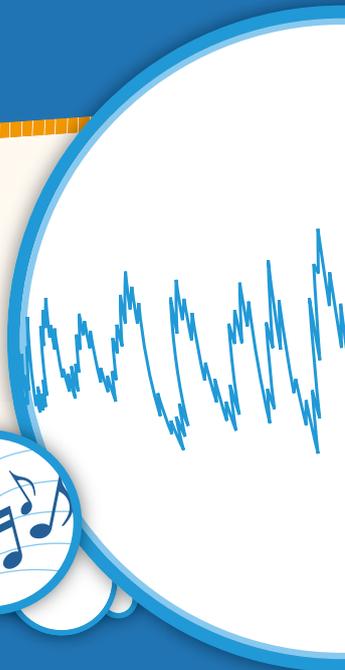


● OngaCREST シンポジウム2014

# 音楽情報処理研究が 切り拓く未来を探る

予稿集





## 開催概要

- 開催名** OngaCREST シンポジウム 2014  
「音楽情報処理研究が切り拓く未来を探る」
- 開催日** 2014年8月23日
- 会場** 明治大学 中野キャンパス
- 主催** OngaCREST プロジェクト  
「コンテンツ共生社会のための類似度を可視化する情報環境の実現」
- 研究代表者 後藤 真孝 (産業技術総合研究所 首席研究員)  
研究分担者 森島 繁生 (早稲田大学 教授)  
研究分担者 中村 聡史 (明治大学 准教授)  
研究分担者 吉井 和佳 (京都大学 講師)



## プログラム

- 13:00-13:15 **研究代表者よりご挨拶**  
後藤 真孝 (産業技術総合研究所)
- 13:15-14:00 **「OngaCRESTプロジェクトの狙いと研究成果の紹介」**  
後藤 真孝 (産業技術総合研究所)
- 14:00-14:15 **「音楽動画の創作・鑑賞支援に関する研究成果の紹介」**  
森島 繁生 (早稲田大学)
- 14:15-14:30 **「音楽動画とインタラクションに関する研究成果の紹介」**  
中村 聡史 (明治大学)
- 14:30-14:45 **「統計的音楽音響信号処理に関する研究成果の紹介」**  
吉井 和佳 (京都大学)
- 15:00-15:45 **招待講演「初音ミクが切り拓いてきた世界、切り拓く未来」**  
伊藤 博之 (クリプトン・フューチャー・メディア株式会社 代表取締役)
- 16:00-18:00 **ポスター・デモ展示**
- 18:15-20:15 **懇親会 (有料、立食形式)**



 OngaCREST シンポジウム 2014

# 音楽情報処理研究が 切り拓く未来を探る

予稿集

# はじめに

OngaCREST シンポジウム 2014「音楽情報処理研究が切り拓く未来を探る」にご来場頂き、どうもありがとうございます。我々の「OngaCREST（音画 CREST/おんがくれすと）プロジェクト」では、音楽の聴き方・創り方の未来を切り拓く技術開発により、音楽の楽しみ方がより能動的で豊かになり、人類と音楽との関係がより良い形で発展していけるようになる情報環境の構築を目指しています。まだ 5 年間の研究プロジェクトの 3 年目の中間段階ですが、これまでの研究成果をみなさまにご紹介するために、この公開シンポジウムを企画いたしました。我々の最新の研究成果を楽しみながら、さまざまなご意見やアドバイスを頂き、みなさまと手を携えて、学術的な貢献をしつつ社会の役に立てるような研究開発を目指していきたいと考えています。

本シンポジウムでは、まずホールにて、OngaCREST プロジェクトを推進する研究代表者と 3 名の研究分担者が、OngaCREST プロジェクトの研究成果を紹介する登壇発表をします。次に、ゲストとしてクリプトン・フューチャー・メディア株式会社の伊藤博之社長をお招きして、技術が生み出す先進的なコンテンツ文化の発展に関する招待講演をして頂きます。その後、ポスター・デモ展示の会場に移動し、OngaCREST プロジェクトの研究成果をさまざまな研究者が実演しながらご紹介します。

我々の研究活動は、多くの方々に支えて頂くことで、これまで大きく発展して行くことができました。日ごろからご支援・ご議論頂くみなさまに深く感謝すると共に、今後も引き続き情熱と信念を持って未来社会において大切な技術を研究開発し、音楽情報処理研究が切り拓く未来をみなさまと共に探っていきます。今後とも、どうぞよろしく願いいたします。



研究代表者 後藤 真孝

---

## 研究代表者 略歴

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）。現在、産業技術総合研究所 情報技術研究部門 首席研究員 兼 メディアインタラクション研究グループ長。IPA 未踏 IT 人材発掘・育成事業 プロジェクトマネージャー等を兼任。日本学士院学術奨励賞、日本学術振興会賞、ドコモ・モバイル・サイエンス賞 基礎科学部門 優秀賞、科学技術分野の文部科学大臣表彰 若手科学者賞、情報処理学会 長尾真記念特別賞、星雲賞【ノンフィクション部門】等、38 件受賞。国際音楽情報検索学会 第 10/15 回国際会議 ISMIR 2009/2014 General Chair、2007～2008 年度 情報処理学会 音楽情報科学研究会 主査、2014～2015 年度 情報処理学会 理事。計算機によって音楽・音声コンテンツを自在に扱える技術の確立を目指し、音の自動理解や音インタフェース、歌声情報処理等の研究を 22 年間推進。

# 目次

はじめに	2
目次	3
OngaCREST プロジェクトの狙いと研究成果の紹介	4
ポスター集	9
コンテンツ共生社会のための類似度を可視化する情報環境の実現	10
Songle: Web 上の楽曲の中身を自動解析する能動的音楽鑑賞サービス	12
Songrium: 音楽コンテンツの関係性を可視化する音楽視聴支援サービス	13
確率的生成モデルに基づく音楽の類似度とありがち度の推定	14
歌声トピックモデルに基づく類似歌声検索とトピック可視化	15
MachineDancing: 機械学習に基づく音楽に連動したダンスの自動生成	16
TextAlive: 音楽に同期した歌詞の Kinetic Typography 制作環境	17
AutoRhythmGuitar and AutoMashUpper	18
歌唱音声の統計的知覚年齢制御	19
歌手映像と歌声の解析に基づく音楽動画中の歌唱シーン検出	20
LyricsRadar: 歌詞の潜在的意味分析に基づく歌詞検索インタフェース	21
半教師あり NMF を利用した音楽信号中のフレーズ検出	22
視聴者コメントに基づく楽曲動画の印象推定とデータセット構築	23
音楽理解とアノテーション分析技術によるサムネイル動画自動生成	24
N 次装飾: 動画共有サイト上の動画に対する装飾とその共有手法	25
自由なテンポで演奏した複数の演奏データから楽曲を生成するシステム	26
VRMixer: 動画と現実の融合による新たなコンテンツの生成	27
統計的機械学習に基づく音楽解析	28
統計的機械学習に基づく言語＝音楽情報処理の横断的展開	29
音楽音響信号中の調波音の周波数特性およびドラムの音色の置換システム	30
音楽音響信号中の歌声 F0 軌跡に対する歌唱表現の転写システム	31
OngaCREST プロジェクト研究参加者一覧	32

# OngaCREST プロジェクト 「コンテンツ共生社会のための 類似度を可視化する情報環境の実現」 の狙いと研究成果の紹介

後藤 真孝

産業技術総合研究所

本研究は、音楽や動画のようなメディアコンテンツを豊かで健全に創作・利用する「コンテンツ共生社会」に貢献できるように、膨大なメディアコンテンツ間の類似度を人々が知ることができる（可視化する）情報環境を実現する技術基盤の構築を目的とする。さらに、創作支援技術と鑑賞支援技術を研究開発することで、コンテンツの創作や鑑賞を人々が能動的に楽しめる社会や、過去のコンテンツに敬意を払う文化、感動体験重視型のコンテンツ文化の実現に寄与することを目指す。

## 1. OngaCREST プロジェクトの狙い

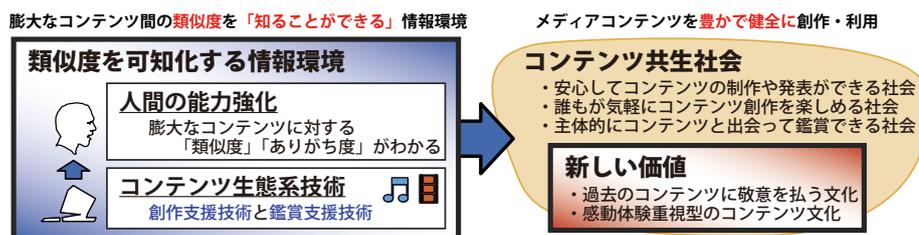
本研究プロジェクト「コンテンツ共生社会のための類似度を可視化する情報環境の実現」（プロジェクト名：OngaCREST（音画 CREST/ おんがくれすと）プロジェクト、実施期間：2011年10月～2017年3月、研究代表者：後藤 真孝、研究分担者：森島 繁生、中村 聡史、吉井 和佳）では、図1のように、過去のコンテンツに敬意を払う文化を築きつつ、感動体験重視型の新たなコンテンツ文化が形成される社会の実現に寄与するために、膨大なコンテンツ間の類似度を人々が知ることができる（可視化する）情報環境を実現することを目的としている。そして、音楽の聴き方・創り方の未来を切り拓く技術開発により、音楽の楽しみ方がより能動的で豊かになる変化を日常生活に起こすことを目指している。本プロジェクトでは、「コンテンツ」として主に音楽あるいは音楽動画（ミュージックビデオ、ダンス動画等）を対象とする。

## 1.1 類似度に関する人間の能力の限界

過去から未来に渡る膨大なメディアコンテンツがデジタル化され、適切な対価で自在にアクセス可能な社会が到来しつつある。これは利便性の面で望ましいが、今後もアクセス可能なコンテンツが単調増加し続けていく結果、コンテンツ制作者にとって、自分の作品が過去も含めた膨大なコンテンツに埋もれやすくなり、視聴してもらおうのがますます困難になっていく。逆にコンテンツ視聴者側から見れば、膨大なコンテンツの中から選択するのが難しくなっていく。さらに、コンテンツの単調増加に伴って似ているコンテンツも単調増加して、本来は盗作ではないにも関わらず、盗作疑惑を招く事例が増えてしまう懸念がある。あらゆる創作は既存コンテンツの影響を受けており、「無自覚に」「何らかの意味で」「部分的に」類似してしまうのは自然であり、本来問題ではないことが多い。

しかし、類似度に関する人間の能力には限界があ

専門家だけでなく一般の人々が「何が似ているのか」「どれぐらいありふれているのか」を知ろうと思えば自在に把握して活用できる技術基盤を構築



対象とする主な時系列メディアコンテンツ：音楽（音楽音響信号等）、音楽動画（ミュージックビデオ、ダンス動画等）

図1: OngaCREST プロジェクトの概要

る。人間は目の前の二つを比較して類似度を判断することはできても、その判断速度には限界があり、似たコンテンツを 100 万個の中から探すことはできない。さらに、人間は過去の経験に基づいて高度な判断ができる一方、ある事象が全体の中でどれぐらい確率的に起こりえるのか、という「ありがち度」の判断には限界がある。例えば、最近よく起きた事象を起きやすいと思いついたり、多数起きている事象でも遭遇しなければ滅多にないと誤解する。アクセス可能なコンテンツをすべて視聴することは量的に不可能なため、全体を俯瞰した適切な判断をすることができない。

そのため、コンテンツが単調増加することによって、もし自分の作品が何かに似ていると糾弾されるリスクが高いと、人々が創作と発表を楽しむ「一億総クリエイター時代」が到来しても、人々が安心してコンテンツの制作や発表をしにくい社会になりかねない。コンテンツは様々な要素から構成されているが、実はありふれている（出現確率の高い）要素にも関わらず、その要素が他のコンテンツに似ているだけで「オリジナリティに問題がある」という勘違いが生じてしまうのは不毛な盗作疑惑だと考えられる。本来、創作活動は過去のコンテンツの積み重ねの上に成り立っており、共通して使われるありがち度の高い要素・表現は、人類共有の知として制作者や視聴者の間で適切に可視化され、活用されるとよい。

## 1.2 人々が類似度とありがち度を活用できる情報環境

そこで我々は、専門家だけでなく一般の人々が「何が似ているのか」「どれぐらいありふれているのか」を知ろうと思えば自在に把握して活用できる技術基盤を構築することが重要だと考え、本プロジェクトで取り組んでいる。それにより、人々が今後も安心してコンテンツの制作や発表を続けられるようにしたい。そして、「ありがち度」の高い事象（例えばコード進行やジャンルごとの慣例的な事象）は人類共有の知として活用できる創作支援技術を実現していくことで、非専門家でも気軽にコンテンツ創作を楽しむようにする。さらに、類似度に基づいて新たなコンテンツと出会える鑑賞支援技術を実現していくことで、能動的にコンテンツと出会って鑑賞できるようにする。

こうした研究開発により、学術論文のようにリ

ファー（参照・引用）され再利用されたら喜びを感じられる新たな音楽文化が支援できれば、過去のコンテンツと共存共栄し、敬意を払う文化を築くことに貢献できる。コンテンツは「他にいかに類似していないか」という新規性だけに価値があるのではなく、人々をいかに感動させ幸せにするかに価値があると、我々は考えている。新規性だけを追求しなくても、コンテンツに感動させる魅力や完成度の高さ等があれば価値があることが重要である。むしろ学術論文等と同様に、多くの作品をリファーしてその土台の上に成り立っているからこそ価値が生まれる状況も望ましい。そうすることで感動体験重視型の新たなコンテンツ文化が形成されることが期待できる。

デジタルコンテンツ社会ではもはや忘却できず、ともすると、単調増加する過去の膨大なコンテンツに未来が押しつぶされかねない。本プロジェクトは、デジタル化による「忘却できない社会」において豊かで持続発展可能な「コンテンツ共生社会」を築く挑戦に位置づけられる。過去のコンテンツと未来のコンテンツとの共生を人々が実感できるようにすると共に、人間とコンテンツとの共生により膨大なコンテンツを楽しめる社会の実現に寄与することを目指していく。

## 2. プロジェクト中間段階での主な研究成果

本研究では、類似度を可視化する情報環境によってコンテンツ共生社会を実現するための技術を、コンテンツの豊かな生態系を育てていくための「コンテンツ生態系技術」と位置付け、音楽コンテンツの創作と鑑賞を総合的に支援する技術の研究開発を進めている。はじめに、2.1 節の鑑賞支援技術の研究開発に先行して着手し、その後、2.2 節の創作支援技術の研究開発に着手した。それと並行して、2.3 節の類似度・ありがち度の推定やそれに資する音楽理解技術、コンテンツ解析技術の研究開発にも持続的に取り組んでいる。

以下では、OngaCREST シンポジウム 2014 の 20 件のポスター題名を『』に囲んで言及しながら、プロジェクト中間段階での主な研究成果を紹介する。

### 2.1 鑑賞支援技術

鑑賞支援技術に関する最も代表的な成果は、『Songle: Web 上の楽曲の中身を自動解析する能動的

音楽鑑賞サービス』[1]と『Songrium: 音楽コンテンツの関係性を可視化する音楽視聴支援サービス』[2]の二つの Web 上のサービスであり、既に一般公開して実証実験を開始している。Songle (<http://songle.jp>) は 2012 年 8 月 29 日に、Songrium (<http://songrium.jp>) は 2013 年 8 月 27 日に、産業技術総合研究所と科学技術振興機構 (JST) が共同でプレス発表をした。Songle は音楽理解技術を用いて、楽曲を「音楽地図」として表現することで、音楽の楽しみ方をより深く、多様化するサービスである。一方 Songrium は Web マイニング技術と音楽理解技術を用いて、音楽コンテンツの関係性をより深く理解して楽しめるようにするサービスである。前者が単一の楽曲の中身を対象とした技術であるのに対し、後者は複数の楽曲間の関係を対象とした技術であり、相補的に連携させて実現した。

音楽理解技術に基づく鑑賞支援技術に関して先駆的な「能動的音楽鑑賞インタフェース」[3]の一連の研究では、様々な事例を、(a) 音楽再生、(b) 音楽加工、(c) 音楽検索・ブラウジングの三つのカテゴリに整理して議論している。上記はその (a) と (c) に該当する。

カテゴリ (c) に該当する他の研究として、歌詞に着目した鑑賞支援技術に位置付けられる『LyricsRadar: 歌詞の潜在的意味分析に基づく歌詞検索インタフェース』[4]を実現した。これは、歌詞テキストのトピック（潜在的な意味）の推定結果に基づいて、様々な歌詞をブラウジングしながら検索できるインタフェースである。音楽動画（ミュージックビデオ等）を対象とした『音楽理解とアノテーション分析技術によるサムネイル動画自動生成』[5]は、多数の音楽動画の中から好みの動画を検索する際の鑑賞支援技術である。Songle のサビ区間検出技術に、動画共有サービス「ニコニコ動画」のコメント分析結果を組み合わせることで、ハイライトシーンを抽出可能とした。

カテゴリ (b) の音楽加工は、既存の楽曲を自分好みにカスタマイズ（パーソナライゼーション）する鑑賞支援技術であり、カジュアルな創作支援技術にも位置付けられる。『歌声トピックモデルに基づく類似歌声検索とトピック可視化』[6]では、歌声のトピック分析に基づいて類似歌声検索を実現しただけでなく、既存楽曲の音高をピッチシフトで変更後に鑑賞すると、歌手の個性が変わってより自分好

みにできる可能性があることを示した。『音楽音響信号中の調波音の周波数特性およびドラムの音色の置換システム』[7]では、既存楽曲の周波数特性とドラム音色を、自分好みの楽曲に近くなるように加工（置換）して鑑賞可能にした。さらに『音楽音響信号中の歌声 F0 軌跡に対する歌唱表現の転写システム』[8]では、既存楽曲中の歌唱を、自分好みの歌唱スタイルになるように加工（転写）して個人的に楽しむことを可能にした。

## 2.2 創作支援技術

誰でも手軽に使える敷居が低く、かつ汎用性と自由度が高い創作支援技術を実現することは難しく、第一段階として、非専門家に対する創作支援インタフェースの土台となる自動生成のための各種基礎技術や、音楽コンテンツの個別要素をインタラクティブに編集可能なインタフェースの研究開発を推進してきた。

楽曲を対象とした創作支援技術に位置付けられる『AutoRhythmGuitar and AutoMashUpper』の一つ目の AutoRhythmGuitar [9] は、ギター伴奏のためのタブ譜を自動生成する手法であり、ギタリスト毎のタブ譜でそのスタイルを機械学習しておくことで、任意のコード進行が与えられたときに、学習したスタイルに沿った伴奏を生成できた。二つ目の AutoMashUpper [10] は既存楽曲を重ね合わせるマッシュアップ音楽の制作支援インタフェースであり、楽曲断片間の類似度等から「マッシュアップ可能性 (mashability)」を計算することで、インタラクティブなマッシュアップ制作を可能にした。一方、『自由なテンポで演奏した複数の演奏データから楽曲を生成するシステム』[11]では、自由なテンポで演奏したクラシック等の個別パートを、収録後に時間伸縮により同期して合成することを可能にした。

音楽動画において、楽曲に連動した映像を対象とした創作支援技術についても取り組んでおり、『MachineDancing: 機械学習に基づく音楽に連動したダンスの自動生成』[12]では、「ダンス語彙」を機械学習して自動獲得することで、3次元 CG キャラクターのダンス動作（骨格構造の位置と回転角の時間変化）を、単なる動作断片の切り貼りでなく、音楽に連動して新たに自動生成することができた。また、歌詞の映像表現を対象とした『TextAlive: 音楽に同期した歌詞の Kinetic Typography 制作環境』

[13] では、単なるカラオケ表示とは違う、歌詞の魅力的なアニメーション表現を制作するインタフェースを実現した。他にも『VRMixer: 動画と現実の融合による新たなコンテンツの生成』では、既存のダンス動画等に自分が入り込んだ動画の制作を支援でき、『N 次装飾: 動画共有サイト上の動画に対する装飾とその共有手法』[14] では、Web ブラウザ上での既存動画の再生に合わせて、効果音や BGM の付与等の簡易装飾を加えることができた。

### 2.3 類似度・ありがち度の推定と音楽理解技術

類似度・ありがち度は、その対象や推定方法に様々な可能性が考えられるが、様々な音楽的要素について個別に推定できることは重要である。そこで、『確率的生成モデルに基づく音楽の類似度とありがち度の推定』[15] では、4 種類の音楽的要素の確率的生成モデルを構築し、モデルからの生成確率を計算することで、「楽曲間の類似度」や「楽曲のありがち度」を推定した。また、入力した楽器演奏フレーズと、混合音中のフレーズとの間の類似度に特化して取り組むことで、『半教師あり NMF を利用した音楽信号中のフレーズ検出』[16] も可能にした。

より深い音楽理解技術の実現を目指した『統計的機械学習に基づく音楽解析』では、音楽の教師なし構造学習の実現を目指して研究開発を進め、無限複合自己回帰モデル [17] や、無限半正定値テンソル分解 [18] を実現した。他にも、コード進行の語彙フリー無限グラムモデルを提案した。さらに『統計的機械学習に基づく言語＝音楽情報処理の横断的展開』では、歌詞の各単語の意味の連続空間への埋め込み [19] や、動画コメントからの感情語抽出に取り組んだ。

歌声の類似度に関しては、前述した歌声トピックモデルに基づく類似歌声検索でも扱っていたが、歌声の別の側面として、知覚年齢の観点から扱う『歌唱音声の統計的知覚年齢制御』[20] にも取り組んだ。また、楽曲の歌声区間の自動検出結果に、その歌手映像から顔や口の動きも検出して統合することで、音楽動画中の歌唱シーンを検出する『歌手映像と歌声の解析に基づく音楽動画中の歌唱シーン検出』[21] も実現した。

以上のような客観的な推定技術だけでなく、音楽動画から受ける主観的な印象の類似度に基づく検索も重要であり、『視聴者コメントに基づく楽曲動画

の印象推定とデータセット構築』では、動画共有サービス「ニコニコ動画」のコメントに基づく印象推定や、音楽動画に対して印象をラベル付けしたデータセットの構築・公開に取り組んでいる。

### 3. 音楽情報処理研究が切り拓く未来を探る

1 章では OngaCREST プロジェクトの狙いと意義を述べたが、それ以外の観点からも本プロジェクトの意義を議論することができる。これまでも文献 [22] では、音楽情報処理研究の進展によって人類の幸せに寄与していくために、「心理的な現状認識の微分値が正な状態」を達成する重要性や、「単位資源（リソース）当たりの価値向上」を考慮する重要性について議論してきた。ここではさらに新たな観点として、「コピー不可能な体験」が持つ価値の重要性を議論する。

本プロジェクトで取り組んでいる、「一億総クリエイター時代」に資する創作支援技術や、能動的音楽鑑賞のためのインタフェースやサービスを生み出す鑑賞支援技術は、実は「コピー不可能な体験」を創出するという重要な意義を持っている。音楽演奏の鑑賞を音楽 CD や DVD 等によってコピー可能な体験としたように、体験をコピーする技術手段を段階的に高度にしながらコンテンツ産業は発展してきた。ところが、流通コストが限りなく 0 に近づくデジタル化が普及した今日、より多くの人々がより多くのコンテンツに出会える素晴らしさがある一方で、コピー可能な受動的な体験の産業上の価値は今後徐々に失われてしまう心配がある。しかし、その人ならではの能動的な体験は、コピー不可能な体験となって価値がある。例えば、何かをクリエイト（創作）する創造的活動は、コピー不可能な体験である。創作結果を見てその創作過程をいくら想像しても、創作したクリエイターと同じ体験・気持ちを得ることはできない。

創作支援技術の発展により「一億総クリエイター時代」を目指すとはいっても、誰でも手軽に一からクオリティの高い音楽コンテンツを創作できるように支援することは容易でない。しかし、「コピー不可能な体験」を創出する観点からは、そこまで実現する必要は必ずしもない。たとえ簡単なカスタマイズ等であっても、その能動的な体験が他の人とは異なり、コピー不可能であれば十分に価値がある可能性がある。

さらに、能動的音楽鑑賞のような鑑賞支援技術も、その能動性が「コピー不可能な体験」を創出することにつながるから重要なのである。音楽再生時の可視化等により理解が深まる体験や、音楽加工によりカスタマイズする体験、音楽検索・ブラウジングの支援によりインタラクティブにコンテンツと出会う体験は、「コピー不可能な体験」の第一歩である。「ニコニコ動画」上でのコメント行為もこのような体験に位置付けられ、動画を受動的に視聴するだけでなく、コメントを能動的に入力した瞬間に、それがコピー不可能なその人だけの体験になる。

コンテンツのデジタル化は進んだが、その潜在的な可能性はまだ充分には引き出されておらず、従来は多量のコンテンツに受動的にアクセスできる量的な変化が中心であった。次の段階は能動的な「コピー不可能な体験」を情報技術の力で実現する質的な変化であり、それこそがデジタル化の本質である。その本質に迫る研究開発を進めなければ、そうした体験を価値の中核とした未来のクリエイティブ産業を創出し、牽引することはできない。このように本プロジェクトでは、「コピー不可能な体験」の価値創出が未来社会において本質であることを認識しながら、多様な創作・鑑賞支援技術を研究開発して学術的、産業的、社会的、文化的に貢献していくことを目指している。

## 参考文献

[1] 後藤 真孝, 吉井 和佳, 藤原 弘将, Matthias Mauch, 中野 倫靖: Songle: 音楽音響信号理解技術とユーザによる誤り訂正に基づく能動的音楽鑑賞サービス, 情報処理学会論文誌, Vol.54, No.4, pp.1363-1372, 2013.

[2] Masahiro Hamasaki, Masataka Goto, and Tomoyasu Nakano: Songrium: A Music Browsing Assistance Service with Interactive Visualization and Exploration of a Web of Music, Proc. WWW 2014.

[3] 後藤 真孝: 音楽音響信号理解に基づく能動的音楽鑑賞インタフェース, 情処研報 2007-MUS-70, 2007.

[4] 佐々木 将人, 吉井 和佳, 中野 倫靖, 後藤 真孝, 森島 繁生: LYRICS RADAR: 歌詞の潜在的意味分析に基づく歌詞検索インタフェース, 情処研報 2014-MUS-102, 2014.

[5] 中村 聡史, 山本 岳洋, 後藤 真孝, 濱崎 雅弘: 視聴者反応と音響特徴量に基づくサムネイル動画の生成手法, 情報処理学会論文誌 TOD, Vol.6, No.3, pp.148-158, 2013.

[6] Tomoyasu Nakano, Kazuyoshi Yoshii, and Masataka Goto: Vocal Timbre Analysis Using Latent Dirichlet Allocation and Cross-Gender Vocal Timbre Similarity, Proc. IEEE ICASSP 2014.

[7] 中村 友彦, 吉井 和佳, 後藤 真孝, 亀岡 弘和: 音楽音響信号中の調波音の周波数特性およびドラムの音色の置換システム, 情処研報 2014-MUS-104, 2014.

[8] 池宮 由楽, 糸山 克寿, 吉井 和佳, 奥乃 博: 混合音中の歌声 FO 軌跡に対する歌唱表現転写システム, 情処研報 2014-MUS-104, 2014.

[9] Matt McVicar, Satoru Fukayama, and Masataka Goto: AutoRhythmGuitar: Computer-aided Composition for Rhythm Guitar in the Tab Space, Proc. ICMC/SMC 2014.

[10] Matthew E. P. Davies, Philippe Hamel, Kazuyoshi Yoshii, and Masataka Goto: AutoMashUpper: An Automatic Multi-Song Mashup System, Proc. ISMIR 2013.

[11] 川名 勇氣, 宮下 芳明: 自由なテンポで演奏した複数の録音データから楽曲を生成するシステム, 情処研報 2014-HCI-157, 2014.

[12] 深山 覚, 後藤 真孝: MachineDancing: ダンス動作データの自動分析に基づく音楽に連動したダンス生成手法, 情処研報 2014-MUS-104, 2014.

[13] 加藤 淳, 中野 倫靖, 後藤 真孝: TextAlive: 音楽に同期した歌詞アニメーションの Kinetic Typography 制作環境, 情処研報 2014-MUS-104, 2014.

[14] 中村 聡史, 石川 直樹, 渡邊 恵太: 個人的な小さな幸せを実現するブラウザ上での動画編集・共有手法, WISS 2013 論文集.

[15] 中野 倫靖, 吉井 和佳, 後藤 真孝: 確率的生成モデルに基づく音楽の類似度とありがち度の推定に関する検討, 情処研報 2014-MUS-104, 2014.

[16] 増田 太郎, 吉井 和佳, 後藤 真孝, 森島 繁生: Query by Phrase: 半教師あり非負値行列因子分解を用いた音楽信号中のフレーズ検出, 情処研報 2014-MUS-102, 2014.

[17] Kazuyoshi Yoshii and Masataka Goto: Infinite Composite Autoregressive Models for Music Signal Analysis, Proc. ISMIR 2012.

[18] Kazuyoshi Yoshii, Ryota Tomioka, Daichi Mochihashi, and Masataka Goto: Infinite Positive Semidefinite Tensor Factorization for Source Separation of Mixture Signals, Proc. ICML 2013.

[19] 持橋 大地, 吉井 和佳, 後藤 真孝: ガウス過程に基づく連続空間トピックモデル, 情処研報 2013-NL-213, 2013.

[20] 小林 和弘, 土井 啓成, 戸田 智基, 中野 倫靖, 後藤 真孝, ニュービグ グラム, サクリアニ サクテイ, 中村 哲: 歌声の知覚年齢に沿った声質制御に向けた音響特徴量の調査, 情処研報 2013-MUS-99, 2013.

[21] 平井 辰典, 中野 倫靖, 後藤 真孝, 森島 繁生: 歌手映像と歌声の解析に基づく音楽動画中の歌唱シーン検出手法の検討, 情処研報 2014-MUS-103, 2014.

[22] 後藤 真孝: 未来を切り拓く音楽情報処理, 情処研報 2013-MUS-99, 2013. (招待講演)

 OngaCREST シンポジウム 2014

# 音楽情報処理研究が 切り拓く未来を探る

ポスター集

# コンテンツ共生社会のための 類似度を可視化する情報環境の実現



**研究代表者：後藤 真孝** (産業技術総合研究所 首席研究員)

音楽を中心とした類似度可視化情報環境の実現と全体統括



**研究分担者：森島 繁生** (早稲田大学 教授)

音楽連動動画を中心とした類似度可視化情報環境の実現



**研究分担者：中村 聡史** (明治大学 准教授)

Web インタラクションを中心とした類似度可視化情報環境の実現



**研究分担者：吉井 和佳** (京都大学 講師)

機械学習を中心とした類似度可視化情報環境の実現

本研究は、音楽や動画のようなメディアコンテンツを豊かで健全に創作・利用する「コンテンツ共生社会」に貢献できるように、膨大なメディアコンテンツ間の類似度を人々が知ることができる（可視化する）情報環境を実現する技術基盤の構築を目的とする。さらに、創作支援技術と鑑賞支援技術を研究開発することで、コンテンツの創作や鑑賞を人々が能動的に楽しめる社会や、過去のコンテンツに敬意を払う文化、感動体験重視型のコンテンツ文化の実現に寄与することを目指す。

## コンテンツが単調増加する社会

膨大なメディアコンテンツがデジタル化され  
適切な対価で自在にアクセス可能な社会になる

### 単調増加する過去の膨大なコンテンツの問題

- ・コンテンツの競争が激化(競争相手が単調増加)
- ・コンテンツの選択が困難

### コンテンツのオリジナリティが危機

- ・似ているコンテンツも単調増加して、本体は盗作ではないにも関わらず、盗作疑惑を招く事例が増えてしまう懸念がある
- ・あらゆる創作は既存コンテンツの影響を受けているのが自然
- ・類似度に関する人間の判断速度や「ありがち度」の判断には限界がある

コンテンツが単調増加することによって人々が安心して  
コンテンツの制作や発表をしにくい社会になりかねない

## 過去と未来の共生

### 過去のコンテンツと共存共栄し敬意を払う文化に貢献

- ・論文のようにリファーマーされ再利用されたら喜びを感じられる新音楽文化

### 感動体験重視型のコンテンツ文化の実現に寄与

- ・コンテンツの新規性のみを追求しても人々は幸せになれないのでは？  
「他にいかに類似していないか」に価値があるのではない
- ・コンテンツは本来、人々をいかに感動させ幸せにするかに価値がある  
感動させる魅力や完成度の高さ等があれば価値があることが重要

### デジタル化による「忘却できない社会」における共生という挑戦

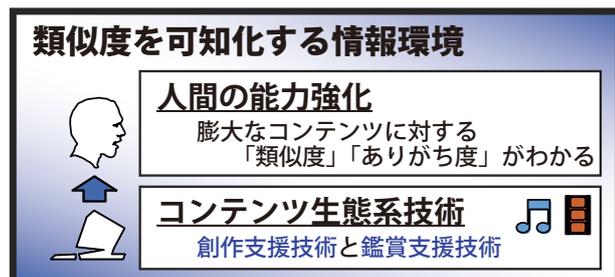
- ・デジタルコンテンツ社会ではもはや忘却できず  
単調増加する過去の膨大なコンテンツに押しつぶされかねない

### 豊かで持続発展可能な「コンテンツ共生社会」の実現を目指して

- ・過去のコンテンツと未来のコンテンツとの共生を人々が実感できるように
- ・人間とコンテンツとの共生により膨大なコンテンツを楽しめる社会に

専門家だけでなく一般の人々が「何が似ているのか」「どれくらいありふれているのか」を知ろうと思えば自在に把握して活用できる技術基盤を構築

膨大なコンテンツ間の類似度を「知ることができる」情報環境



メディアコンテンツを豊かで健全に創作・利用

### コンテンツ共生社会

- ・安心してコンテンツの制作や発表ができる社会
- ・誰もが気軽にコンテンツ創作を楽しめる社会
- ・主体的にコンテンツと出会って鑑賞できる社会

### 新しい価値

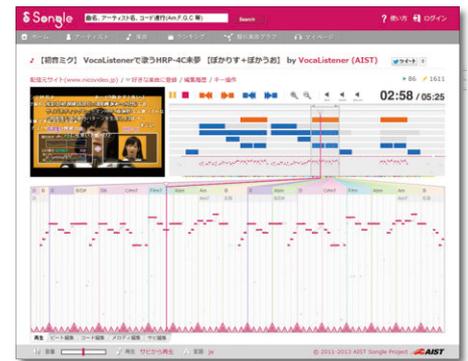
- ・過去のコンテンツに敬意を払う文化
- ・感動体験重視型のコンテンツ文化

対象とする主な時系列メディアコンテンツ：音楽（音楽音響信号等）、音楽動画（ミュージックビデオ、ダンス動画等）

## これまでの主要な研究成果

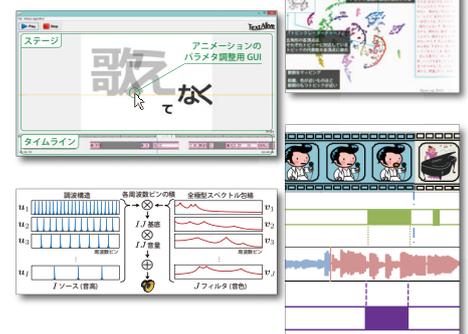
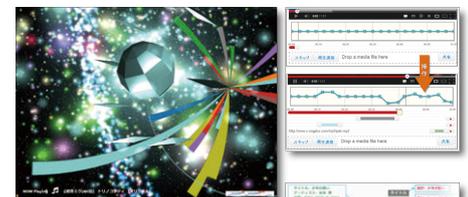
### 鑑賞支援技術

- 『Songle: Web 上の楽曲の中身を自動解析する能動的音楽鑑賞サービス』  
後藤 真孝, 吉井 和佳, 川崎 裕太, 井上 隆広, 中野 倫靖
- 『Songrium: 音楽コンテンツの関係性を可視化する音楽視聴支援サービス』  
濱崎 雅弘, 石田 啓介, 後藤 真孝, 中野 倫靖
- 『LyricsRadar: 歌詞の潜在的意味分析に基づく歌詞検索インターフェース』  
佐々木 将人, 吉井 和佳, 中野 倫靖, 後藤 真孝, 森島 繁生
- 『音楽理解とアノテーション分析技術によるサムネイル動画自動生成』  
中村 聡史, 山本 岳洋, 後藤 真孝, 濱崎 雅弘
- 『歌声トピックモデルに基づく類似歌声検索とトピック可視化』  
中野 倫靖, 吉井 和佳, 後藤 真孝
- 『音楽音響信号中の調波音の周波数特性およびドラムの音色の置換システム』  
中村 彦彦, 吉井 和佳, 後藤 真孝, 亀岡 弘和
- 『音楽音響信号中の歌声 F0 軌跡に対する歌唱表現の転写システム』  
池宮 由楽, 糸山 克寿, 吉井 和佳



### 創作支援技術

- 『AutoRhythmGuitar and AutoMashUpper』  
M. McVicar, M. Davies, P. Hamel, K. Yoshii, S. Fukayama, M. Goto
- 『自由なテンポで演奏した複数の演奏データから楽曲を生成するシステム』  
宮下 芳明, 川名 勇氣
- 『MachineDancing: 機械学習に基づく音楽に連動したダンスの自動生成』  
深山 覚, 後藤 真孝
- 『TextAlive: 音楽に同期した歌詞の Kinetic Typography 制作環境』  
加藤 淳, 中野 倫靖, 後藤 真孝
- 『VRMixer: 動画と現実の融合による新たなコンテンツの生成』  
平井 辰典, 中村 聡史, 森島 繁生, 湯村 翼
- 『N 次装飾: 動画共有サイト上の動画に対する装飾とその共有手法』  
中村 聡史, 石川 直樹, 渡邊 恵太



### 類似度・ありがち度の推定と音楽理解技術

- 『確率的生成モデルに基づく音楽の類似度とありがち度の推定』  
中野 倫靖, 吉井 和佳, 後藤 真孝
- 『半教師あり NMF を利用した音楽信号中のフレーズ検出』  
増田 太郎, 吉井 和佳, 後藤 真孝, 森島 繁生
- 『統計的機械学習に基づく音楽解析』  
吉井 和佳, 持橋 大地, 後藤 真孝
- 『統計的機械学習に基づく言語=音楽情報処理の横断的展開』  
持橋 大地, 吉井 和佳, 後藤 真孝
- 『歌唱音声の統計的知覚年齢制御』  
小林 和弘, 戸田 智基, 中野 倫靖, 後藤 真孝, G. Neubig, S. Sakti, 中村 哲
- 『歌手映像と歌声の解析に基づく音楽動画中の歌唱シーン検出』  
平井 辰典, 中野 倫靖, 後藤 真孝, 森島 繁生
- 『視聴者コメントに基づく楽曲動画の印象推定とデータセット構築』  
山本 岳洋, 中村 聡史

音楽の聴き方・創り方の未来を切り拓く技術開発により、音楽の楽しみ方がより能動的で豊かになる変化を日常生活に起こすことを目指す。

# Songle: Web 上の楽曲の中身を自動解析する能動的音楽鑑賞サービス

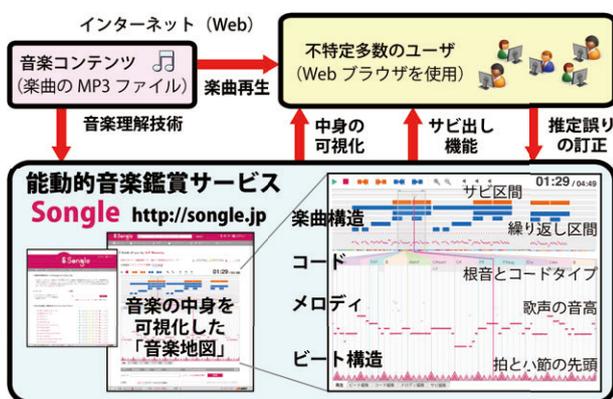
後藤 真孝 吉井 和佳\*1 川崎 裕太 井上 隆広 中野 倫靖

産業技術総合研究所 \*1 京都大学

## Songle (ソングル) とは

### 能動的音楽鑑賞サービス Songle (http://songle.jp)

- ・人々が音楽理解技術の力でポピュラー音楽をより深く理解して楽しめるようにするサービス
- ・4つの代表的な音楽的要素(サビ、ビート、メロディ、コード)を自動解析



- ・ニコニコ動画、YouTubeの音楽動画、ピアプロ、SoundCloud、MP3に対応  
75万曲以上を自動解析済み(閲覧時の音楽は元のサイト上で再生)

### 自動解析誤りをみんなで訂正してユーザ体験の向上に貢献可能

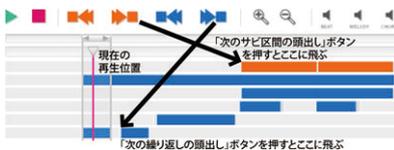
- ・候補選択等により誰でも誤り訂正して保存・共有可能なインタフェース



## Songleの様々な機能

### サビ出し機能

- ・楽曲中のサビ区間や繰り返し区間への頭出しが可能



### ビジュアライザ画面表示機能

- ・音楽的要素に密接に連動して表示内容が変化(音楽動画再生時は未対応)



### コード進行検索機能

- ・コード名の系列による音楽情報検索  
同一コード進行(音楽家共通のボキャブリ)をもつ複数の楽曲を聴き比べ

## Songle 外部埋め込みプレーヤ

### 外部のWebページ内に小型Songleプレーヤを埋め込み可能

- ・ユーザがホームページやブログなどでSongle上の楽曲を紹介できる
- ・楽曲構造の可視化機能とサビ出し機能付き



### Songle 外部埋め込みプレーヤのAPIを提供中

- ・プログラミングができれば誰でも音楽に連動した世界を実現可能に
- ・JavaScriptサンプルソースコード中に使用方法を例示

### 利用例: 音楽に連動したWebページの背景アニメーション

- ・背景に楽曲の再生に連動して変化する視覚的効果(音楽連動図形)を付与



### 利用例: 音楽に連動した照明制御(実世界デバイス制御)

- ・楽曲の再生中に音楽地図に連動してDMX512対応照明機器を制御可能



### 外部利用例: 初音ミク、鏡音リン等の音楽連動アニメーション

- ・クリプトン・フューチャー・メディア株式会社が2012年8月の初音ミク5周年記念、12月の鏡音リン・レン5周年記念で公開



### 外部利用例: Melvieで音楽に連動して動的に映像演出

- ・Web上の音楽を映像で彩るソーシャルVJサービスMelvie(http://melvie.jp)
- ・中嶋 誠氏(東京大学大学院)が中心となって産業技術総合研究所と共同で実現



### 外部利用例: V-Sidoで人型ロボットが音楽に連動してダンス

- ・人型ロボット制御ソフトウェアV-Sido(ブシドー)のSongle対応特別版
- ・音楽地図に基づいて動的に振り付けを変えながらロボットがダンス可能
- ・V-Sido開発者の吉崎航氏がSongle外部埋め込みプレーヤを活用し後藤真孝と連携して実現



## Songleの目指す世界

音楽コンテンツのデジタル化がもたらす真の価値を引き出す音楽に連動した世界を人々に届ける

謝辞: 櫻井 稔氏(Webデザインとビジュアライザ)、藤原 弘将氏、Matthias Mauch氏(音楽理解技術等)

# Songrium: 音楽コンテンツの関係性を可視化する音楽視聴支援サービス

濱崎 雅弘 石田 啓介 後藤 真孝 中野 倫靖

産業技術総合研究所

## Songrium(ソングリウム)とは

### 音楽の関係性を利用した音楽視聴支援サービス

- 多様な関係性を意識しながら、Web上で新しい音楽コンテンツに出会うことができるサービス
- Webマイニングによる音楽コンテンツの自動収集・分類、音楽情報処理技術による音楽コンテンツの自動解析、それらのデータを用いた様々な可視化インタフェース



## Web-native Music

### Webで発表、共有、視聴され、派生作品が生まれる音楽

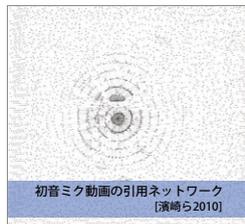
- すべてがWeb上にあるため誰でも聴ける、統計や機械処理に向く
- ニコニコ動画上で広がりを見せる VOCALOID 音楽はその最先端事例

### ニコニコ動画上の VOCALOID オリジナル楽曲は 11 万曲以上、その派生動画は 53 万動画以上



### 派生関係が巨大なネットワークを構築

- 新しいコンテンツを生み出す原動力に音楽のオープンコラボレーション
- コンテンツだけでなく関係性も生まれる  
歌ってみた、踊ってみた、演奏してみた、描いてみた、MMDをつけてみた
- 一方で膨大かつ多様になり過ぎて、全体が見渡しにくくなっている問題も



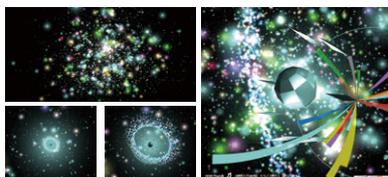
### Web-native Musicのための音楽視聴インタフェースの必要性

➡ 関係性を可視化する音楽視聴支援サービス Songrium

#### 新機能

### Songrium 3D

音楽星図を三次元 CG で表現。一人称視点での可視化により、コンテンツ表示と全体表示を両立。動画再生時にはビートや楽曲構造と連動した視覚的演出が表示される。



## Songriumの様々な機能

### 音楽星図

音響特徴量の類似関係をもとにオリジナル楽曲を平面配置。マウスによるパン・ズーム操作で 11 万曲の楽曲を俯瞰できる。また、フィルター機能である条件を満たした楽曲だけを連続再生したり、サビ出し機能でサビから再生したりできる。



### 惑星ビュー

派生動画の様子が俯瞰でき、さらにそこからオリジナル楽曲の特徴も知ることができる。



### 矢印タグ

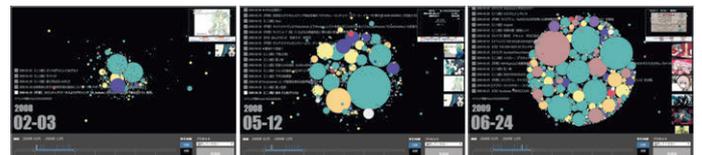
音楽コンテンツ間の関係に名前を付ける、新しいソーシャルギンギン。ユーザはタグ付けされた楽曲間の関係(つながり)をたどることで、様々な楽曲と出会う。

### YouTube バージョン

YouTubeでも音楽星図を利用できる。国内外のプロアーティストやアマチュアミュージシャンのミュージックビデオが楽しめる。

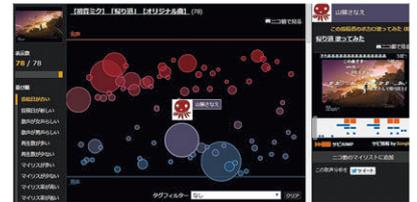
### バブルプレーヤ

指定期間に投稿された楽曲群を連続サビ再生するとともに、音楽コミュニティの成長過程として、楽曲群が続々と投稿されていく様子を可視化する。



### 歌声分析

音楽動画中の歌声の音響的特徴から男女度(男声・女声らしさを示す値)を自動推定し、その結果を可視化する。楽曲を聴く際に「歌声を選ぶ」という新しい視聴体験が可能に。



### ブラウザ拡張

ニコニコ動画や YouTube 上での音楽鑑賞時に Songrium の各機能を利用できる、ブラウザ用プラグイン。(現在は Google Chrome のみに対応)



## Songriumの目指す世界

- 音楽コンテンツの膨大さ・多様さを楽しむ視聴体験
- 音楽のつながりが可視化され活用できる「音楽の Web」の実現

# 確率的生成モデルに基づく音楽の類似度とありがち度の推定

中野 倫靖 吉井 和佳\*1 後藤 真孝

産業技術総合研究所 \*1 京都大学

## 研究背景

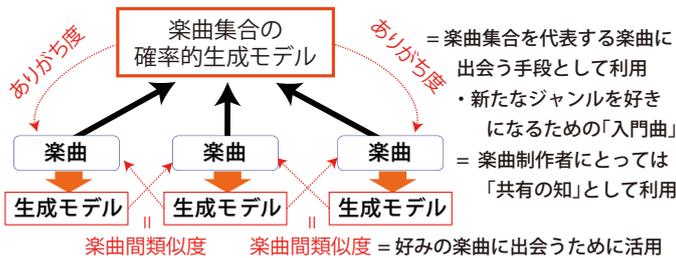
### 音楽がデジタル化されアクセス可能な楽曲が単調増加

- ・人々が音楽の「何が似ているのか」「どれくらいありふれているのか」を知ろうと思った時に容易に知るための手段の実現
- ・過去の楽曲と共存共栄し、敬意を払う文化を築くことへの貢献  
例) 論文のように引用され再利用されたら喜びを感じられる音楽文化  
カバー曲の制作やニコニコ動画でのN次創作においては引用が一般的新規楽曲を発表する際に他の楽曲への引用を記述することは稀

## 音楽要素の確率的生成モデル

### 歌声と伴奏を含む音楽音響信号の音楽要素の生成モデル

- ・各音楽要素(音響特徴量や和音進行)がどのような形で出現しやすいかその確率(生成確率)を計算できるモデル
- ・生成確率によって「楽曲間の類似度」や「楽曲のありがち度」を推定する



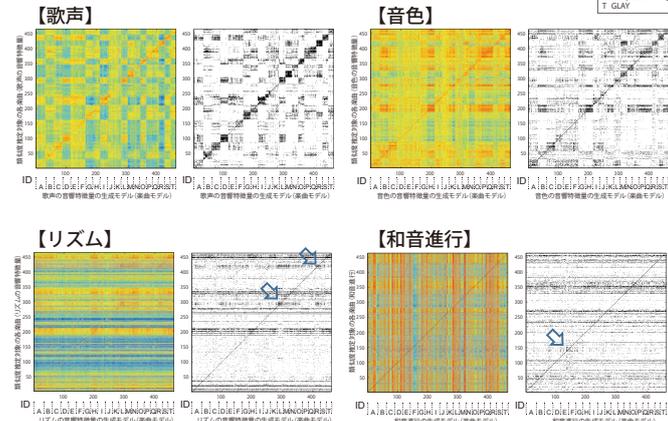
※現時点では、音楽要素を生成したり楽曲を作ったりすることはできない(将来的には生成できるよう発展できる可能性がある)

## 実験 A: 類似度

### 楽曲毎の生成モデルをポピュラー音楽 3278 曲で学習

- ・オリコン上位 20 位以内(2000-2008)
- ・楽曲数が多い上位 20 アーティストの楽曲 463 曲

- A 浜崎あゆみ
- B 嵐
- C ミュージック
- D 水木南衣
- E 椎名林檎
- F BoA
- G 嵐
- H L'Arc~en~Ciel
- I 嵐
- J 嵐
- K SORRY
- L 嵐
- M CHEMISTRY
- N GACKT
- O GARNET CROW
- P TOKIO
- Q 嵐
- R 平井堅
- S Every Little Thing
- T GLAY



### 作品の発表時に盗作疑惑を招く事例が増えてしまう懸念

- ・類似度に関する人間の能力の限界に起因  
楽曲全てを聞いて全体を俯瞰した適切な判断を行うことは不可能
- ・あらゆる楽曲は既存曲の影響を受けている  
「無自覚に」「何らかの意味で」「部分的に」類似してしまうのは自然
- ・自分の作品が何かに似ていると糾弾されるリスクが高いと  
安心して楽曲の制作や発表をしにくい社会になりにかねない

### 過去の楽曲に敬意を払う文化、感動体験重視型の音楽文化へ

- ・「他に類似していないか」という新規性だけを追求するのではなく過去の楽曲と共存共栄し、人々を感動させる魅力や完成度の高さ等を重視

## 分析対象と生成モデル

### ボーカルの歌声

- ・線形予測メルケプストラム係数 (LPMCC)、 $\Delta F0$
- ・[生成モデル] 潜在的ディリクレ配分法 (Latent Dirichlet Allocation: LDA)

### 楽曲中の音色

- ・メル周波数ケプストラム係数 (MFCC)、 $\Delta MFCC$ 、 $\Delta$  パワー
- ・[生成モデル] 潜在的ディリクレ配分法 (Latent Dirichlet Allocation: LDA)

### リズム

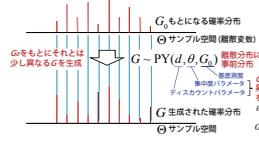
- ・Fluctuation Pattern (FP)
- ・[生成モデル] 潜在的ディリクレ配分法 (Latent Dirichlet Allocation: LDA)

### 和音進行

- ・8 種類の代表的な和音とその 12 種類の根音 (+和音がない区間)  
= major, major 6th, major 7th, dominant 7th, minor, minor 7th, diminished, augmented
- ・[生成モデル] 可変長 Pitman-Yor 言語モデル (VPYLM)

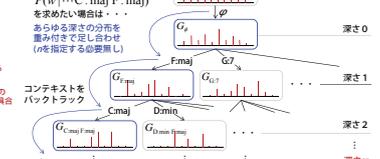
### Pitman-Yor過程

- ・PY: 確率分布上の確率分布  
- ティリクレ過程 (DP) の一般化  
- ある確率分布  $G$  に対する事前分布として利用可能  
例:  $G$  が離散分布の場合



### 可変長 Pitman-Yor 言語モデル (VPYLM)

- 和音ごとに異なるコンテキスト長を許容  
・無限の深さを持つ木の生成モデル



## 実験 B: ありがち度

### RWC 研究用音楽データベース (ポピュラー音楽) 100 曲を推定

- ・実験 A で学習したポピュラー音楽 3278 曲の生成モデル(楽曲集合のモデル)

【歌声】	【音色】	【リズム】	【和音進行】	含まれていた和音進行
1位 No.60 飯島柚子	No.15 小澤克之	No.6 オリゲン	No.56 橋本まさし	Cmaj Gmaj Fmaj Cmaj
2位 No.70 松坂珠子	No.90 井口慎也	No.81 ドナ・バーク	No.82 井口慎也	Cmaj Fmaj Gmaj Cmaj
3位 No.45 森元康介	No.99 井口慎也	No.29 西一男	No.41 小澤克之	Amin Fmaj Gmaj Cmaj
4位 No.20 市川えり	No.55 鐘木朗子	No.8 フィーバース	No.84 井口慎也	Cmaj Fmaj Cmaj
5位 No.42 森元康介	No.73 西一男	No.60 M&Y	No.54 凜	Cmaj Fmaj Gmaj Cmaj Gmaj Fmaj Gmaj Fmaj

## 今後の展望

- 音楽要素毎の類似度やありがち度の統合
- 類似度やありがち度を活用した鑑賞支援・創作支援技術

# 歌声トピックモデルに基づく 類似歌声検索とトピック可視化

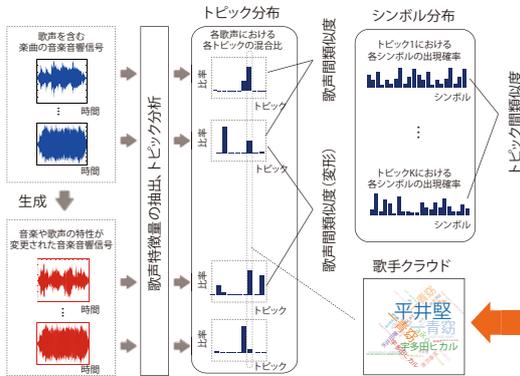
中野 倫靖    吉井 和佳\*1    後藤 真孝

産業技術総合研究所    \*1 京都大学

## 歌声トピックモデル

音楽音響信号からボーカルの歌声を定量的にモデル化

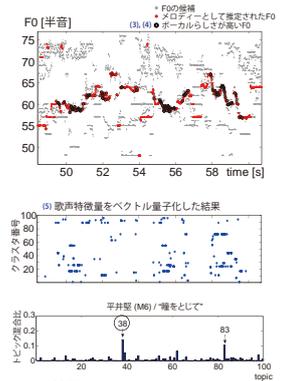
- Latent Dirichlet Allocation (LDA) に基づくトピック分析
- 複数の歌声から潜在的な意味(トピック)を学習
  - [従来]一つの歌声からの特性分析、モデル化
- 「何と何が似ているか」だけでなく、「どう似ているか」まで分析可能



手法：特徴抽出・モデル化・類似度計算

- 「何と何が似ているか」を推定する

- (1) メロディーの F0 推定: PreFest
- (2) 再合成: 20 次倍音までの正弦波合成
- (3) 歌声特徴の抽出: LPMCC (12),  $\Delta F0$  (1)
- (4) 高信頼度フレーム選択: 歌声・非歌声 GMM
- (5) ベクトル量子化: k-means 法 (k = 100)
- (6) LDA 学習:
  - 周辺化 Gibbs サンプラー (トピック数 100)
  - 事前分布 (ディリクレ分布のハイパーパラメータ)
  - $\alpha$  (トピック混合比): 1     $\beta$  (シンボル混合比): 0.1
- (7) 類似度計算: 対称カルバック・ライブラ距離 (KL2) の逆数



## 【音高シフト】ボーカル声質の変形

- 検索対象を変形して探索空間を拡張する
  - [従来] 既にあるものから検索する
- 性別を超えた歌声情報検索が可能となる
  - 例) 「一青窈の歌声を 3 半音下げると平井堅に似ている」を自動推定

## 【歌手クラウド】歌声トピックの意味を歌手名で可視化

- 「どう似ているか」を分析する
- 各トピックを代表する歌手名を大きく表示
- トピックの意味をイメージしやすくなることができる



## 評価実験

ポピュラー音楽で歌声検索の性能を評価

- ボーカルが 1 名である 12 アーティストの楽曲で評価
- オリコン上位 20 位以内 (2000-2008)

男性歌手

- M1 ASIAN KUNG-FU GENERATION
- M2 BUMP OF CHICKEN
- M3 福山雅治
- M4 GLAY
- M5 氷川きよし
- M6 平井堅

女性歌手

- F1 aiko
- F2 JUDY AND MARY
- F3 一青窈
- F4 東京事变
- F5 宇多田ヒカル
- F6 矢井田瞳

本手法の有効性を確認

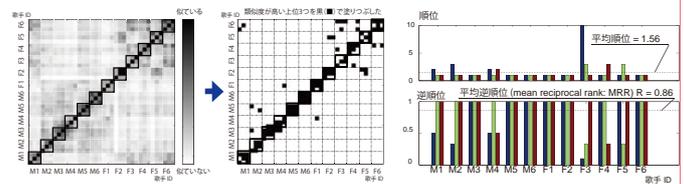
- トピックの混合比は、歌手名同定とトピックの意味の可視化に適用できる
- 類似した歌声を性別非依存に検索可能
  - 「一青窈を 3 半音下げると平井堅に似ている」ことを自動推定して実証

## 今後の展望

- トピック分析結果の定量的な分析: 歌声を特徴付ける要素
- F0 軌跡の変化等の歌い方のモデル化: 歌い方特徴量の追加
- 手法の拡張: 離散化せずにモデル化、トピック数の無限化

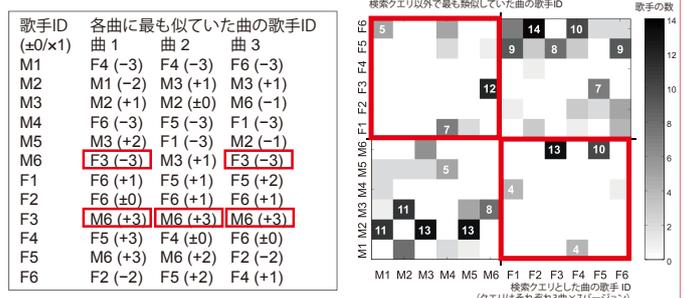
## 実験 A: 歌手名同定

- 36 曲 = 歌手 12 人 (男性 6 人、女性 6 人) × 各 3 曲



## 実験 B: 性別を超えた類似歌声検索

- 音高シフト: -3 半音 ~ +3 半音 (7 種類)
- 252 曲 = 12 人 × 3 曲 × 7 種類



# MachineDancing: 機械学習に基づく音楽に連動したダンスの自動生成

深山 覚 後藤 真孝

産業技術総合研究所

## MachineDancing とは

音楽に連動した3次元CGキャラクターのダンスを自動生成

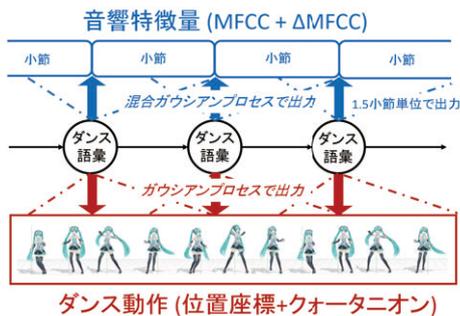
- ・ダンス制作にかかる多大な時間や労力などの困難を克服
- ・人手では限界のある、状況に応じた多様な動作の動的な生成を実現

楽曲とダンスが対応付いたデータからダンス動作を学習し生成

- ・ダンス断片の切り貼りでないため楽曲に応じて新たな動作を生成可能
- ・学習データを変えれば、それに伴って異なったダンスを生成

「ダンス語彙」の概念を考案しダンス動作の確率モデルを構築

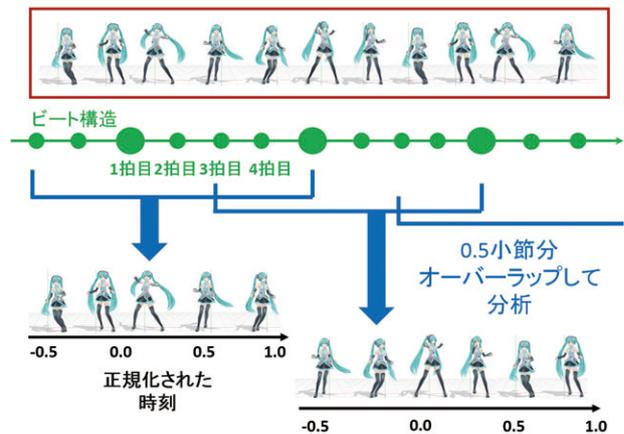
- ・「ダンス語彙」: 類似した特徴を持つダンス動作の確率的な生成源
- ・ダンス語彙を隠れ状態とする隠れマルコフモデルでダンス動作を学習
- ・ダンス動作(位置座標+クォータニオン)と音響特徴量(MFCC+ΔMFCC)のダンス語彙との関係を(混合)ガウシアンプロセスでモデル化



## ダンス動作の機械学習

ダンス動作をビート構造に基づいた分析区間に区分して分析

- ・長さ1.5小節ごとに0.5小節分オーバーラップさせながら逐次的にダンス動作を区分することで、区間の接続部分を含めてダンス動作を分析
- ・楽曲間で異なるテンポに対応するため1小節の長さを基準に時刻を正規化



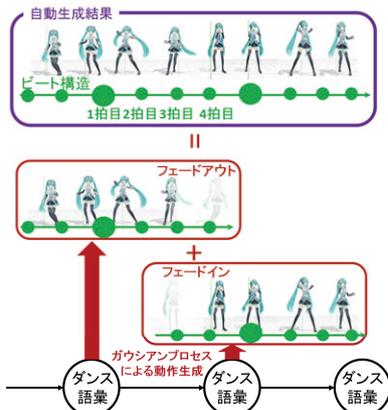
確率モデルによってクラスタリングを行いダンス語彙を学習

- ・ガウシアンプロセスに基づいて、共通した特徴をもつダンス動作同士を集めその特徴を持つ動作を確率的に生成できるダンス語彙を学習
- ・連続時間でダンス動作を出力できる確率モデルとすることで、分析区間中のダンス動作の異なるデータポイント数に対応可能

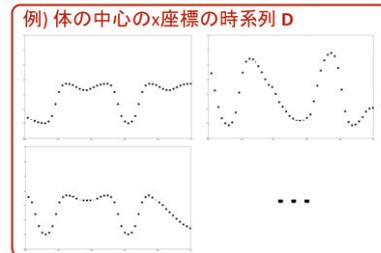
## 音楽と連動したダンス生成

入力した楽曲の音響特徴量のもとで確率的にダンス動作を生成

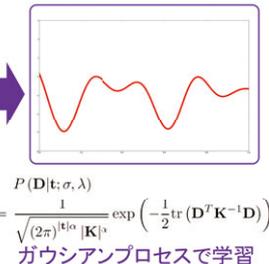
- ・学習した確率モデルに基づいて尤度最大のダンス語彙の系列を探索
- ・求めたダンス語彙の列からガウシアンプロセスにより動作を生成
- ・ダンス語彙から生成されたダンス同士をそれぞれ3-4拍目で動作を補間し1曲を通じて滑らかに連なる動作を生成
- ・同じ楽曲でも確率モデルの分散に基づき異なったダンス動作を生成可能



類似するダンス動作のクラスタ



学習されたダンス語彙



## MachineDancing の今後

インタラクティブにダンスを制作できるインタフェース構築

- ・本研究のダンス自動生成手法を応用してダンス動画を手軽に制作可能に

楽曲のより深い理解を反映したダンスの自動生成

- ・音楽理解技術を活用しながら新たな音響特徴量や機械学習手法を検討して楽曲の盛り上がりに関連し構成をもつダンスを自動生成

# TextAlive: 音楽に同期した歌詞の Kinetic Typography 制作環境

加藤 淳      中野 倫靖      後藤 真孝  
産業技術総合研究所

## Kinetic Typography とは

- 文字情報を動画としてアニメーションさせ、視聴者の印象に残す表現手法
- 印刷物上で文字を綺麗に配置する **Typography の動画版**
- 研究では**コミュニケーションツール向けの応用**が多い  
例) インスタントメッセージ、携帯電話の SMS など
- 商業的には**テレビや映画のタイトルなどで多く実用**されている  
例) テレビ番組のタイトルやテロップがアニメーションする動画

音楽動画で Kinetic Typography を使えば、歌詞に込められたメッセージがより伝わる

a) 単純なカラオケ表示

歌えなくて

b) Kinetic Typography (提案システムによる出力)



## Kinetic Typography の難しさ

### ① 歌声と文字情報の高度な同期

- 歌詞は、歌声の発声開始～終了の間だけ表示されているわけではない
- 歌詞中の文字、単語、フレーズそれぞれに画面上で目立つ瞬間がある
- 通常は**全てのタイミングを手作業で設計する必要がある**、手間がかかる

### ② 空間的な意味を持つパラメタ調整

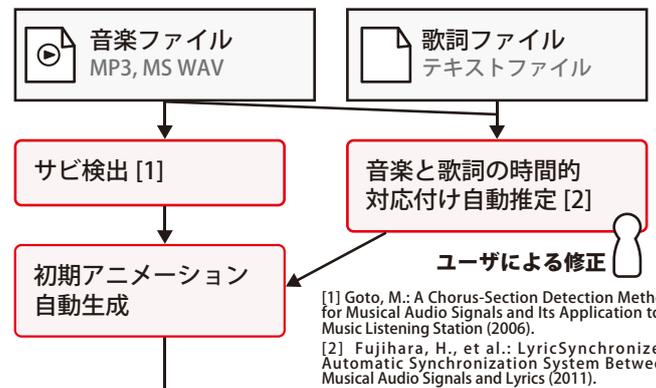
- Kinetic Typography にはいくつも典型的なパターンがあり、個別に見れば調整すべきパラメタの数は多くなく、しかも空間的な意味を持っている  
例) 初期位置、アニメーションの向き、文字を敷き詰める領域の形状など
- 通常はキーフレームでの位置など、**個別具体的な調整を行う必要がある**

## TextAlive (テキストアライブ) とは

- 音楽に同期した歌詞アニメーションの **Kinetic Typography 制作に特化**
- Kinetic Typography を制作する上での難しさを解決する **2 手法を実装**

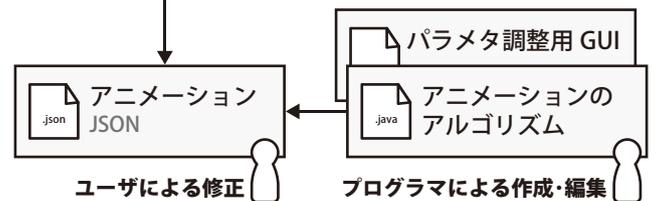
### ① 音楽+歌詞=Kinetic Typography 自動生成

- 音楽情報処理技術により、歌詞の Kinetic Typography に必要なパラメタを全て推定して、**初期アニメーションを自動生成**
- 歌詞表示のタイミング推定が誤っていても**インタラクティブに修正**できる



### ② Kinetic Typography の直感的な編集用 GUI

- ユーザは**アニメーションをステージ上の GUI で直感的に調整**できる
- プログラマは**アルゴリズムとパラメタ調整用 GUI を容易に実装**できる



Kinetic Typography に特化した GUI でアニメーションを自分好みにできる



内蔵ソースコードエディタで書いたコメントを元に調整用 GUI が生成される

```

1  1|/**
2  2| * Line height
3  3| * @ui Slider(0, stageHeight)
4  4| */
5  5| public float lineHeight = 50;
6  6|
7  7|
8  8|
9  9|
10 10|/**
11 11| * Horizontal offset
12 12| * @ui MoveHandle(stageWidth - offsetX, stageHeight - 20, x)
13 13| */
14 14| public int offsetX = 30;
15 15| class textalive.autoui.MoveHandle

```

歌詞をより魅力的に伝える 動画表現の大衆化

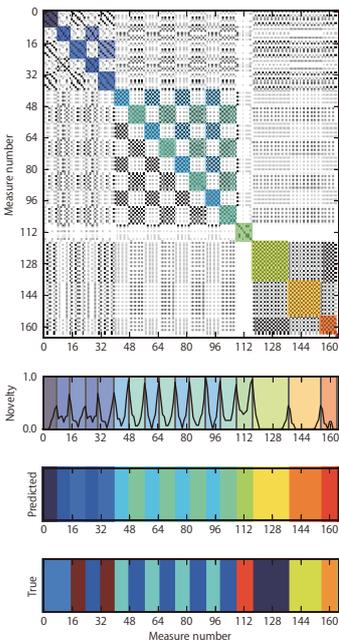
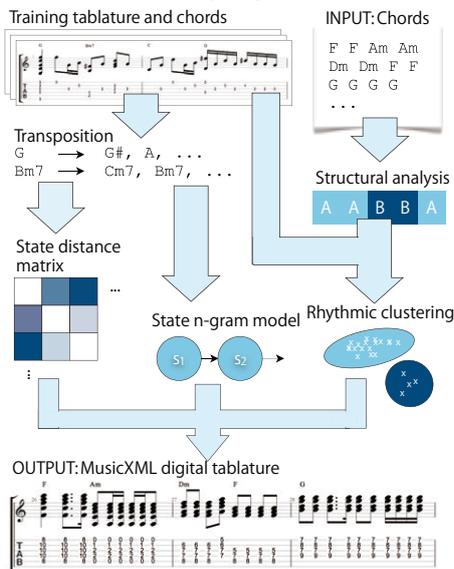
# AutoRhythmGuitar and AutoMashUpper

M. McVicar M. Davies P. Hamel K. Yoshii\*1 S. Fukayama M. Goto

National Institute of Advanced Industrial Science and Technology (AIST) \*1Kyoto University

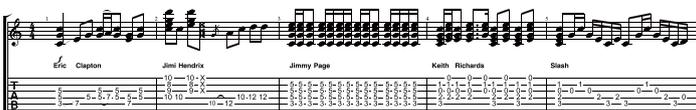
## AutoRhythmGuitar

- AutoRhythmGuitar automatically composes rhythm guitar parts in tablature notation from an input chord sequence.
- The sequence is automatically segmented into sections.



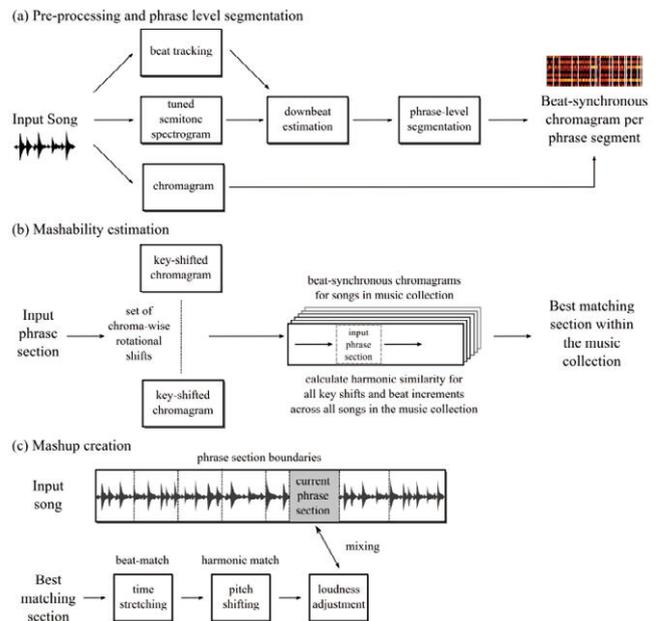
- Training rhythms are then clustered into an appropriate number of clusters based on the structural segmentation.
- n-gram models are then built for each chord type (major, minor, dominant 7th, ...) directly in the tablature space.
- Tablature is then produced by a random walk over the model, biased by the distance between states to increase tab playability.

Below: example output in the style of 5 different guitarists over a C major chord.

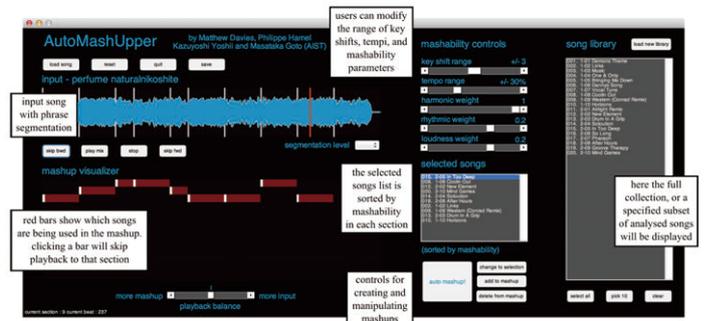


## AutoMashUpper

- AutoMashUpper creates “mashups” of existing songs, creating new compositions without the need for musical training.
- Beat-synchronous chromagrams per phrase describe the evolution of the pitch classes in a particular song.
- The “mashability” between an input song and each song in a database is then calculated.



- Mashability is calculated not only from rhythmic and harmonic similarity but also spectral balance.



AutoMashUpper has four modes of operation:

1. **Album/artist mode** - mashes with a fixed album or artist.
2. **Style mode** - mashes restricted to genre: J-pop, rock, etc.
3. **Forced mashup mode** - mashes up two given songs.
4. **Musician mode** - uses isolated recordings such as vocals or bass guitar and mashes to existing songs.

# 歌唱音声の統計的知覚年齢制御

小林和弘 戸田智基 中野倫靖\*1 後藤真孝\*1 G. Neubig S. Sakti 中村哲

奈良先端科学技術大学院大学 \*1 産業技術総合研究所

## 歌唱音声の声質制御

統計的手法に基づく声質変換 [Toda et al., 2007]

歌手の身体的制約を超えた声質での歌唱を実現



少量の適応データを用いて任意の歌手への声質変換を実現 [Doi et al., 2012]

研究目的

歌手が自身の声質を自在に制御できる技術の実現



発表内容

個人性を保持した知覚年齢に基づく声質制御



## 知覚年齢に基づく声質制御法

重回帰GMMに基づく歌声声質変換 (MR-GMM)

$$P(X_t, Y_t | \lambda^{(MR)}, w^{(s)}) = \sum_{m=1}^M \alpha_m \mathcal{N} \left( \begin{bmatrix} X_t \\ Y_t \end{bmatrix} ; \begin{bmatrix} \mu_m^{(X)} \\ \mu_m^{(Y)}(s) \end{bmatrix}, \begin{bmatrix} \Sigma_m^{(XX)} & \Sigma_m^{(XY)} \\ \Sigma_m^{(YX)} & \Sigma_m^{(YY)} \end{bmatrix} \right)$$

$$\mu_m^{(Y)}(s) = \mathbf{b}_m^{(Y)} w^{(s)} + \bar{\mu}_m^{(Y)}$$

出力歌手の平均ベクトル  $\mu_m^{(Y)}(s)$  = 代表ベクトル  $\mathbf{b}_m^{(Y)}$  × 知覚年齢スコア  $w^{(s)}$  + バイアスベクトル  $\bar{\mu}_m^{(Y)}$

知覚年齢スコア  $w^{(s)}$  により出力平均ベクトルが決定

問題点：特定の歌手の個人性を持つ声質を表さない

個人性を保持した重回帰GMMに基づく歌声声質変換 (Modified MR-GMM)

$$\mu_m^{(Y)}(s) = \mathbf{b}_m^{(Y)} w^{(s)} + \bar{\mu}_m^{(Y)}$$

出力平均ベクトル  $\mu_m^{(Y)}(s)$  = 入力歌手の知覚年齢  $w^{(i)}$  による  $\mathbf{b}_m^{(Y)} w^{(i)} + \bar{\mu}_m^{(Y)}$  +  $\Delta w$  の差分

$$= \mathbf{b}_m^{(Y)} w^{(i)} + \bar{\mu}_m^{(Y)} + \mathbf{b}_m^{(Y)} \Delta w$$

入力歌手の特定モデルへと変更

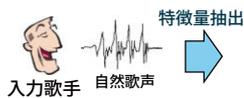
$$\mu_m^{(Y)}(s) \simeq \hat{\mu}_m^{(Y)} + \mathbf{b}_m^{(Y)} \Delta w$$

修正した出力平均ベクトル = 入力歌手の特定モデル +  $\Delta w$  歳の年齢変動

歌手の知覚年齢を基準に差分知覚年齢スコアに基づいて年齢操作

## 知覚年齢に寄与する音響特徴量

音響特徴量

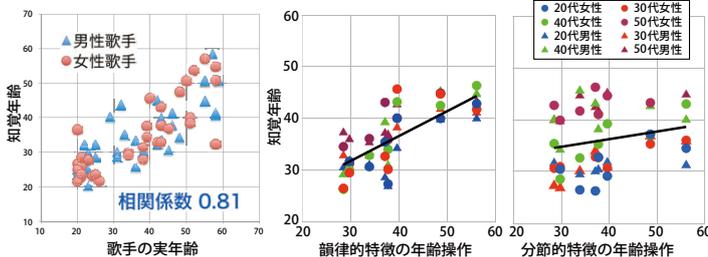


韻律的特徴

- 基本周波数
- パワー

分節的特徴

- スペクトル包絡
- 非周期成分



20代男性8人による知覚年齢推定実験

知覚年齢に関する調査結果

歌声の知覚年齢と歌手の実年齢に高い相関

音響特徴量の寄与調査

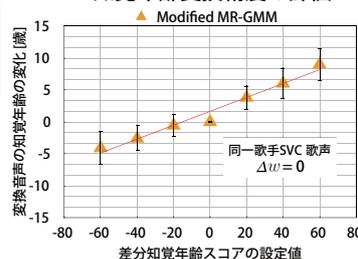
特徴量操作をした場合

	知覚年齢の寄与	歌手の個人性	知覚年齢の変化	歌手の個人性
韻律的特徴	大きい	多い	大きい	失う
分節的特徴	小さい	少ない	小さい	保持される

分節的特徴の操作による知覚年齢制御の実現

## 実験的評価

知覚年齢変換精度の評価

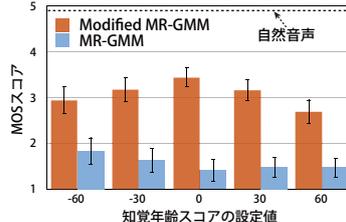


知覚年齢制御の実現

実験環境

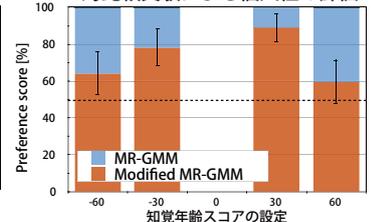
歌声データベース	AIST/ハミングデータベース
メルケプストラム, 非周期成分	24次元 (1st-25th), 5周波数帯
事前収録歌手	54名 (20代~50代)
評価歌手	各年代の男女16名 (オープンテスト)
学習データ	25曲/人
被験者	20代男性8名
評価歌手の組合せ	各年代, 性別が含まれる2セットに分割

5段階MOSによる自然性の評価



自然性の高い変換音声

対比較実験による個人性の評価



個人性保持の実現

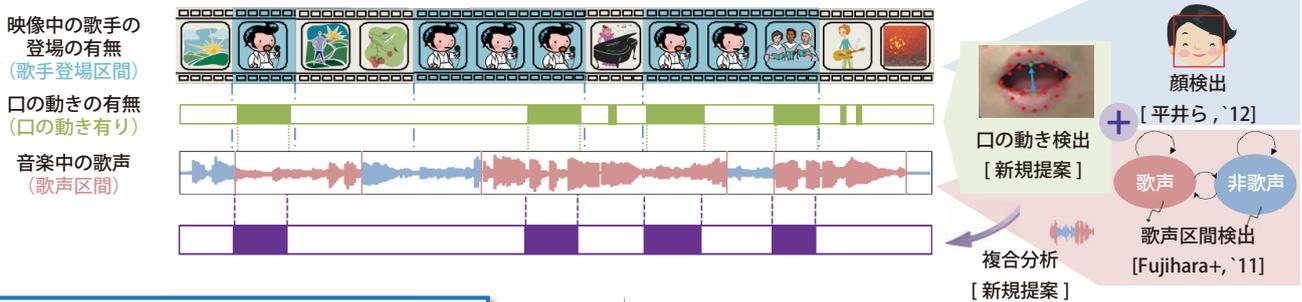
# 歌手映像と歌声の解析に基づく 音楽動画中の歌唱シーン検出

平井 辰典 中野 倫靖\*1 後藤 真孝\*1 森島 繁生

早稲田大学 \*1 産業技術総合研究所

## 目的

音楽動画中の歌唱シーンを音と映像の分析結果の統合によって自動検出



## 1. はじめに

### 研究背景

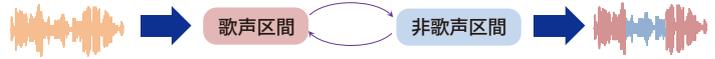
動画共有サービスにおける音楽動画の人気  
→YouTubeの歴代再生回数上位 30 作品中 29 作品が音楽動画  
そのうち 26 作品がソロ歌手による Music clip / ライヴ(歌唱)動画

ポピュラー音楽において歌手は中心的な役割を担う  
歌手に注目して動画を検索 / 鑑賞する視聴者は多いはず



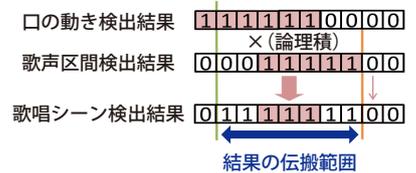
そこで → 音楽動画中の歌唱シーンを自動検出する手法を検討

3. 歌声区間検出手法 (Vocal Activity Detection : VAD) (Fujihara et al. '11)  
歌声区間と非歌声区間の間の状態遷移を HMM によりモデル化



4. 検出結果の統合

口の動き検出と  
歌声区間検出  
の結果を統合  
(結果の時間連続性を  
仮定して結果を伝搬)



## 2. 歌唱シーン検出手法

### 用語の定義

- ① 歌手登場区間: 映像中で、歌唱の有無にかかわらず歌手が登場する区間
- ② 歌声区間: 音楽中に歌声が含まれている区間(コーラスも含む)
- ③ 歌唱シーン: 映像中で歌手が歌っており、対応する歌声が聞こえる区間

### 各検出手法

1. 顔検出手法(平井ら, '12)

階層的 ASAM + 動画フレームの時間連続性に基づく顔検出手法 (Irie et al. '11)



2. 口の動き検出法 (Mouth Aperture Detection : MAD)

唇の特徴点間の距離に基づく口の動きの推定



(同一ショット内の口の動きの標準偏差が一定以上なら歌唱していると推定)

## 3. 歌唱シーン検出精度

### 実験条件

実験動画: Music clip 9 作品 + 演奏動画 1 作品

### 実験結果

	歌唱シーン検出精度		
	適合率	再現率	F 値
顔検出	0.570	0.869	0.672
口の動き	0.609	0.823	0.677
歌声区間	0.632	0.732	0.672
統合検出	0.662	0.759	0.690

統合検出の精度が最も高かった。

適合率・再現率の一方のみを向上させる複合も可能。

## 4. まとめ

歌手映像と歌声の解析に基づく音楽動画中の歌唱シーン検出手法を検討した。今後、より精度が向上する複合方法の検討と共に、歌唱シーン以外の音楽動画中のイベント認識手法へと拡張。

# LyricsRadar: 歌詞の潜在的意味分析に基づく歌詞検索インターフェース

佐々木 将人 吉井 和佳\*1 中野 倫靖\*2 後藤 真孝\*2 森島 繁生

早稲田大学 \*1 京都大学 \*2 産業技術総合研究所

## 研究背景と目的

歌詞は重要 → 歌詞のもつ意味に基づいた検索

森, 日常の音楽聴取における歌詞の役割についての研究, 対人社会心理学研究, 10, pp. 131-137, (2010).

## 従来システムとその問題点

- 歌詞中に登場する語句を表層的にテキスト全文検索

例. 涙

「涙」という単語をクエリとして入力すると、失恋の「涙」と喜びの「涙」が同時に表示される



ユーザの検索意図を十分反映できない

Uta-Net (<http://www.utanet.com/>) 検索画面より

## 提案手法

### 歌詞のもつトピックを利用

- 潜在的ディリクレ配分法 (Latent Dirichlet Allocation) により歌詞のトピックを推定する
- 歌詞のトピック推定における問題点
  - 一つの歌詞に登場する単語数が少ない
  - トピック推定は文書の単語数に依存



単に歌詞を文書とみなしてもトピックを正確に推定することは困難

### 事前処理 (歌詞のテキスト解析)

- 歌詞 例. 走れあの夕日に向かって
- 形態素解析 例. 走れ/あの/夕日/に/向/か/っ/て (MeCab を利用)
- 名詞・動詞・形容詞の抽出 例. 走れ/夕日/向/か/っ
- 原形処理 例. 走る/夕日/向/か/う → idf による重み付け

### 1. 歌詞の選別とトピック推定 100語以上の歌詞のみを扱う

ポピュラー音楽 (J-POP) 21845 曲 → 100語以上 → 6902 曲 → LDA

### 2. アーティスト毎にトピック推定

- (1) 各アーティストの全歌詞をまとめて一つの歌詞とみなす
- (2) 100語以上の歌詞のみを扱う
  - 21845 歌詞
  - 3755 アーティスト
  - 2848 文書 (2848 アーティスト)



## LyricsRadar

歌詞の潜在的な意味をクエリとして 検索を行うインターフェース



### 1. トピックレーダーチャート

- 各歌詞の潜在的なトピックの比率を五角形内に着色して可視化
- トピック数は5 (操作性とのトレードオフからバランスを見極めて5に決定)

### 2. 二次平面へのマッピング

- すべての歌詞をそのトピックの類似度に応じて二次元平面上にマッピング
- ユーザがインタラクティブに歌詞を探ることができる機能



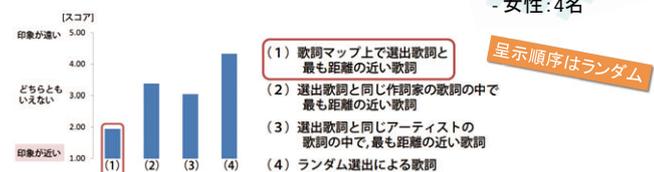
### 3. トピックレーダーチャートへの直接入力機能

- トピック比率の5個の値を図形として直接変形操作することでそれをクエリとしてトピック比率が最も近い歌詞を検索できる機能



## 評価

- ランダムに選出した歌詞1曲(以降, 選出歌詞)に対し 被験者 20代: 21名
  - 下記の4曲の歌詞(以降, 比較歌詞)と比較し - 男性: 17名
  - 二つの歌詞から受ける印象を五段階で評価 - 女性: 4名



## まとめ

歌詞の潜在的なトピックの比率に基づいて, 歌詞を検索できるインターフェース LyricsRadar を提案した. 歌詞の深層的な意味をトピックレーダーチャートとして表現することで, トピックの可視化とインタラクティブで多様な入力手段を可能とした.

- 今後の課題
  - 個人性を考慮したユーザ適応型UIの実現
  - 階層的なトピック分析によるトピック分析機能の高度化

\* R. Adams et al., Tree-Structured Stick Breaking Processes for Hierarchical Data, Proc. NIPS, 23, pp. 19-27, (2010).

# 半教師あり NMF を利用した 音楽信号中のフレーズ検出

増田 太郎 吉井 和佳\*1 後藤 真孝\*2 森島 繁生

早稲田大学 \*1 京都大学 \*2 産業技術総合研究所

## 研究の背景と目的

ある楽器で演奏したフレーズと同じフレーズが、  
どの曲のどの時間位置に含まれるかを検索したい！

背景：音響信号から直接楽曲を検索する研究の興隆

- Query-by-Humming システム：ユーザの歌唱に基づく曲検索
  - データベース中に MIDI ファイルなどの楽譜情報が必要
- フィンガープリントシステム：楽曲そのものをクエリとする曲検索
  - 楽曲の一部の構成要素(楽器など)を手掛かりに探すことは困難

目的：楽器で演奏したフレーズに基づく楽曲検索



- ・楽譜情報不要
- ・主旋律以外のメロディも検索可能

楽器で演奏したフレーズをクエリとする検索への需要

- ・楽曲名を知らない / 忘れてしまった場合でも、その楽曲の特徴的なフレーズを演奏するだけで、直感的に楽曲検索可能
- ・思いついたフレーズが、既存の楽曲中でどう編曲されているかを学べる

## 評価実験

実験条件

- Exact-match：クエリと全く同一の音響信号が楽曲中に含まれる場合
- 楽器変化：楽曲中とは異なる楽器で演奏した場合
- テンポ変化：同一の楽器、楽曲中よりも 20% 速いテンポで演奏した場合
  - ・データベース = 4 曲
  - ・クエリ = 10 種類

実験結果

比較手法に比べて、提案手法がより優れた検索性能を実現

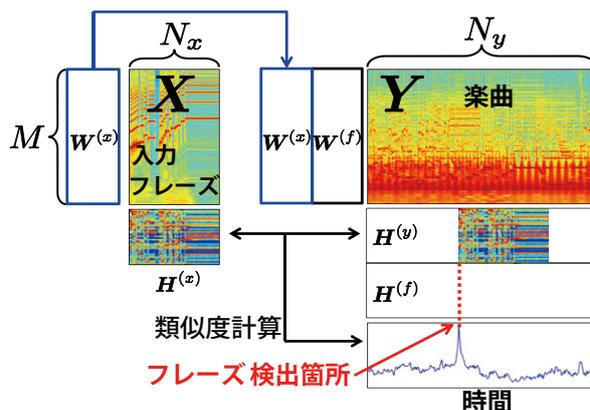
	F-measure( 平均値 ) [%]		
	条件 1	条件 2	条件 3
スペクトル DP	3.6	6.2	2.7
MFCC	29.0	0	0
クロマグラム	43.1	23.0	14.7
提案手法	57.9	36.5	23.4

- ・従来手法…他の楽器の存在により特徴量が歪められ、性能が落ちる
- ・提案手法…楽器やテンポの変更への頑健性が課題

## フレーズ検索手法

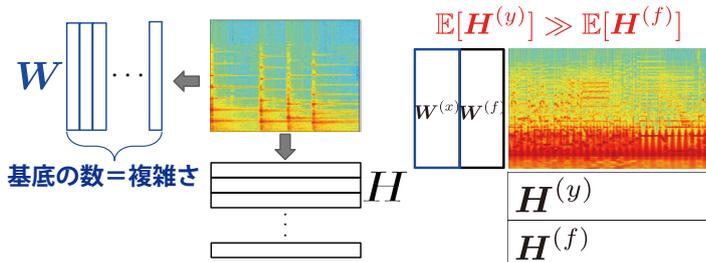
入力音と楽曲の一部の要素との類似度を計算する手法が必要

- ・クエリのスペクトログラムに GaP-NMF [Hoffman et al. 2010] を適用し、基底スペクトル及びアクティベーションを推定
- ・検索対象の楽曲に対し、クエリの基底を固定することで、**固定した基底に対応するアクティベーション**を計算(半教師あり NMF)
- ・クエリ及び楽曲から得たアクティベーション間の相関係数を類似度と定義
  - 類似度のピークをフレーズ検出箇所とする



ノンパラメトリックベースの利点を活用

- ・適切な基底の数 = 複雑さを自動的に決定
- ・固定 / 非固定の音量に対し、事前分布を調節する



## まとめ・今後の課題

楽器フレーズを入力とする楽曲検索手法を提案

- ・クエリのスペクトル基底を辞書とする半教師あり NMF
- ・固定した基底のアクティベーション類似度に基づくフレーズ検出
- ・計算時間以外の点では、従来手法を上回る検索性能を確認

今後の課題

- ・楽器、テンポの変動に対する検索精度の向上
- ・計算コスト削減
- ・大規模データベースに対する評価実験

# 視聴者コメントに基づく楽曲動画の印象推定とデータセット構築

山本 岳洋 中村 聡史\*1  
 京都大学 \*1 明治大学

本研究の目的

## 印象に基づく楽曲検索の実現

### 楽曲から受ける「印象」

- ・「爽やか」、「切ない」、「可愛い」、「昭和っぽい」、「酒が飲みたくなる」・・・
- ・初音ミクの有名な曲の中で、**爽やかな感じの曲が聴きたい!**
- ・この曲と、**似た印象の曲が聴きたい!**

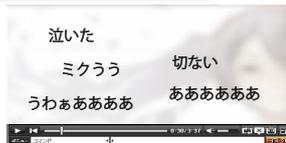
### 何が問題か?

- ・楽曲の印象に関するタグはわずか **5%**  
 - 楽曲の印象を推定する技術の必要性
- ・評価用データセットの **欠如**  
 - 印象を扱った、標準的なデータセットの必要性

## コメントに基づく印象推定

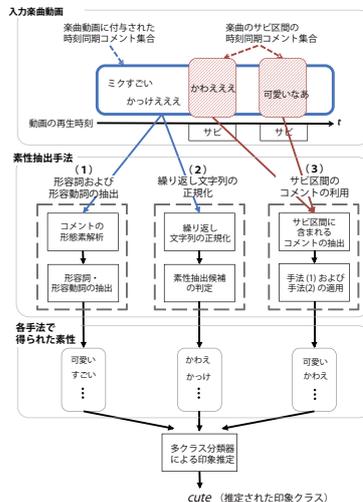
### アプローチ：視聴者コメント

- ・コメント = 視聴者の反応
- ・既存のアプローチ  
 - 音響特徴量や歌詞の利用



### コメントからの素性抽出

- ・形容詞・形容動詞
- ・繰り返し文字の正規化  
 かわいいいいい  
 きたああああああああ  
 感情を強く表す表現  
 [Brody and Diakopoulos 2011]
- ・サビ区間の利用  
 仮説：楽曲のサビが楽曲に対する印象を決定



## 印象データセットの構築

### コミュニティで共有できる評価データの必要性

- ・印象に基づく楽曲動画の検索アルゴリズムの性能を評価したい
- 現状では評価データを1から作成する必要
- ある程度の規模の評価値を集めることは高コスト

### 楽曲動画に対する印象データセットの構築・公開

## 楽曲動画印象データセット

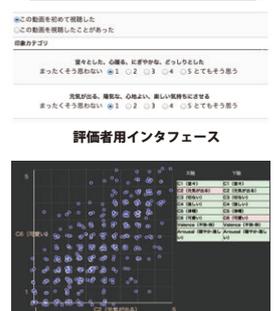
### データセットの概要

- ・楽曲動画数: **500 件**
- ・各動画あたり最低 3 名の評価者を割り当



### 既存研究に基づいた 8 印象クラス

- ・堂々
  - ・元気が出る
  - ・切ない
  - ・激しい
  - ・滑稽
  - ・可愛い
  - ・Valence
  - ・Arousal
- [MIREX] (堂々, 元気が出る, 切ない, 激しい, 滑稽)
- [Russel1980] (可愛い, Valence, Arousal)



データセットを利用した楽曲検索システム

## 評価実験

### 7 印象クラス計 1,314 本の動画を対象

形容詞が有効な特徴, 他の特徴と組み合わせることで分類精度向上

	形容詞	正規化	サビ	全ての組合せ
マクロF値	.627	.555	.526	.659
マイクロF値	.661	.556	.540	.676

歌詞や音響特徴と比べてコメントが分類に有効. コメントと音響等の組合せによる精度向上

	コメント	歌詞	音響	全ての組合せ
マクロF値	.651	.365	.301	.664
マイクロF値	.673	.441	.380	.687

下記 URL にてデータセットを公開中!

<http://nkmr.io/mood/>

### データセット第 2 弾

- ・音楽のみ, 映像のみ, 音楽+映像の 3 種類のデータに対して, 同様の評価データを構築
- ・映像の有無が音楽に対する印象に与える影響の分析

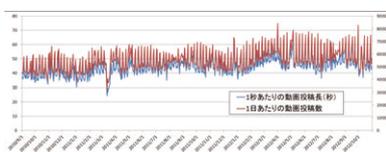
# 音楽理解とアノテーション分析技術によるサムネイル動画自動生成

中村 聡史 山本 岳洋\*1 後藤 真孝\*2 濱崎 雅弘\*2  
 明治大学 \*1 京都大学 \*2 産業技術総合研究所

## 背景と研究目的

膨大な動画から目的とする動画を探すのは困難

- 日々投稿される動画の数は膨大  
1日に投稿される動画は5500本程度  
1秒間に47秒分の動画が投稿される
- 動画に出会うことができない  
機会を損失している



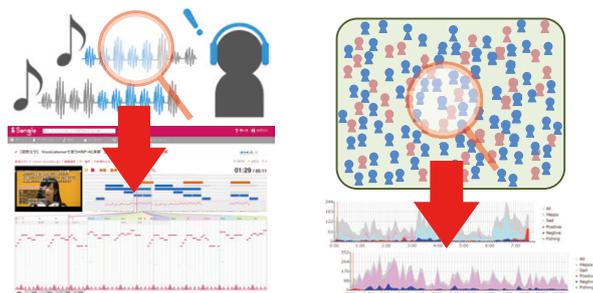
サムネイル動画を自動生成してユーザに提示

- 動画要約ではなく、動画の中でもっとも魅力的な15秒を抽出して提示
  - 視聴判断のための魅力的な15秒を如何にして抽出するか？
- サムネイル動画は推薦動画の候補や、検索結果ランキング上で短時間で試聴するかどうかを判断するためや、デイリーランキング動画等で使用可能

## アプローチ

音楽理解技術とソーシャルアノテーション分析技術の融合によりここぞというシーンを抽出してユーザに提示

- 音楽理解に基づくサビ検出技術
- ソーシャルアノテーション分析技術による視聴者の盛り上がり検出技術



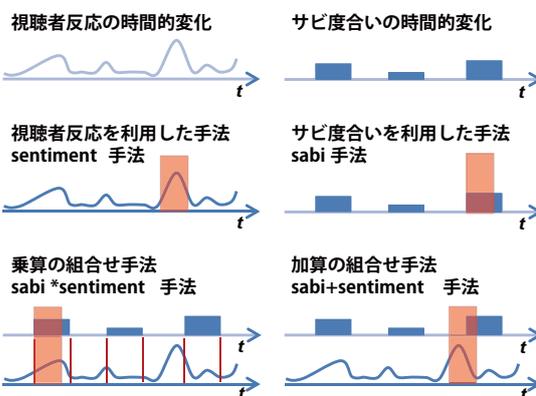
楽曲動画の理解

魅力的な15秒の抽出

## 提案手法

仮説

- サビ部分は楽曲的に盛り上がる部分ではないか？
- 視聴者が盛り上がっているシーンは、盛り上がるシーンではないか？



最もスコアの高い15秒を抽出してサムネイル動画を生成

## 結果

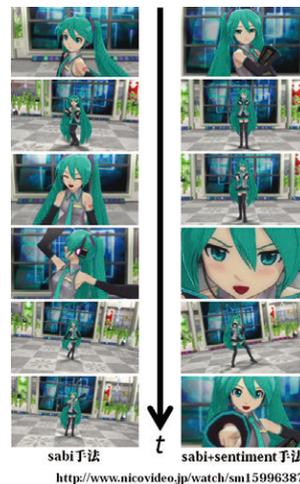
評価実験により有効性の検証

- 比較のために中央15秒を抽出する middle 手法とコメント量が多い部分を抽出する comment 手法を用意
- 12人が29件以上の動画(29動画×6手法)をスコアリング
- sabi + sentiment 手法がもっともよい結果(独立はすべての手法が異なるシーンをサムネイル動画として抽出したもの)
- サビの少し前から再生すると評価が高くなる傾向(左下表)
- sabi + sentiment 手法は人の心に響くシーンを抽出する傾向あり(右下图)

手法	重複: 145	独立: 64
middle	3.96	3.70
comment	3.29	3.36
sentiment	3.54	3.69
sabi	4.16	4.19
sabi*sentiment	4.04	4.30
sabi+sentiment	4.28	4.63

$$\delta = \epsilon_{sabi} - \epsilon_{sabi+sentiment}$$

開始時間の差	sabi	sabi+sentiment
$\delta \leq -15$	4.40	4.37
$-15 < \delta < -1$	3.96	4.65
$-1 \leq \delta \leq 1$	4.02	4.00
$1 < \delta \leq 15$	4.00	4.00
$15 \leq \delta$	4.26	4.55



http://www.nicovideo.jp/watch/sm15996387

# N次装飾: 動画共有サイト上の動画に対する装飾とその共有手法

中村 聡史   石川 直樹\*1   渡邊 恵太

明治大学   \*1 東京農工大学



見ているYoutubeをその場で編集  
自分好みにアレンジ、そして共有できる  
N次装飾プラットフォーム

ウェブ上で完結する手軽な動画編集が  
個人の繊細な創造性を汲み取り  
N次創作を加速する

## N次装飾とは

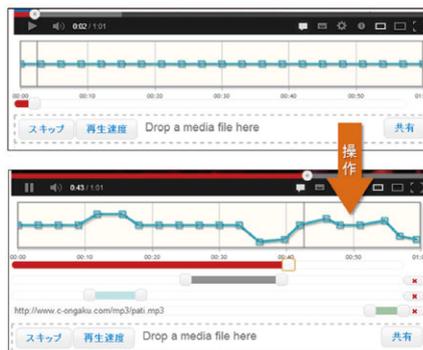
「この動画は良いものだ。ただ、ここをちょっと変えたい」そんなちょっとした編集(装飾)で動画を自分の好みにし、満足度を高められます。

他者の動画や音楽へのリスペクトとして、アレンジを行うN次創作行為がウェブ上では盛んですが、このような行為は、個人が一度動画をダウンロードし、専用の編集ソフトを利用しなければなりません。N次創作されたコンテンツは高い魅力を持つものの、制作のためのモチベーションやプロセスの敷居が高いため、だれでも手軽にできるものではありませんでした。

そこで本研究ではブラウザ上で動画プレイヤーに手を加えることによって、簡易的な装飾を施せるシステムを開発しました。

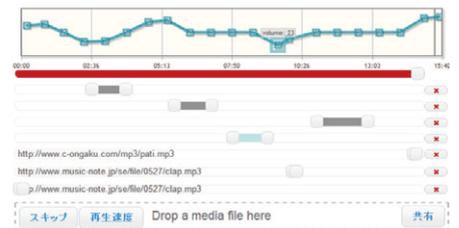
## 装飾のためのUI

動画プレイヤーの下に、音量設定、付加した音楽コンテンツのタイムラインがいつでも追加できます。またその開始と終了位置を設定できるので、BGMだけでなく効果音としての付加も可能です。



## ウェブすべてが素材

一般的な動画編集ソフトは、どこで何の素材を利用しているかの参照情報が明確でない場合があります。本システムは装飾で使うコンテンツも他のウェブ上から利用する仕組みです。アドレスを貼り付けるだけで様々なメディアを素材として利用でき、同時参照も明確になるメリットがあります。



★ 詳細はこちら。体験&利用可能です。  
<http://nkmr-lab.org/app/VideoMixer/>

# 自由なテンポで演奏した複数の演奏データから楽曲を生成するシステム

宮下 芳明 川名 勇気

明治大学

## 概要

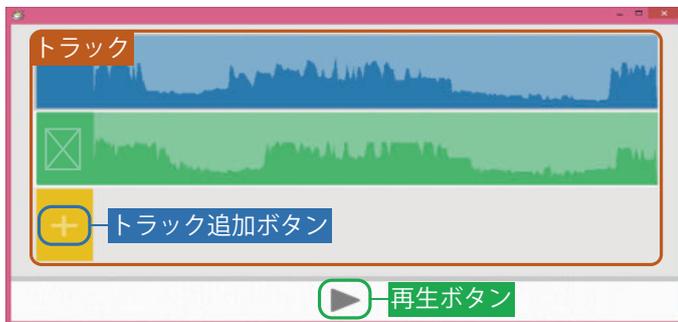
これまでの楽曲の制作手法は、ガイドリズムのクリック音や、既存の演奏のテンポに合わせて演奏を行う必要があり、奏者ひとりひとりのテンポの緩急によって生じる音楽表現は犠牲にされてきた。

各パートの奏者が自由なテンポで録音した演奏データを統合する、ユーザが選んだパートのテンポに合わせて合奏を行っているような楽曲を生成できるシステムを提案する。

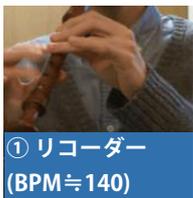
演奏データの数だけ楽曲のテンポを選ぶことができるため、テンポによって生じる音楽表現について想像の幅を広げることが可能となっている。

## システムインタフェース

- ・演奏データを読み込ませるための「トラック」
- ・トラックを追加するための「トラック追加ボタン」
- ・楽曲を生成して再生するための「再生ボタン」

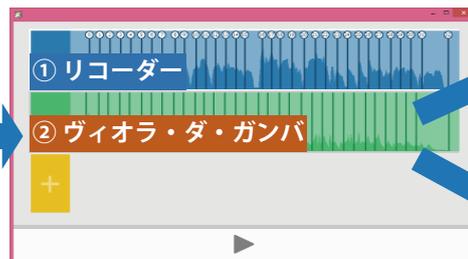


## 創作例



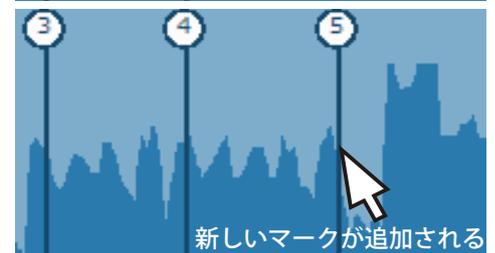
元となる楽譜 (メヌエット ト長調)

2つの演奏を録音した上でシステムを用い、マーク付けを行って音楽を生成。テンポの異なる2種類の合奏曲を生成可能。



## 楽曲の生成方法

1. トラック追加ボタンで演奏データを新しいトラックに読み込ませる  
(読み込みが成功すると演奏データのスペクトログラムが作成される)
2. 1.を繰り返してすべての演奏データを各トラックに読み込ませる
3. テンポを合わせるための「マーク付け」を各トラックに対して行う



4. 楽曲の基準のテンポの演奏データを「指揮データ」として選択する
5. 再生ボタンを押すことで、マーク内のテンポが指揮データに合わせて再生し、音楽を生成・再生する

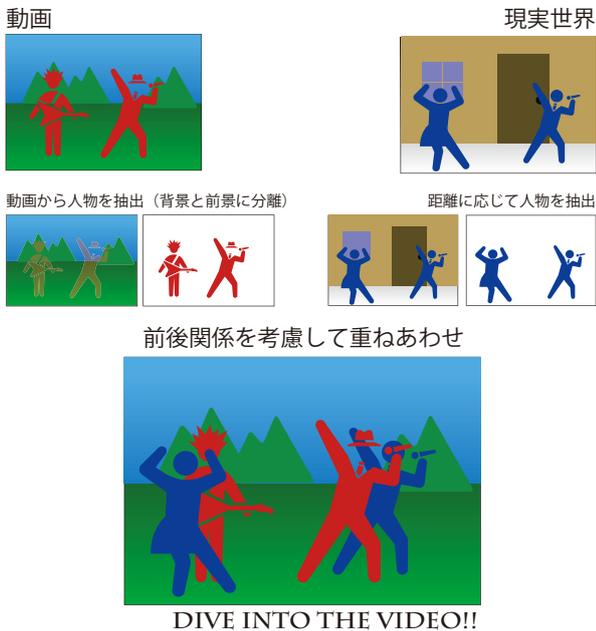
# VRMixer: 動画と現実の融合による 新たなコンテンツの生成

平井 辰典 中村 聡史\*1 森島 繁生 湯村 翼\*1  
早稲田大学 \*1 明治大学

## VRMixer とは

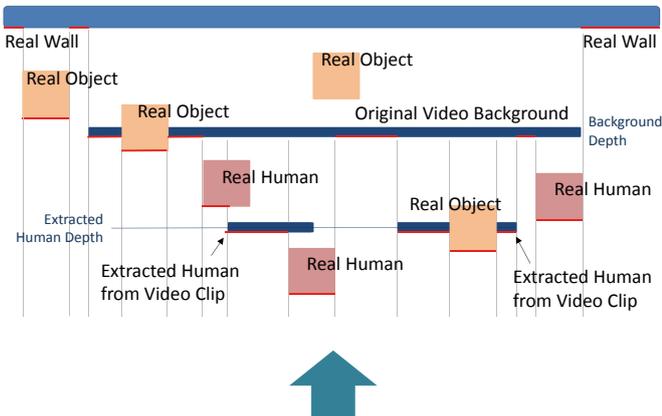
### VRMixer

任意の動画の中に入り込み、登場人物とのインタラクションを  
とっているかのような感覚を実現する動画と現実世界の融合手法  
**動画から人物を抽出 + 現実世界から人物を抽出**  
抽出されたそれぞれの人物を抽出して融合



### 提案手法

深度センサ付きカメラにより実世界の人物画像を距離情報付きで抽出  
動画からの人物領域の自動切り出しにより動画の背景と前景を分離  
↓  
前後関係を考慮して背景、切り出した人物、現実世界を空間に配置  
位置合わせについてはキーボード操作により手作業で行う



## 人物領域の自動検出

### 人物領域の自動検出手法

1. 人物領域と背景領域をおおまかに検出するため  
動画の各フレームに対して以下を適用
    - ・顔検出手法 (ASAM: Active Structure Appearance Model)
    - ・人物検出手法 (HOG: Histogram of Oriented Gradients)
  2. 人物領域と背景領域の色分布を変分ベイズ GMM によって学習
  3. 動画の各ピクセルの色が人物領域と背景領域のどちらの色に近いかを  
基に 3 次元グラフカットで領域のセグメンテーションを行う
  4. ノイズ除去のために膨張収縮処理及びラベリング処理を行う
- (顔検出が適用できない動画については人物領域の抽出不可)

## プロトタイプシステム

### システム構成

PC + Kinect + openFrameworks

### 動画は事前に処理し、前景動画と背景動画をそれぞれ生成

(前景動画の背景部分は青色で塗りつぶす)  
深度カメラから計算された距離に応じて現実世界のオブジェクトを配置

### 下記の順で描画を行う

1. 背景動画を描画
2. 背景動画と前景動画の間の距離にあるオブジェクトを描画
3. 前景動画を描画 (青色の部分は透過処理)
4. 前景動画より前の距離にあるオブジェクトを描画

## システムの運用結果

### ユーザのフィードバック

- ・ 踊ってみた動画と一緒に踊ることの楽しさ
- ・ 間違えている部分がかげがすぐに分かるためダンス練習に最適
- ・ 憧れのアイドルと一緒に踊れることの喜び
- ・ 足元と空間の一部を共有することによる不思議な一体感など

# 統計的機械学習に基づく音楽解析

吉井 和佳 持橋 大地\*1 後藤 真孝\*2

京都大学 \*1 統計数理研究所 \*2 産業技術総合研究所

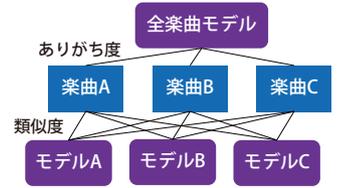
## 音楽データの教師なし構造学習を目指して

確率的な枠組みを用いて類似度・ありがち度の計算に客観的な裏付け(エビデンス)を与えたい

類似度・ありがち度: ある確率モデル(個別楽曲モデル・全楽曲モデル)から与えられた音楽データが生成される確率  
確率が高い=予測しやすい→確率モデルの学習に用いたデータに対して類似度・ありがち度が大きい

WEB上に存在する大量の音楽データから音楽に内在する構造を教師なしで学習したい

音楽データ(信号・記号データ)の生成過程を、理論上は「無限の複雑さ」をもつノンパラメトリックベイズモデルで表現  
実際には有限の音楽データが与えられると、それを説明するのに必要な「実効的な複雑さ」が自動的に決定→構造学習が可能



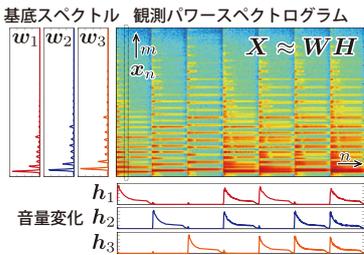
## 音楽音響信号に対するノンパラメトリックベイズ学習

音楽音響信号が高々有限個の「部品」から構成されていると仮定し、音楽音響信号の構造を教師なし学習

何を部品とみなすかによって異なる確率モデルが定式化→部品の個数が未知であるのでノンパラメトリックベイズモデル+変分ベイズ法などの最適化技法

### 非負値行列分解

(Nonnegative Matrix Factorization: NMF)



非負ベクトルを非負ベクトルの凸結合で近似

$$x_n \approx \sum_{k=1}^K w_k h_{kn} \stackrel{\text{def}}{=} y_n$$

観測行列  $X = [x_1, \dots, x_N] \in \mathbb{R}^{M \times N}$

基底行列  $W = [w_1, \dots, w_K] \in \mathbb{R}^{M \times K}$

音量行列  $H = [h_1, \dots, h_K]^T \in \mathbb{R}^{K \times N}$   
全て非負ベクトル

Bregmanダイバージェンスの最小化 = ある特定の確率モデルの最尤推定

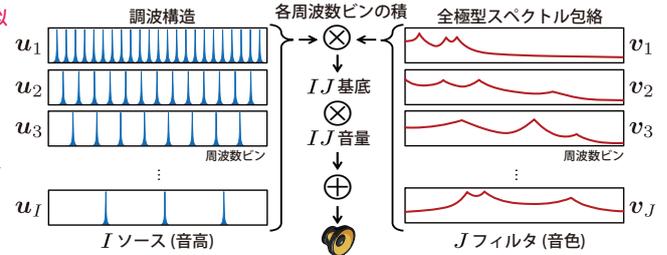
$$D_\phi(x_n|y_n) = \phi(x_n) - \phi(y_n) - \phi'(y_n)^T(x_n - y_n) \quad \phi: \text{凸関数}$$

Pros: ガンマ過程事前分布を導入することで基底数の無限が可能 [Hoffman 2010]

Cons: 音色単位の分解が不可能・周波数成分が全て独立であるという強い仮定

### 無限複合自己回帰モデル

(Infinite Composite Autoregressive Model: iCAR)

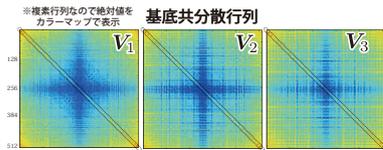


音楽音響信号は無限個の音高・音色の組み合わせから生成されていると仮定

$$x_n \approx \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} (u_i \odot v_j) h_{ijn} \stackrel{\text{def}}{=} y_n$$

音高で分離: 多重音F0推定  
音色で分解: 楽器パート分離 [Yoshii 2012]

### 無限半正定値テンソル分解 (Infinite Positive Semidefinite Tensor Factorization: iPSDTF)



すべての行列が対角行列であれば  
PSDTFはNMFに帰着

PSDTFは周波数成分間の相関を考慮可能

半正定値行列を半正定値行列の凸結合で近似

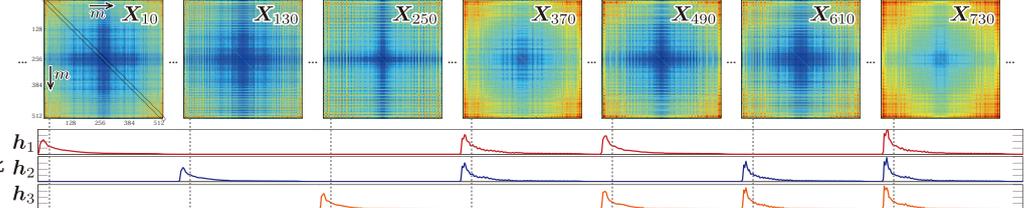
$$X_n \approx \sum_{k=1}^{\infty} V_k h_{kn} \stackrel{\text{def}}{=} Y_n$$

観測テンソル  $X = [X_1, \dots, X_N] \in \mathbb{R}^{M \times M \times N}$

基底テンソル  $V = [V_1, \dots, V_K] \in \mathbb{R}^{M \times M \times K}$

音量行列  $H = [h_1, \dots, h_K]^T \in \mathbb{R}^{K \times N}$

テンソルデータ: 局所的な共分散行列(複素スペクトルとその共役な複素スペクトルとの直積)の集合



Bregman行列ダイバージェンスの最小化 = ある特定の確率モデルの最尤推定

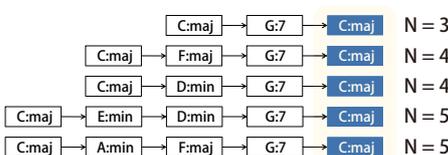
$$D_\phi(X_n|Y_n) = \phi(X_n) - \phi(Y_n) - \text{tr}(\nabla\phi(Y_n)^T(X_n - Y_n)) \quad \phi: \text{行列凸関数}$$

異なる音高の楽器音(C,E,G)を混合した音響信号を用いて分離実験 [Yoshii 2013]  
ピアノ・ギター・クラリネットの平均でSDR/SIR/SARともに4.0[dB]程度改善

## 楽譜情報に対するノンパラメトリックベイズ学習

重要な記号データのひとつであるコード系列に着目し、その背後にあるコード進行の確率モデルを教師なし学習

N-gramモデルにおけるNの値が可変・コードの語彙を恣意的に決めたくない→ノンパラメトリックベイズモデル+マルコフ連鎖モンテカルロ法などの最適化技法



### 語彙フリー無限グラムモデル

(Vocabulary-Free Infinity-gram Model)

各コードについて最適なNを推定可能→コードパターンの発見  
理論上は無限語彙を扱うことが可能であるので [Yoshii 2011]  
将来的に新しいコードラベルが追加されても影響を受けない

# 統計的機械学習に基づく 言語＝音楽情報処理の横断的展開

持橋 大地      吉井 和佳\*1      後藤 真孝\*2

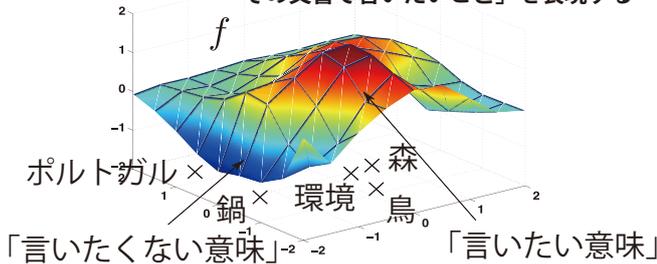
統計数理研究所      \*1 京都大学      \*2 産業技術総合研究所

## 単語の意味の連続空間埋め込み

動機：歌詞や楽曲の存在する空間を探索するための統計的モデルが必要→トピックモデル  
LDA などの従来モデルではなぜいけないのか？  
・文書の表現が確率分布の単体上に限定  
・混合モデルでは「意味の重なり」が表現不可能

## ナビゲーションのための 連続空間トピックモデル (CSTM) ガウス過程による文書の意味のモデル化

ガウス過程  $f \sim GP(0, K)$  によって  
「その文書で言いたいこと」を表現する



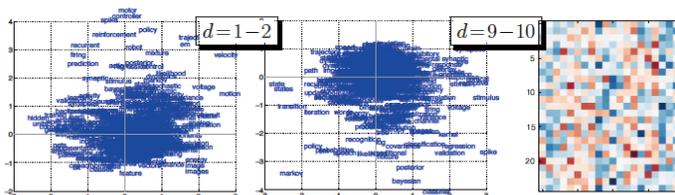
## 文書の確率的生成モデル

$f \sim GP(0, K)$      $K(v, v') = \phi(v)^T \phi(v')$   
 $\alpha(v|d) = \alpha_0 G_0(v) e^{f(\phi(v))}$     言いたい意味の単語が  
 $G \sim DP(\alpha)$     出やすくなるようスケールアップ  
 $w_1 w_2 \dots w_N \sim G$     文書ごとに単語分布に対する  
 ディリクレ過程をきめ細かに制御

## 実験結果

Perplexity	データ	LSTM	RSM	SDM	LDA
	NIPS	1383.66	1290.74	1638.94	1648.3
	KOS	1632.35	1396.61	1936.25	1730.7
	毎日新聞	466.83	622.69	582.37	507.39

## Latent Embedding (NIPS, $d=10$ )



## 動画コメントからの感情語抽出

動画のコメント：辞書にない文字列の連続感情に依存（どこまでが単語かすら分からない）  
文字列を単語に分けつつ、感情語と機能語に分類するための2クラスの隠れセミマルコフモデルを提案  
・確率的生成モデル  
For  $n = 1 \dots N$   
 (1) 文字列言語モデル A または B を確率的に選択  
 (2) 選んだモデルから単語文字列を生成  
 (3) 出力文字列の末尾に追加  
 ・上の逆問題を MCMC で解く

## 解析例（赤が感情語）

リリイいいねえ・・・声 以外は  
 がくぼに聞き入ってしまう件 wwwばねえながくぼ www  
 オレンジに すりゃよかったのに  
 調教 次第ではすこく 伸びそうな声  
 オレンジが赤がよかった  
 この場合 緑 www が良かった b  
 もう好きな 食べ物 神のアクア でよくな www  
 地に落ちた 神にしか見え ない  
 リリイ可愛い ww

## 得られた感情語彙量の例（数字は出現回数）

"positive"の場合	"sad"の場合	"happy"の場合	"negative"の場合
10537 www	30301 ..	15149 www	5322 www
9929 あああ	20949 ..	2579 ww	3775 これは
8505 きた	8454 ...	1775 ktr	3376 !
7850 !	2822 !	1058 www	2889 w
7687 w	2286 ?	895 神曲	2554 ..
5472 かわいい	1658 ★	769 吹いたwww	2355 ひどい
5317	1594 t	743 萌	2171 ごめんなさい
5034 ..	1346 泣ける	651 w	2104 あ
4465 すげ	1272 が・	459 吹いた	1657 はは
4185 好きだ	1255 w	410 わろた	1438 ww
4015 ww	1143 www	399 !	1320 よ
3576 神曲	1066 マスター	369 笑い	1148 ?
3480 神	1065 涙	362 え	1123 え
3422 え	1028 な・	343 面白い	1056 はい
3314 ( √ )	984 ああ	313 ww	962 ね
2965 好き	953 ざわ	308 笑える	903 ぎゃああ
2922 いいね	940 か・		

## 今後の展開：統計的音楽生成

### 確率的文法に従って音楽を生成できる可能性

a red to expected say the city lindner random suffer out found note various future solution  
 (S (NP (-NP (-NP (DT "a")  
 (ADJP (JJ "red")  
 (VP (TO "to")  
 (VP (VBN "expected")  
 (VP (VBP "say") (NP (DT "the") (NN "city"))))))))  
 (NNP "lindner")  
 (NNP "random")  
 (VP (-VP (VB "suffer") (RP "out")  
 (NP (-NP (-NP (-NP (VBN "found") (NN "note")  
 (JJ "various")  
 (NN "future")  
 (NN "solution"))))

Penn Treebank PCFG  
 に基づくランダム  
 ウォーク文生成の例

# 音楽音響信号中の調波音の周波数特性 およびドラムの音色の置換システム

中村 友彦 吉井 和佳\*1 後藤 真孝\*2 亀岡 弘和

東京大学 \*1 京都大学 \*2 産業技術総合研究所

## システムの概要

### 加工による音楽理解の促進

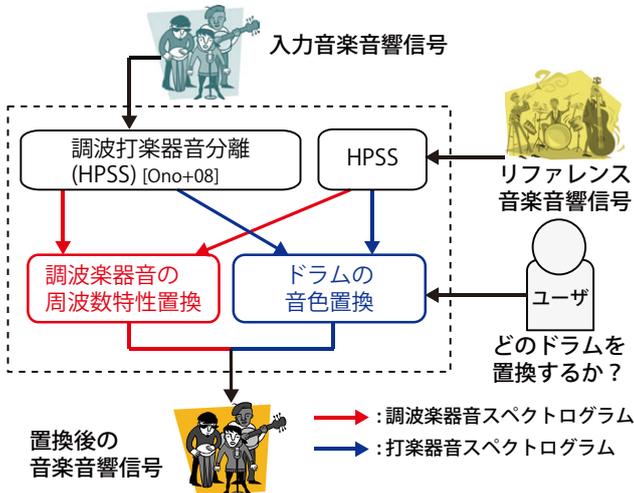
音楽を聴くだけでなく、加工して深く理解  
既存楽曲を編集して、自由にリミックス曲を作成可能

### 音楽音響信号間で音色情報を置換

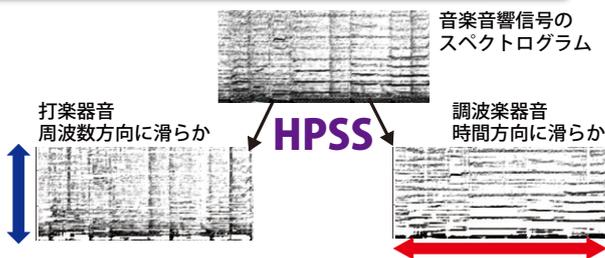
入力およびリファレンスとして多重音の音楽音響信号が使用可能

調波楽器音と打楽器音の両方を置換可能

楽譜情報がなくても置換可能



## 調波打楽器音分離 [Ono+08]



## 調波楽器音の周波数特性置換

### 入力の調波楽器音成分の振幅スペクトルを変形

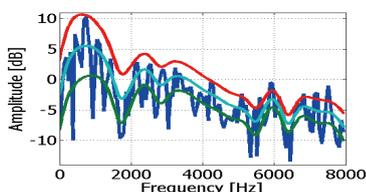
調波楽器音スペクトルからボトム・トップエンベロープの推定 [亀岡 +06]

入力のエンベロープがリファレンスのエンベロープに似るように

入力の振幅スペクトルを変形

トップエンベロープ  
調波構造を近似的に表現

ボトムエンベロープ  
歌声の子音などに対応



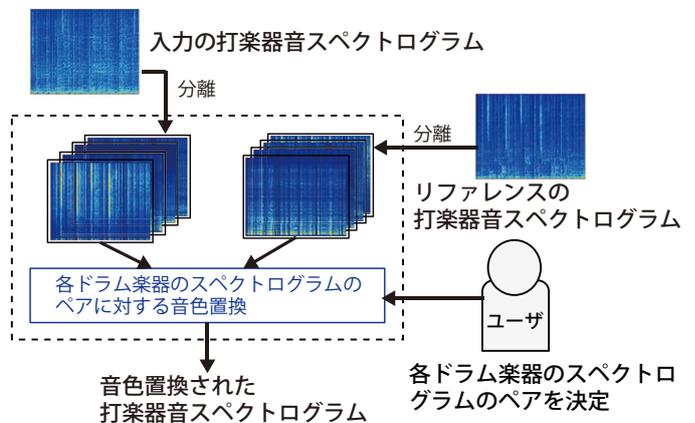
## ドラムの音色置換

### リファレンスのドラムの音色を用いて

### 入力のリズムの打楽器音複素スペクトログラムを構築

HPSS で得られた打楽器音の複素スペクトログラムを、非負値行列因子分解と Wiener フィルタにより各ドラム楽器の複素スペクトログラムに分離

入力とリファレンスの各ドラム楽器音のペア同士で音色転写



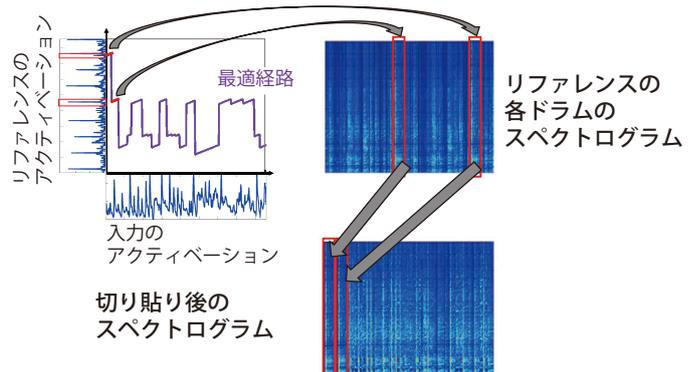
### 切り貼り法

入力とリファレンスのスペクトログラムを時間アライメント

音色の差異に頑健な特徴量：非負値行列因子分解のアクティベーション

動的計画法により効率的に最適経路を導出

最適経路にしたがって、リファレンスの複素スペクトログラムを切り貼り



## 主観評価実験

### 11人の被験者による5段階MOS評価

Q1:「調波楽器音の音色がリファレンスから入力に適切に置換されているか」

Q2:「ドラムの音色がリファレンスから入力に適切に置換されているか」

### システムが適切に動作していることを確認

Q1とQ2どちらについてもMOS値が1に比べ有意差有り

(t検定:  $p < 0.01$ )

# 音楽音響信号中の歌声 F0 軌跡に対する歌唱表現の転写システム

池宮 由楽

糸山 克寿

吉井 和佳

京都大学

## 歌唱表現転写システムとは

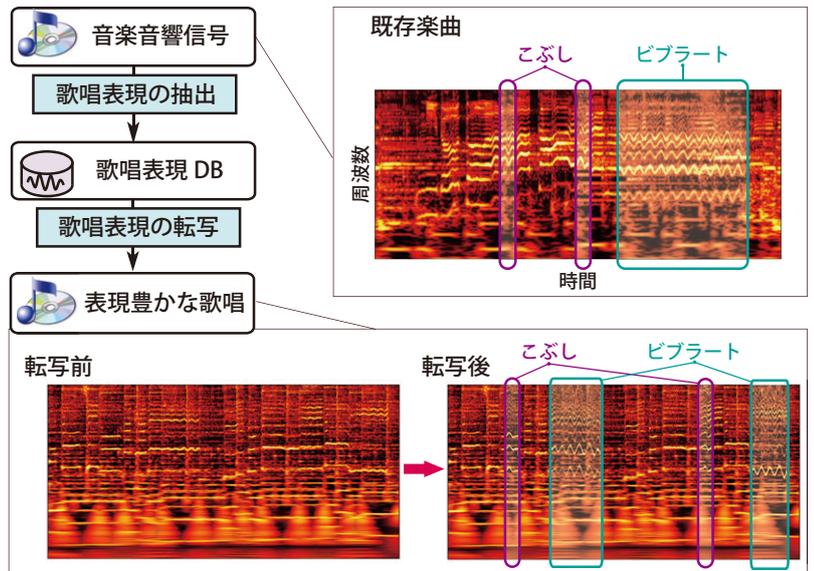
音楽音響信号から歌手の**歌唱表現**を抽出し別の歌手の歌唱や歌声合成器に豊かな表現を転写するシステム

**歌唱表現** 歌声 F0 軌跡の特徴的な変動

ビブラート  こぶし   
グリッサンド 

## こんなことができるように

「美空ひばりのようなこぶしで初音ミクに歌ってほしいでも本格的なパラメータチューニングは難しそう」  
「スピッツのグリスタウンが好きでたまらないこの曲もスピッツのような歌い方で聞いてみたい」



## 歌唱表現の抽出

### 歌唱F0推定

音響信号中に含まれる歌声のF0軌跡を見つける

時間周波数領域での最適時系列探索問題として定式化

歌声 F0 軌跡の滑らかさを課すマルコフモデル  
周波数探索範囲の制限

音楽音響信号 + 音高列

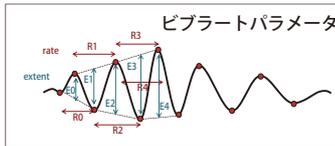
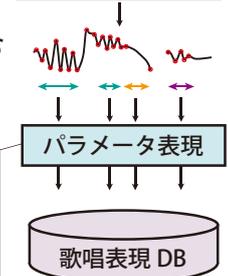


### 歌唱表現同定

推定された歌声F0軌跡から歌唱表現を見つける

各歌唱表現を操作可能なパラメータ表現に落とし込む  
パターンマッチングにより歌唱表現を同定

特徴点検出  
歌唱表現同定



## 音楽音響信号への歌唱表現転写

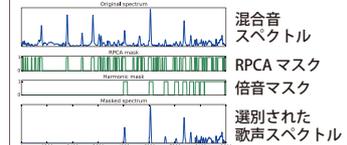
### ユーザによる歌唱表現転写

任意歌唱の任意箇所に対して歌唱表現を転写する

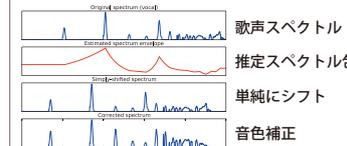
ユーザが F0 存在範囲・歌唱表現転写箇所をスペクトログラム上で提示 (GUI 画面)



RPCAによるマスク & F0を用いた倍音マスク



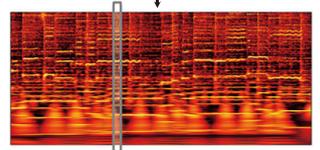
スペクトル包絡を用いた音色補正



音色の自然性に関する聴取実験  
赤: 音色補正なし  
緑: 音色補正あり

音楽音響信号

定Q変換

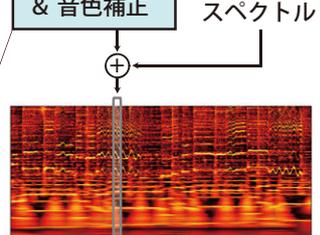


スペクトル選別

歌声スペクトル

音高シフト & 音色補正

伴奏スペクトル



位相復元 & 定Q逆変換

表現が転写された音楽音響信号



## OngaCREST プロジェクト研究参加者一覧

---

<b>後藤グループ</b>	後藤 真孝	(産業技術総合研究所)	2011年10月～
	濱崎 雅弘	(産業技術総合研究所)	2011年10月～
	中野 倫靖	(産業技術総合研究所)	2011年10月～
	深山 覚	(産業技術総合研究所)	2013年4月～
	加藤 淳	(産業技術総合研究所)	2014年4月～
	Matthew McVicar	(産業技術総合研究所)	2013年9月～
	Graham Percival	(産業技術総合研究所)	2014年6月～
	川崎 裕太	(産業技術総合研究所)	2011年12月～
	石田 啓介	(産業技術総合研究所)	2012年4月～
	井上 隆広	(産業技術総合研究所)	2012年11月～
	田中 一大	(産業技術総合研究所)	2012年4月～
	戸田 智基	(奈良先端科学技術大学院大学)	2012年4月～
	栗原 一貴	(津田塾大学)	2011年10月～
	藤原 弘将	(産業技術総合研究所)	2011年10月～2012年8月
	Philippe Hamel	(産業技術総合研究所)	2012年11月～2013年5月
Matthew Davies	(産業技術総合研究所)	2013年3月～2013年12月	

---

<b>森島グループ</b>	森島 繁生	(早稲田大学)	2011年10月～
	大矢 隼士	(早稲田大学)	2012年4月～
	張 卓鵬	(早稲田大学)	2014年7月～
	平井 辰典	(早稲田大学)	2011年10月～
	岩本 尚也	(早稲田大学)	2014年2月～
	福里 司	(早稲田大学)	2014年4月～
	佐々木 将人	(早稲田大学)	2012年4月～
	増田 太郎	(早稲田大学)	2012年4月～
	岡田 成美	(早稲田大学)	2014年4月～
	赤堀 渉	(早稲田大学)	2014年7月～
	朝比奈 わかな	(早稲田大学)	2014年7月～
	佐藤 晴紀	(早稲田大学)	2014年7月～
	松田 聡子	(早稲田大学)	2013年10月～
	前島 謙宣	(早稲田大学)	2011年10月～2014年5月
	濱崎 皓介	(早稲田大学)	2012年4月～2014年3月

---

<b>中村グループ</b>	中村 聡史	(明治大学)	2011年10月～
	山本 岳洋	(京都大学)	2011年10月～
	佃 洸撰	(京都大学)	2011年10月～
	宮下 芳明	(明治大学)	2012年10月～
	渡邊 恵太	(明治大学)	2013年4月～
	大島 遼	(明治大学)	2013年11月～
	湯村 翼	(明治大学)	2013年12月～

---

<b>吉井グループ</b>	吉井 和佳	(京都大学)	2011年10月～
	糸山 克寿	(京都大学)	2014年4月～
	池宮 由楽	(京都大学)	2014年4月～
	持橋 大地	(統計数理研究所)	2012年4月～
	中村 友彦	(東京大学)	2014年4月～



OngaCRESTシンポジウム2014 (<http://ongacrest.jp/symposium2014>)  
「音楽情報処理研究が切り拓く未来を探る」予稿集

科学技術振興機構 (JST) 戦略的創造研究推進事業 (CREST)  
研究領域「共生社会に向けた人間調和型情報技術の構築」  
平成23年度採択研究課題

「コンテンツ共生社会のための類似度を可視化する情報環境の実現」



---

発行日 2014年8月23日

代表者 後藤 真孝

編集者 加藤 淳 笠井 志麻

発行 OngaCREST プロジェクト <info@ongacrest.jp>

独立行政法人 産業技術総合研究所

情報技術研究部門 メディアインタラクション研究グループ

〒305-8568 茨城県つくば市梅園 1-1-1 中央第二

TEL 029-861-3309 FAX 029-861-3313

---