析調査を行なった結果、料理・掃除・ランニングなど身体の反復行動が一定のリズムを刻んでおり、弾く・叩くといった演奏の行為との共通することが分かった。殊に掃除は音楽との結び付きが強く、音楽を聴きながら掃除をしたり極端な例では箒をギターに見立て遊ぶケースから、掃除と演奏行為を繋げることを考えた。

以上から考案したのが演奏掃除デバイス「SOJIC」となる。その内の一つは箒の形状をとっている。上部ダイアルで演奏ジャンルを選択、スタートボタンを押すとカウントインが始まる。地面と接触する先端部にセンサーを内蔵し、演奏ボタンを押しながら履くことで擬似的に高度な演奏体験を得ることができ、演奏ボタンの押し変えて流れるフレーズが変わることで音楽を盛り上げる。他シリーズも同様の構造を持ち、メロディやリズムまたはベースなどの違ったロールとの掛け合いで、複数機間でのセッションが可能。ズレが生じた場合はリセットボタンで中断できる。

SOJIC の導入には日常の退屈な部分が解消されることだけではなく、娯楽化や能動化が期待できる。また、演奏能力のない人間に演奏の快感を与え、単調な日々を楽しく演出することが可能となる。



図 1: SOJIC の全貌

遷移確率を用いた自動作曲

藤原 和弘, 高橋 智一, 鈴木 昌人, 青柳 誠司

アーティストの楽曲から抽出した音高と音長の遷移確率を使用した作曲方法を紹介する。遷移確率とは、次にとどの音へ進行するかを確率としたものである。様々な楽曲から遷移確率を抽出し蓄積することで、人間の作った曲に近づけることを目指している。

(1) 遷移確率の抽出

遷移確率とは、ある音高（音長）の次に、どのような音高（音長）が続くかの確率である。ハ長調に移調した楽曲から音高、音長それぞれの遷移を抽出し、2音毎の関係を調べ、遷移確率を導出する。図1に本稿で実際に用いた遷移確率の一部を示す。

(2) 遷移確率を使用した作曲

前項で導出した遷移確率を使用し、実際に作曲を行う。第一音の音高は、ユーザーによって選択できるようになっている。選択された音高によって、作成される曲の調も決定される。第一音の音長は、楽曲の先頭の音から確率で選択される。第二音以降の音高、音長は遷移確率により決定される。

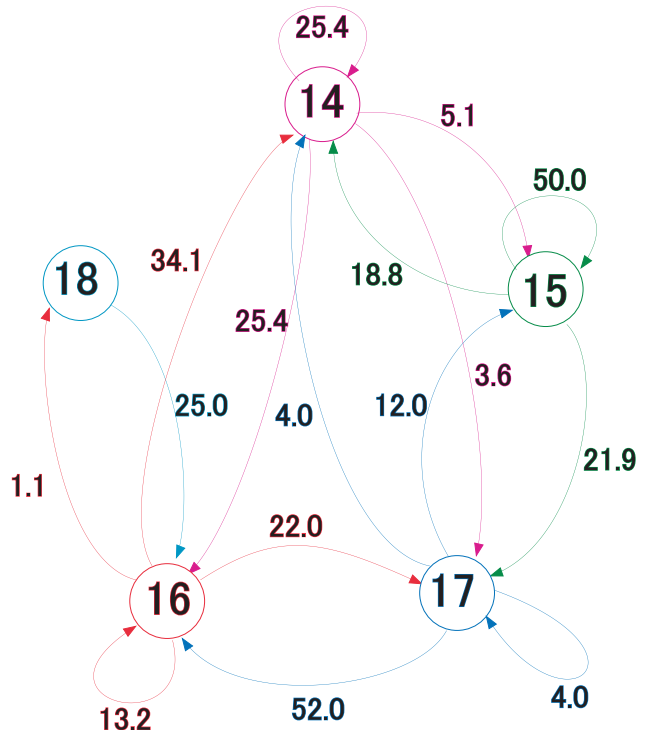


図 1: 音高の遷移確率（一部）

スコア追従に基づく 演奏-映像インタラクション

前澤 陽, 中村 吉就

音楽を演奏者と視聴者が楽しむ上で、演奏と、映像シーケンス等の演出は重要な要素である。演出は、ある楽曲に対して緻密に作りこむことで、より効果が発揮されると考えられる一方で、演出用コンテンツを作り込み過ぎると、演奏者がそのコンテンツに合わせて演奏しなければいけなくなり、演奏者が演奏体験を楽しめなくなる恐れがある。

(1) 演奏・映像同期システム

我々は演奏者が事前に指定した楽曲を演奏すると、それに同期して演出用の映像コンテンツを再生するようなシステムを提案する。このとき、演奏のテンポや、想定される演奏からの逸脱度合いにより、再生される動画が切り替わる。これにより、演奏者はコンテンツの制約を気にすることなく、気兼ねなく楽器を演奏できる。更に、演奏のニュアンスに応じて映像を切り替えることにより、視聴者にとっては、その演奏でしか味わえない演出を体験することができ、特定の演奏に対する付加価値が高まると期待される。

http://www.y2lab.com/project/score_alignment/に詳細を記載してある。

(2) システム構成

システムは、図1に示すように、スコア追従モジュールと、スコア追従モジュールの出力結果から映像を再生するシステムで構成されている。映像再生は、スコアの位置に応じた映像を再生する。これに加え、演奏されているテンポや、スコア追従の「信頼度」を入力することで、演奏が「速い」「想定外である」といった具合で変化した場合に、適切な映像コンテンツに切り替えることを特長としている。

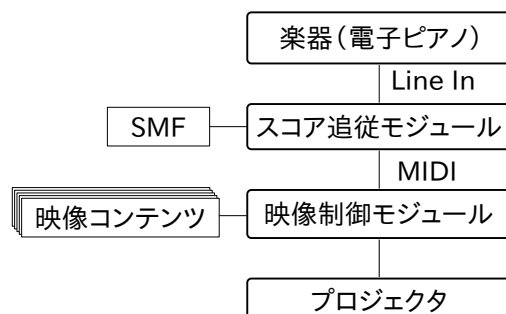


図1: 本システムの構成。

謝辞 デモコンテンツの利用を許諾して下さった多数の関係者様に感謝する。

LiVo: 母音列への歌詞アライメントによる リアルタイム歌唱合成演奏のための 歌詞操作インタフェース

山本 和彦

歌詞と旋律を同時に入力することによって、リアルタイムに歌唱合成を用いた演奏表現を可能にするための実用的なインタフェースを紹介する。こうした試みは過去にも行われてきた [1][2] が、一般的な楽曲を演奏するために十分な入力速度を達成するには至らなかった。そこで本研究では、ユーザによって連続的に入力された母音列に歌詞をアライメントして出力することでこの問題を解決する。

(1) システム概要

図1に、システムの概要を示す。本システムでは、ユーザはまず演奏前に楽曲の歌詞を入力し、システムはそれを分析して事前データとして記憶しておく。演奏時にはユーザは、片手で母音列を入力し、もう片方の手で音高を入力する。システムは入力された母音列から歌詞のどの位置を演奏しているのかを推定し、その歌詞を出力する。最適な歌詞位置を推定するため、任意の歌詞位置移動や入力ミスに対しても堅牢に動作する。日本語では母音の数は 高々5, 6 個であり、片手の各指とほぼ一対一に対応させることができる。また、母音キーは鍵盤に割り当てられているため、五線譜で練習することもできる。これにより入力速度と表現の自由度を兼ね揃えた実用的なインタフェースを実現した。

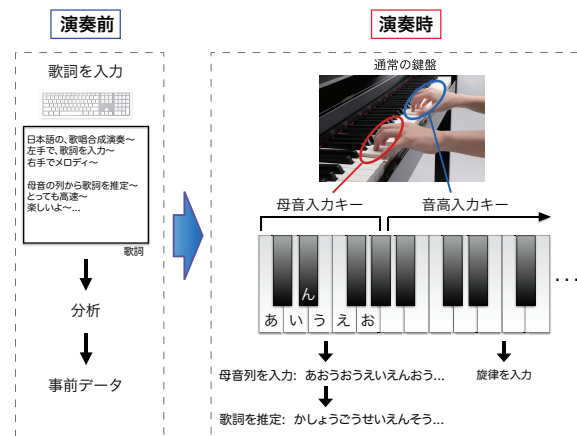


図1: LiVo システム概要

参考文献

- [1] 山本和彦, 加々見翔太, 濱野桂三, 柏瀬一輝: リアルタイム日本語歌唱鍵盤楽器のための文字入力インタフェースの開発, 情報処理学会論文誌, Vol.54 No.4, p1373-1382 (2013).
- [2] 佐近田 展康: 「兄弟式リアルタイム音声合成演奏システム」の概要と背景, 名古屋学芸大学メディア造形学部研究紀要 (2013).