2008－MUS－74 (11)
2008－SLP－70 (11)
2008／2／8

デモンストレーション: 音楽・音声言語情報処理の研究紹介

西村 竜一　和歌山大学システム工学部
伊藤 丈一　北陸先端科学技術大学院大学知識科学研究科
内村 佳典　名城大学理工学研究科
川添 正人　名城大学理工学研究科
剣持 秀紀　ヤマハ株式会社 サウンドテクノロジー開発センター
浜中 雅俊　筑波大学, 科学技術振興機構 さきがけ
宮本 賢一　東京大学大学院情報理工学系研究科
梅本 暁　早稲田大学理工学研究科
森勢 将雅　和歌山大学大学院システム工学研究科
中野 倫靖　筑波大学大学院図書館情報メディア研究科
大石 康智　名古屋大学情報科学研究科
高橋 量衛　名古屋大学情報科学研究科
野池 賢二　株式会社トランス・ニュー・テクノロジー
戸田 智基　奈良先端科学技術大学院大学 情報科学研究科
梶 克彦　NTT コミュニケーション科学基礎研究所

あらまし 本デモセッションでは，音楽・音声言語情報処理に関する研究分野の発展に向けて，研究事例をデモンストレーション形式で紹介する．

Demonstrations

Ryuichi Nisimura　Faculty of Systems Engineering, Wakayama University
Joichi Ito　School of Knowledge Science, Japan Adnvanced Institute of Science and Technology
Yoshinori Uchimura　Graduate School of Science and Engineering, Meijo University
Masato Kawazoe　Graduate School of Science and Engineering, Meijo University
Hideki Kenmochi　Center for Advanced Sound Technologies, Yamaha Corporation
Masatoshi Hamanaka　University of Tsukuba / Presto, Japan Science and Technology Agency
Kenichi Miyamoto　Gradualte School of Information Science and Technology, The University of Tokyo
UMEMOTO Akira　Gradualte School of Science and Engineering, Waseda University
Masanori Morise　Graduate School of Systems Engineering, Wakayama University
Tomoyasu Nakano　Graduate School of Library, Information and Media Studies, University of Tsukuba
Yasunori Ohishi　Graduate School of Information Science, Nagoya University
Ryoei Takahashi　Graduate School of Information Science, Nagoya University
Kenzi NOIKE　Trans New Technology, Inc.
Tomoki Toda　Graduate School of Information Science, Nara Institute of Science and Technology
Katsuhiko Kaji　NTT Communication Science Laboratories

Abstract Toward further progresses of researches in the field of music information processing and spoken language processing, we introduce case studies of demonstrations.

はじめに

西村 竜一 (和歌山大)

2007年を振り返ってみると，ソフトウェアの分野で最も大きな話題と言えば，バーチャル・シンガー「初音ミク」の登場だったように思う．発売とともに，インターネットコミュニティを中心に瞬く間に人気を集め，社会に大きなインパクトを与えた．音テクノロジの研究開発では，皆が，新たなブレイクスルーを目指し，日々奮闘努力をしている．「初音ミク」は，音声合成の応用が成し遂げた大きな成功例として，若手研究者を中心に，音の研究・開発に携わる多くの人々に希望を与えてくれた．同時に，研究開発の成果のアピールを工夫することが大切であると考えさせられる出来事であった．

今回，音楽情報科学研究会（SIG-MUS）と音声言語処理研究会（SIG-SLP）の合同研究会では，音楽と音声分野の双方の研究者からデモを広く募った．その結果，14件（内4件は一般とデモの両方で発表）もの多数の申込みがあった．まずは，発表者の皆様に感謝を述べたい．

いずれの発表も非常に興味深く，次の成功を予感させるような内容になっている．特に，「歌声」をテーマにした発表が多いのが印象的である．歌声は，音楽・音声分野のどちらの研究者にとっても大変興味深いテーマであると言って良いだろう．双方の分野の研究者が集まることで，有意義な議論が行われることと思う．

また，デモセッションでは，一般の口頭発表では伝えきれないシステムの面白さや，システム開発に込められた裏側を知ることができる．これからシステムの開発を考えている人や，まだ発展途上なシステムの開発者の方々には，システムを世の中に出していくのにはどのようなアプローチが良いのかを考える機会にして欲しいと思う．今回のデモセッションでは，まだブラッシュアップの段階のものから，既にWeb上で発表されたり，ビジネスに直結しているシステムまで幅広いラインアップがなされている．先人たちの工夫や経験を知ることで，冒頭で述べたような，研究開発の成果を世界にアピールする方法として有益な知見が得られると思う．このため，今回，デモセッションを前半と後半に分けて行うことにした．限られた時間内ではあるが，発表者が，自身の展示と関連の深いデモを横目でしか見られないという状況は回避したつもりである．密度の濃い意見交換にご協力いただき，今後の活動に役立てていただきたい．

最後に，今回のデモセッションにより，MUS研・SLP研の双方の研究者が互いを意識し，分野横断的な研究がますます促進されるようになれば幸いに思う．

以下に，発表概要を掲載する．紙面の都合上，表紙ページにおける代表発表者以外のお名前はやむなく省略させていただき，各半ページの原稿の中での連名にしていただくことにした．指導者や共同研究者の皆様には大変恐縮であるが，何卒ご了承いただきたい．

音楽的特徴量と作曲者の主観評価の関連性を用いたフレーズ作成支援システムの構築

伊藤丈一，伊藤直樹，西本一志 (北陸先端大)

デモセッションで展示するシステム:mu-cept は、試行錯誤的にフレーズを作成している段階で，”自分の嗜好を満たし、かつテーマに沿ったフレーズを作ることができない”という状況からユーザを脱却させることを目的としている．ユーザは midi リアルタイム入力でフレーズを mu-cept に入力し，フレーズの入力直後にフレーズに対する好みの度合いとテーマとの合致度についての主観評価を入力する．このプロセスを繰り返してフレーズ作成を進めていく．mu-cept 内部では，ユーザが入力した主観評価を目的変数，各フレーズから自動抽出された音楽的特徴量を説明変数とし重回帰分析を行っている．ユーザは行き詰ったとき，画面上の help!ボタンをクリックし，mu-cept に助けを求める．すると，重回帰分析で得られた，好みの度合いとテーマとの合致度の変化に大きく影響する音楽的特徴量に基づき，ユーザにとって主観評価の高いフレーズを作成するために有用と思われるアドバイスが，例えば”音数を減らした方がよい”といった形で提示される．ユーザは mu-cept から提示された情報を行き詰まり脱却の示唆として活用し，フレーズ作成を継続する．

声質制御への応用を目的とした、声道断面積関数の分析

内村 佳典，坂野 秀樹，板倉 文忠 (名城大)

声道断面積関数による声質制御方式について研究を行っている．ここでいう声道断面積関数とはKellyの音声生成モデルに基づくもので，声道を断面積一定の微小音響管の従属接続で表現したものである．これは，音声の特徴量として知られるPARCOR係数から求めることができる．提案法では，ある音声から求めた声道断面積関数の値を変化させることで声質を制御する．PARCOR分析では，音源と放射の特性を声道の特性に含めて考えているため，声道断面積関数を求める際には，あらかじめ音声波形から音源と放射の特性を取り除いておく必要がある．今回は，その方法として，1978年に中島らにより提案された適応逆フィルタ法を用いた．他の方法では，声質制御後の合成音声が不自然になることがあったが，適応逆フィルタ法を用いることでこれが改善された．デモセッションでは，ある音声波形から求めた声道断面積関数の値を区間ごとに定数倍することにより声質を制御するシステムを紹介する．

テンポの変化による影響を考慮した歌唱音声合成に関する検討

川添 正人，坂野 秀樹，板倉 文忠 (名城大)

既存の歌声合成システムでは速く歌わせた部分において，音声を切り貼りしたように聞こえる，あるいは，実際より滑舌が良すぎるように聞こえるなど，子音部において不自然に聞こえる場合がある．このような問題を解決すべく，楽曲のテンポの変化による歌声への影響を考慮した歌声合成手法について検討している．実際の歌唱音声をスペクトル分析した結果，楽曲のテンポが速くなるにつれ，スペクトルの時間変動が小さくなることが観測された．これは，声道の形状が急に変化できないために，調音が不完全になることに起因すると考えられる．既存の歌声合成手法による合成音では，このことがあまり考慮されておらず，これが不自然に聞こえる原因の1つとなっている．そのため，提案手法では，既存の手法による音声に対し線形予測分析を行って声道断面積関数を推定し，楽曲のテンポが速い場合には，この声道断面積関数を時間軸に関して平滑化し，音声を再合成する．これにより，楽曲のテンポによるスペクトル変動量の変化を再現している．デモンストレーションでは既存の手法による歌唱音声と，提案手法による歌唱音声を提示する．

歌声合成システム VOCALOID

剣持 秀紀，大下 隼人 (ヤマハ)

Vocaloid は、ヤマハが開発した素片連結型歌声合成技術およびその応用商品の総称である。Vocaloid の第2バージョンである Vocaloid2 を応用した「初音ミク」は音楽制作向けソフトウェアとしては異例の売上を記録した。Vocaoid は、歌声に特化した専用の GUI(スコアエディタ)、実際の歌手から取り出した歌声の音声素片を含む歌手ライブラリ、そして合成エンジンから構成される。合成エンジンは、ユーザが指定した音符に合うように、素片を周波数軸上でピッチ変換、音色のあわせ込みを行い、楽譜に合うようにタイミングを調整して連結し出力する。デモセッションでは、Vocaloid の入力画面で歌詞と音符を入力して歌声を合成し、簡単な楽曲制作のデモンストレーションを行う。さらに、Vocaloid2 の「リアルタイム演奏機能」、すなわちキーボードを用いて「歌う」機能もデモンストレーションする。また、合成エンジンについて、歌声に特化した場合に必要だった工夫を中心に内部処理の簡単な説明を行う。

電気モーフ

浜中 雅俊 (筑波大学，科学技術振興機構 さきがけ)

電気モーフは，2つのメロディの中間的なメロディを生成するメロディモーフィング手法 [1] を用いたデモンストレーションである．見た目は DJ ミキサーのようであるが，使用しているのは，マスタースライドボリュームのみである．DJ ミキサーでは，マスタースライドボリュームによって，2つの音源の音量比が変化するが，「電気モーフ」では，メロディ自体が変化する (図 1)．

(1) 電気モーフの全体像

電気モーフは，2小節のメロディをループ再生する．そして，ボリュームのスライダーが一番左にあるときはメロディAを，一番右にあるときはメロディBを再生する．また，スライダーを左右に動かすと，メロディモーフィング手法によって生成された A と B の中間のメロディをループ再生する．具体的には，文献 [1] の手法により複数生成された A と B の中間のメロディを用いて，スライダーを左から右に移動させるに従い，メロディAの特徴が減少し，メロディBの特徴が増加したメロディが再生されるようにした．ユーザはスライダーを左右に動かすことで，モーフィングによるメロディの変化を楽しむことができる．

(2) メロディモーフィング手法の改良

文献 [1] のモーフィング手法は，メロディAの特徴を反映させる度合いを決めるパラメータと，メロディBの特徴を反映させる度合いを決めるパラメータの2つを独立に設定するものであった．それには，多くのメロディが出力できるメリットがある反面，操作性が悪いというデメリットがあった．電気モーフでは，A の値の増加に反比例して B の値が小さくなるようにすることで，1つのスライダーで，モーフィングできるようにするした．

また，文献 [1] の手法では，パラメータが同じ設定でも，join の計算の結果，複数のメロディが生成される．そこで，電気モーフでは，生成された複数のメロディを楽曲分析器 FATTA[2] を用いて分析し，FATTA によって最も安定度が高いと判定されたメロディを出力する．

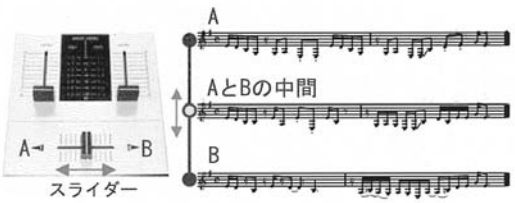

図 1: 電気モーフ

参考文献

[1] 浜中 雅俊，平田 圭二，東条 敏：タイムスパン木に基づくメロディモーフィング，情報処理学会研究報告 2008-MUS-74, 2008.

[2] Matstoshi Hamanaka, Keiji Hirata, and Satoshi Tojo. FATTA: Full Automatic Time-span Tree Analyzer, Proceedings of ICMC2007, Vol.1, pp.153–156, 2007.

リアルタイム調波音・打楽器音分離システム

宮本 賢一, 亀岡 弘和, 小野 順貴, 嵯峨山 茂樹 (東大)

近年、多重ピッチ推定など様々な音楽信号分析技術が開発されてきているが、ポピュラー音楽のように調波音と打楽器音が混合した信号においては、こうした分析が難しいと考えられる。この問題に対し我々は、楽器や楽譜の固有情報を用いずに、単一チャネル音楽信号から調波音的な成分と打楽器的な成分を分離する手法を開発した。この手法は、音楽信号処理における前処理や音楽加工など、多くの応用が期待される。本発表では、実時間で調波音・打楽器音分離を行い、それらを任意のバランスで合成して再生するシステムを紹介する。

(1) 調波音・打楽器音分離手法の提案

我々は、調波音・打楽器音の混合した音楽における時間周波数スペクトログラムが、周波数方向にはブロードだが時間方向に急峻な打楽器的な成分と、逆に周波数方向には急峻だが時間的に滑らかな調波音的な成分から成る点に着目し、この性質を満たすように時間周波数マスクを設計して分離を行う手法 [1][2] を提案した。

(2) リアルタイム分離システムの実現

[2] では時間周波数領域における EM アルゴリズムを用いた時間周波数マスクの反復推定手法を提案しているが、前述した調波音・打楽器音の性質を微分的に定義することで、局所的な分析領域での推定でも比較的良い性能で分離できる。これを利用し、分析時間区間の移動とパラメータ反復更新を交互に行なうことで、実時間での調波音・打楽器音分離を実現した。図 1 に提案システムの GUI 画面を示す。本システムでは、リアルタイムに分離した調波音・打楽器音のパワースペクトルを表示し、両者の音量バランスを調整しながら再生するという加工機能を実現した。

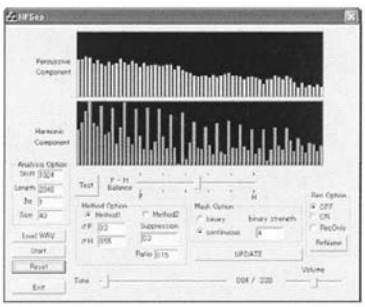

図 1: 調波音・打楽器音分離システムの GUI 画面

参考文献

[1] 宮本 賢一, 立薗 真理, ルルー ジョンナトン, 亀岡 弘和, 小野 順貴, 嵯峨山 茂樹, "スペクトログラム 2 次元フィルタによる調波音・打楽器音分離," 日本音響学会秋季研究発表会講演集, pp. 825-826, Sep, 2007.

[2] 宮本 賢一, 亀岡 弘和, 小野 順貴, 嵯峨山 茂樹, "スペクトログラムの滑らかさの異方性に基づいた調波音・打楽器音の分離," 日本音響学会春季研究発表会講演集, Mar, to appear, 2008.

Select&Voice + FlexibleShortcuts：GUI とのアナロジーに基づいた音声インタフェース

梅本暁，熊井朋之，中野鐵兵，小林哲則（早大）

ユーザ主導の効率的な音声インタフェースとして提案している，Select&Voice[1] と FlexibleShortcuts[2] を実装したシステムを紹介する．本システムでは，提示／選択／入力という GUI の基本概念に基づいたデータ入力操作と，連続キーワード入力をサポートしたメニュー選択操作を組み合わせた音声インタフェースを定義し，従来の対話型インタフェースでは困難であった初心者に対する容易性と熟練者に対する効率性を同時に実現した．

(1) Select&Voice を利用したデータ入力

GUI コンポーネントを音声拡張した音声 UI コンポーネント (voice widget) を定義し，これを共通部品として組合せた音声フォーム (voice form) を構築する．アプリケーションでは実行対象となる機能毎に voice form を用意する．実行時には，voice widget がそれぞれ入力可能な項目として提示され，ユーザは，入力対象とする項目を選択してから，それに対する音声入力を行う．また，異なる入出力デバイス環境においても一貫した選択／入力の音声操作を定義し，様々な使用環境をサポートする．特に車の運転中など画面が見れない状況でも，図 1 に示すような音声フィードバックを用いてシステムの状態を提示することにより，音声インタフェースの利点であるアイズフリーインタフェースを実現する．

(2) FlexibleShortcuts を利用した機能選択

操作対象となる voice form の選択手段として，FlexibleShortcuts を利用した voice menu を定義した．ここでは，システムで実行可能な複数の機能が木構造で表現される．各ノードにキーワードが，葉ノードには実行対象の機能，voice form が割り当てられる．voice menu はノード選択用の UI として用意され，ユーザはノード選択を繰り返すことで実行する機能を選択する．複数のキーワードを用いた音声によるノード選択がサポートされ，これが任意のノードに対する柔軟な音声ショートカットとして機能する．また，特定機能を一意に選択するためのキーワード系列を学習することで，実行したい機能の効率的な選択が可能となる．

図 1: Select&Voice (a) 上下ボタンで項目を選択，システムは選択された項目名と入力されている値を読み上げる．(b) 発話ボタンを押しながら選択した項目に対し音声入力をする．(c) 認識結果を表示し，読み上げる．

謝辞 本研究は，経済産業省，平成 18,19 年度戦略的技術開発委託費「音声認識基盤技術の開発」の一部として実施された．

参考文献

[1] 梅本暁，他：GUI とのアナロジーに基づいた音声インタフェースの提案と評価，音講論集，pp.63-66, Sept. 2007.

[2] 熊井朋之，他：機能構造と連続キーワード入力を利用した音声インタフェースのユーザビリティ評価，音講論集，pp.67-70, Sept. 2007.

音声入力 Web システム w3voice とそのアプリケーション

西村竜一, 三宅純平, 河原英紀, 入野俊夫（和歌山大）

Web ブラウザ上で動作する Web アプリケーションに，音声入力の機能を提供する w3voice システムと，それを用いて試作した音声 Web アプリケーションを紹介する．我々は，本システムの基本コンポーネントをフリーのソフトウェアとして公開しており，Web 開発者が自らの Web サイトに音声入力インタフェースを容易に追加できるようツール整備を行っている．同時に，プロジェクトの Web サイト（http://w3voice.jp/）にて，試作アプリケーションの公開実験を行っている．実際のインターネットユーザが音声認識や自動対話の音声 Web アプリケーションにアクセスしたログや発話を記録することで，音声インタフェースの利用実態の調査，分析を行うことが研究の目的である．

図 1 は試作した音声入力 Web サイトである．通常の Web サイトのコンテンツに加えて音声入力パネルが配置される．利用者は，音声入力パネルをマウス等で操作し，発話を入力する．録音中は音声入力パネルがレベルメータとして動作し，確実な入力をサポートする．録音の後，発話データは，Web サーバに転送され，音声認識，加工等の処理がなされる．最後に，動的に生成されたコンテンツが Web ブラウザ上に表示される．

本システムは，Java アプレットの実装による音声入力パネルと CGI プログラムから構成されている．アプリの利用に，特別なプログラムのインストールを要求せず，普段利用の一般的な Web ブラウザと OS 環境をそのまま使うことができる．通信プロトコルは HTTP を応用しており，Web にアクセスできる環境でならばどこからでも本システムを利用できる．また，PCM 録音された波形信号をサーバサイドで処理するアーキテクチャに設計した．このため，音声認識等の特定の用途に限らず，さまざまな音アプリケーションのフロントエンドとして本システムを適用することが可能である．

図 1: w3voice システムの実行画面（くだもの通信販売サイト: http://w3voice.jp/shop/）と音声入力パネル

参考文献

[1] 西村 竜一, 三宅 純平, 河原 英紀, 入野 俊夫: 音声入力・認識機能を有する Web システム w3voice の開発と運用, 情報処理学会研究報告, 2007-SLP-68-3, 2007.

STRAIGHT の歌唱への応用 ～合唱とモーフィング～

森勢将雅，吉田有里，高橋徹，西村竜一，入野俊夫，河原英紀 (和歌山大学)，豊田健一，片寄晴弘 (関西学院大学)

STRAIGHT[1] を用いることで，自然性を大きく損なうことなく，音声の基本周波数を操作することができる．この方法を用いることで，男性，女性歌手 8 名の単音の歌唱音声から，合唱を合成した．この作品は，RENKON04 において，音声合成システムによる歌唱作品の中で最も高く評価された [2]．ここでは，RENKON04 で用いられた作品と，その制作方法について紹介する．

歌唱モーフィングは，図 1 のインタフェースのように，歌声の声質・歌い回し平面での制御を狙っている．2 名の女性ボーカルを声質・歌い回しについてモーフィングを行い，合成された歌唱の歌手性について主観評価することで，歌手の識別に重要な手がかりは声質であることを示した [3]．これらは，情報処理学会の推薦論文として報告されている [4]．ここでは，評価に用いたデモを紹介するとともに，歌手の個性が，声質と歌い回しのどちらに現れるかを体験していただく．

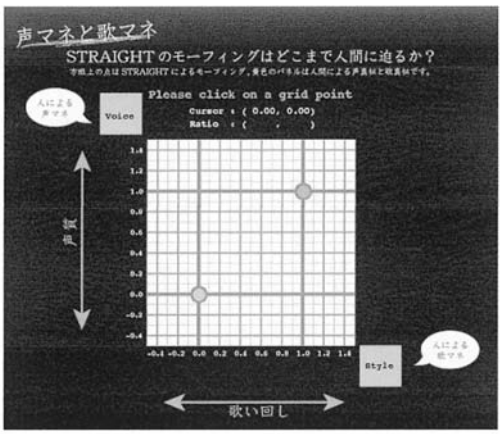

図 1: 歌唱モーフィングのインタフェース

謝辞 本研究の一部は，CrestMuse プロジェクトの支援を受けて行われた．

参考文献

[1] 河原英紀, “Vocoder のもう一つの可能性を探る–音声分析変換合成システム STRAIGHT の背景と展開–，” 日本音響学会誌，Vol.63，No.8，pp.442-449 (2007).

[2] Hideki Kawahara, Hideki Banno, Masanori Morise, Yumi Hirachi, A cappella synthesis demonstrations using RWC music database, Proc. NIME04, pp.130-131, 2004.

[3] 河原 英紀, 生駒 太一, 森勢 将雅, 高橋 徹, 豊田 健一, 片寄 晴弘, “歌唱モーフィングに基く音質と歌い回し転写の知覚的検討,” インタラクション 2007, 一般講演論文, pp.113–120, March, 2007.

[4] 河原 英紀，生駒 太一，森勢 将雅，高橋 徹，豊田 健一，片寄 晴弘，“ モーフィングに基づく歌唱デザインインタフェースの提案と初期検討，” 情報処理学会論文誌, Vol.48，No.12, pp.3637-3648 (2007)

歌唱力向上支援インタフェース MiruSinger へのブレス位置呈示機能の追加

中野倫靖 (筑波大)，後藤真孝 (産総研)
緒方淳 (産総研)，平賀譲 (筑波大)

著者らは以前、歌を「見る」歌唱力向上支援インタフェース MiruSinger を提案した [1]。MiruSinger は、ユーザの歌唱音声の音高 (F_0) 軌跡とビブラート[1] 区間を、視覚呈示する機能を持っていた。本デモでは、歌唱者のブレス (息継ぎ) 位置を視覚呈示する機能を追加したので紹介する。また、音楽 CD 中のボーカルのブレス位置も、ラベル付けして利用できるようにした。

(1) 機能

MiruSinger は、歌唱音声の F_0 軌跡を見ながら歌ったり (歌って見る)、聴いたり (聴いて見る)、描いたり (描いて見る) できる。これらの機能にブレス位置を見る機能を追加した (図 1)。マイク入力されたユーザの歌唱音声を *User*、音楽 CD 中のボーカルの歌唱音声を *Ref.* とすると、現在の実装は以下のような機能を持つ。

歌って見る User と Ref. の F_0 とビブラートを見る
＋ラベル付けされた Ref. のブレス位置を見る機能

聴いて見る 過去の User と Ref. の F_0 とビブラートを見る
＋過去の User と Ref. のブレス位置を見る機能

描いて見る Ref. の F_0 推定結果をマウス操作で修正できる
＋ Ref. のブレス位置をラベル付けできる機能

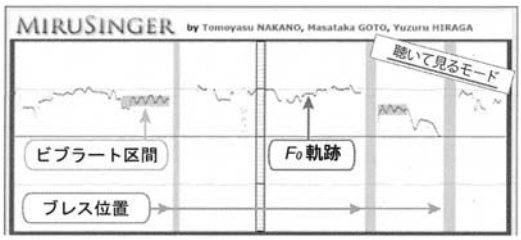

図 1: MiruSinger の実行画面 (F_0 描画部分のみの表示)。画面は「聴いて見るモード」で、F_0 等は右→左に流れる。

(2) 実現方法

無伴奏の歌唱音声からブレス位置を自動検出する技術が必要である。現在の実装では、HMM によって「ブレス/無音/歌声」の音響モデルを構築して、ブレスを自動検出した [2]。また、音楽 CD 中のブレス位置はラベル付けする必要があるが、将来は自動検出を目指す。

参考文献

[1] 中野倫靖, 後藤真孝, 平賀譲：MiruSinger: 歌を「歌って/聴いて/描いて」見る歌唱力向上支援インタフェース. 情報処理学会 インタラクション 2007 論文集 (インタラクティブ発表), pp.195-196, 2007.

[2] 中野倫靖, 緒方淳, 後藤真孝, 平賀譲：無伴奏歌唱におけるブレスの音響特性と自動検出, 日本音響学会 2008 年春季講演論文集, 2008. (掲載予定)

[1] ビブラートとは、主に音を伸ばすときに周期的に音高を変化させる（揺らす）歌唱テクニックである。

歌声と話し声の自動識別システム

大石康智 (名古屋大学)，後藤真孝 (産総研)
伊藤克亘 (法政大学)，武田一哉 (名古屋大学)

音声認識システムが扱うことのできる発話の対象は，人間が日常のコミュニケーションに用いる多様な発話様式の中の，ごく一部（読み上げや講演など）に限定されている．対象を拡大するために，発話様式や個人性に対応するための技術が必要である．本研究の目的は，感情音声や歌声のような発話様式の違いを特徴付ける物理的あるいは信号的性質を明らかにすることであり，その第一歩として，通常の話し声 (読み上げや講演音声など) との違いを聞き分けやすい歌声を研究対象に取り上げた．

聴取実験より，人間は 2 秒の歌声，話し声を 100%識別可能であり，特に短時間のスペクトル特徴，韻律の時間変化が識別の手がかりとなることを確認した．これらの知見を客観的に評価するために，識別特徴量を提案して自動識別実験を行ったところ，2 秒の歌声，話し声の音声信号に対して 84.7%の識別性能が得られた [1].

本報告では，提案する自動識別手法を利用して，リアルタイムに識別結果が提示される歌声と話し声の自動識別システムを実装する．識別システムの実現方法を図 1 に示す．スペクトル包絡を表す MFCC12 次までの係数とその時間変化 ΔMFCC，基本周波数 (F0) の時間変化 ΔF0 を識別特徴量に利用する．事前に大規模な歌声・話し声データベースを利用して，この特徴ベクトルの分布を混合ガウス分布 (GMM) によって学習する．システムに音声が入力されると，入力音声を特徴ベクトル列に変換し，歌声 GMM，話し声 GMM に対する特徴ベクトルの尤度を比較することによって，識別結果が出力される．

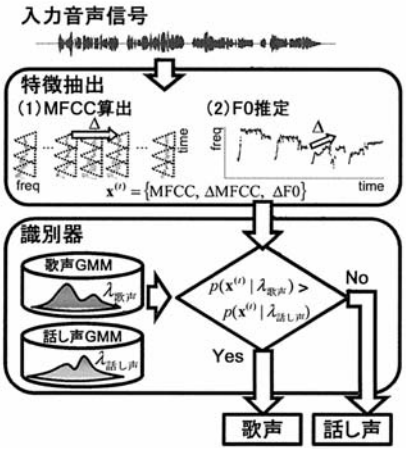

図 1: 歌声と話し声の自動識別システムの概要

参考文献

[1] 大石康智，後藤真孝，伊藤克亘，武田一哉：スペクトル包絡と基本周波数の時間変化を利用した歌声と朗読音声の識別，情報処理学会論文誌，Vol.47, No.6, pp.1822–1830, 2006.

楽曲を解説したテキストと音響特徴量との関連付けを利用した楽曲推薦システム

高橋 量衛, 大石 康智, 武田 一哉 (名古屋大学)

ユーザが Web ページを閲覧しているときに，自動的に BGM を流す新しい楽曲推薦システムを目的として，テキスト (文章) をクエリーとする楽曲検索システムを紹介する．入力テキストに対して楽曲を検索するために，テキストに含まれる語彙と楽曲とを関連付けることが必要である．語彙の共起関係に基づくドキュメント空間と，音響的特徴空間とを利用した関連付け手法 [1] を提案し，楽曲検索 (推薦) システムを構築した．概要を図 1(左図) に示す．

(1) テキストと音響特徴量との関連付け手法

ドキュメント空間と音響的特徴空間との関連付けは，以下のように二つの特徴ベクトルを利用して実現する．楽曲を解説したテキストに出現する語彙の頻度に基づき，テキストをベクトルで表現する (文書ベクトルと呼ぶ)．また，楽曲を信号処理することにより得られる音響特徴量の頻度分布に基づき，楽曲の音響的特徴をベクトルで表現する (音響ベクトルと呼ぶ)．これら二つの特徴ベクトルを変換行列により関連付ける．

(2) 検索方法

図 1(右図) に検索結果例を示す．ユーザが入力したテキストに対して，次のように楽曲を検索する．

1. テキストクエリーに出現する語彙の頻度に基づき，テキストを文書ベクトルで表現する．
2. 変換行列を用いて文書ベクトルから音響ベクトルを推定する．
3. 入力テキストから推定した音響ベクトルと，検索対象曲の音響ベクトルとの距離により楽曲を選択する．

テキストクエリーに出現する語彙に基づいて楽曲が検索され，さらに音響的に類似した楽曲をも選曲することが可能となる．

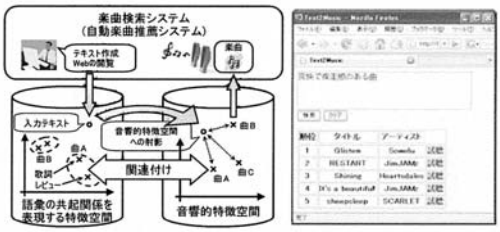

図 1: 楽曲を解説したテキストと音響的特徴との関連付けを利用した楽曲推薦 (検索) システム，(左図:概要，右図:検索結果例)．

参考文献

[1] 高橋量衛, 大石康智, 武田一哉：Web から収集した楽曲を説明するテキストと楽曲の音響特徴量との関連づけに関する検討, 情処研報, 2007-MUS-72, pp.85–90, 2007.

歌声素材生成 Web ツール “ぼーか郎”

野池賢二 (株式会社トランス・ニュー・テクノロジー)

歌声を含む音楽素材を手軽に生成し，コンテンツ制作に利用しやすい形式で得られる Web ツール “ぼーか郎” を紹介する．“ぼーか郎” は,「歌う TiMidity」[1] の仕組みを基礎とし，パラメータ割当てや声素材の調整・追加を施した，ソフトウェアサンプラーによる歌声生成 Web ツールであり，次の特徴がある．

- Web ツールであるため，OS やハードウェアなどのクライアント環境を問わずに使用できる．
- 多様な入力方式が可能である．
  - MML(mml2mid，はてなダイアリー MML)
  - ピアノロール (VOCALOID Editor を利用)
  - 五線譜 (Finale やスコアメーカーなどを利用)
  - SMF の Text Event，Lyric Event に歌詞情報を記述
- 多様な出力形式を用意している．
  - MP3, WAV, AIFF
  - 携帯電話用 3GP(着信音として設定可能)
  - VOCALOID VSQ, MIDI 形式
  - SMF(歌声情報がプログラムチェンジ情報として埋め込まれている)
- 既に用意してある「普通の声」,「ささやき声」のほかに，音節を単位とした声素材を用意することよる声色の追加が可能である．ユーザによる声素材の作成・提供を支援するツールの提供も予定している．
- ポリフォニー音源であるので，和音や複数旋律からなる歌声の作成が容易である．また人間の発声モデルを仮定していないため，隣り合う音節の発音時間が重なっていてもよく，通常の打ち込みと同様の手法で滑らかに歌うデータを作成できる．
- マルチティンバー音源であるので，複数の声色や，楽器音を含んだ素材の作成が容易である．
- すべて無料で使用できる．

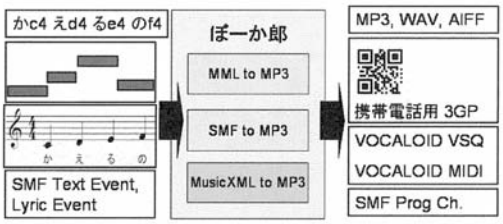

図 1: ぼーか郎の入出力

ぼーか郎は 図 1 に示した入出力の一部を除いて http://noike.info/~kenzi/cgi-bin/mml2mp3/ から，いつでも自由に利用できる．研究色の薄いツールではあるが，素材の作成以外に，歌声生成システムとしての最低品質基準として他研究から参照されれば幸いである．

参考文献

[1] 本本雅彦：歌う TiMidity, http://www.ohnolab.org/~kimoto/timidity/, 2004.

固有声変換に基づくボイスチェンジャー

戸田智基（奈良先端科学技術大学院大学）

ボイスチェンジャーは，入力された音声の声質を変換する装置であり，映画の吹き替えや発声障害者補助など様々な用途への応用が期待される．これまでに，簡易な音声パラメータ操作（例えば，周波数軸伸縮処理や基本周波数シフト処理）に基づく方式が実現されているが，実際に得られる変換音声の声質は入力音声の声質に大きく依存する．そのため，所望の声質への変換を実現するのは極めて困難である．これに対して，統計的手法に基づく声質変換法を導入することで，より幅広い声質への変換が実現できる．ここでは，その一例として，固有声変換 [1] に基づくボイスチェンジャーを紹介する．

(1) 固有声変換

固有声変換は，事前に収録された多数話者の音声データを事前知識として活用することで，所望の話者間における統計的声質変換モデルを効果的に構築する技術である [1]．この技術により，例えば，言語情報を一切必要とせずに，所望の話者による極少量の音声データのみを用いて，変換モデルを構築できる．また，少量のパラメータ操作により，変換音声の声質を自在に制御できる．

(2) 固有声変換ソフトウェア

固有声変換ソフトウェアは，任意の人の声をある特定の人の声へと変換する多対一変換用と，ある特定の人の声を任意の声質へと変換する一対多変換用の二つからなる．図 1 に固有声変換ソフトウェアの GUI 画面を示す．

多対一固有声変換ソフトウェア：入力された音声をアニメのキャラクターのような声へと変換する．変換先となるキャラクター声は 30 種類用意されている．

一対多固有声変換ソフトウェア：男性的／女性的などの声質表現語スライダーの操作により，ある特定話者の音声を様々な声へと変換する [2]．特定話者のサンプルとして，男性 1 名及び女性 1 名に対するデモが用意されている．なお，50 文からなる学習用音声データを用意すれば，各ユーザー専用の変換モデルを構築できる．

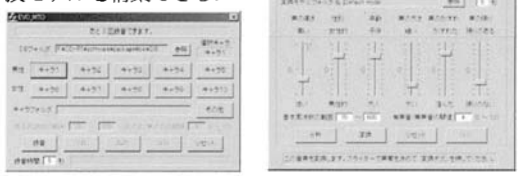

図 1: 固有声変換ソフトウェア（左：多対一固有声変換ソフトウェア，右：一対多固有声変換ソフトウェア）

謝辞　本ソフトウェアは情報処理推進機構 (IPA) 未踏ソフトウェア創造事業の援助を受けて開発したものである．共同開発者である奈良先端科学技術大学院大学の大谷大和氏，中村圭吾氏，関本英彦氏（現在はオムロン株式会社）に感謝する．

参考文献

[1] T. Toda et al., *Proc. ICASSP*, pp. 1249-1252, Hawaii, USA, Apr. 2007.

[2] K. Ohta et al., *Proc. 6th ISCA Speech Synthesis Workshop*, pp. 101-106, Bonn, Germany, Aug. 2007.

おわりに

梶 克彦 (NTT)

今回の合同研究会でのデモセッションは，当初追加募集を前提として募集日時を設定していたが，そのような配慮は全く必要がなかった．結局（多少の根回しもあり）一次締切の時点で 14 件にものぼる応募をいただいた．一般発表についても同様に，パラレルセッションを組まざるを得ないほど多くの申し込みをいただいた．このように音楽・音声の研究者の多くが，今回の合同研究会を発表の場に選んでいただいたことからも，お互いの研究分野での交流が重要であることを共通意識として持っていただけていることが伝わってきた．

今回は前述のように多くの一般発表の申し込みがあったこともあり，デモセッションを昼間のセッションとして組み込むことができなかった．そこでデモセッションをイブニングセッションと位置づけ，夕食懇親会後に行われることにした．音楽情報科学研究会では，デモセッションを夕食懇親会後に行うことが恒例となりつつあり，毎回大いに盛り上がりをみせている．今回も実際に体験してみたいと思わせる内容のデモが集まっているので，懇親会での盛り上がりそのままにデモセッションに移っていただき，大いに熱い議論をしていただけるだろうと期待している．

また今回のデモセッションで刺激を受けた方は，ぜひ次回以降のデモセッションの機会に積極的に参加していただきたい．システムを構築するにあたり，主観的な意見は非常に有意義なものとなる．多くの人にシステムを触ってもらうことで，システム設計者自身では気付かなかったような点が指摘され，さらにシステムの良い所を伸ばし，欠点を修正することにつながるだろう．また，直接はシステムに結び付かない要素技術の研究であっても，その研究の応用としてシステムを構築し，実際に動くモノとしてデモをすることで，研究内容の把握を促進し，応用の可能性について議論することができる．

今回デモ発表をしていただく方の中には，若手の研究者が多い．さらに，研究会の枠を超えて，多くの若手研究者の方々に今回のデモセッションの運営に協力していただいている．異なる研究分野間の交流は多角的な視点からの意見を得ることができ，また新たな刺激を生む貴重な機会であるので，今回のデモセッション運営協力を，今後の異なる研究分野間で積極的に交流するきっかけにしていただきたいと思う．

音声言語の研究は日常生活のコミュニケーションに，音楽は日常の娯楽・芸術に密着した研究分野であり，それぞれ一般社会と非常に関連の深い研究分野であるといえる．実際，音楽・音声に関する研究成果の多くが，一般社会に浸透しているシステムにつながっている．今回発表されたデモシステムはもちろん，今回のデモセッションに触発された研究についても，将来一般社会に浸透し，豊かな社会を創る立役者となってくれることを願っている．