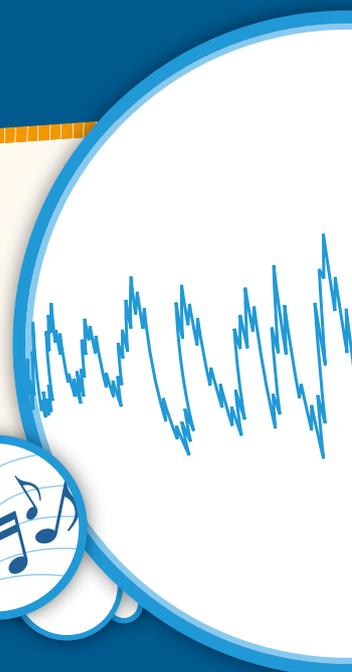


● OngaCREST シンポジウム2016

音楽情報処理研究が 切り拓いた世界

予稿集





開催概要

- 開催名** OngaCREST シンポジウム 2016
「音楽情報処理研究が切り拓いた世界」
- 開催日** 2016年8月27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)
- 15:00-15:30 **招待講演「OngaCRESTプロジェクトの今後への期待」**
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 2016

音楽情報処理研究が 切り拓いた世界

予稿集

はじめに

OngaCREST シンポジウム 2016 「音楽情報処理研究が切り拓いた世界」にご来場いただき、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。2011年10月に始まった研究プロジェクトの最終年度になりましたので、2年前の「OngaCREST シンポジウム 2014」に引き続き、これまでの研究成果をみなさまにご紹介する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んでいただきながら、さまざまなご意見やアドバイスをいただき、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今後も推進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、「OngaCREST プロジェクトの今後への期待」と題した招待講演をしていただきます。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・エンジニアが実演しながらご紹介します。

我々の研究活動は、2011年に研究プロジェクト「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」が科学技術振興機構 (JST) 戦略的創造研究推進事業 CREST に採択され、多くの方々に支えていただくことで、これまで大きく発展して行くことができました。その結果、新たに5年間の研究プロジェクト「次世代メディアコンテンツ生態系技術の基盤構築と応用展開」が JST 戦略的創造研究推進事業 ACCEL に採択され、OngaCREST プロジェクトの研究成果を最大限に生かしてより一層インパクトの大きな技術の実現を目指した研究開発を、2016年8月から開始できることになりました。日ごろからご支援・ご議論いただ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いた世界をより一層豊かにして、さまざまな貢献をしていきたいと考えています。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。JST ACT-I「情報と未来」研究総括、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、43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24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 & Songle Widget: 楽曲の中身を自動解析する能動的音楽鑑賞サービスとそれに基づく音楽連動制御	12
Songrium: 膨大な音楽コンテンツを俯瞰的に鑑賞できる音楽視聴支援サービス	13
Songrium コメント分析: 視聴者の時刻同期コメントに基づく音楽動画検索支援サービス	14
TextAlive: 音楽に同期した歌詞アニメーション制作支援サービス	15
Songmash: マッシュアップ音楽の創作支援サービス	16
Songroid: 音楽に連動したダンスの分析・生成技術に基づくダンスアニメーション鑑賞サービス	17
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とその応用	18
Dancing Snap Shot System: 個性あるキャラクタ自動生成とダンス音楽動画創作支援システム	19
ユーザの好みを反映した 3D ダンス編集・創作支援	20
楽曲の歌詞情報を考慮した手話ダンス創作支援システム	21
物理ベースのキャラクタ 2 次動作生成と音楽シンクロキャラクタ生成	22
音楽と映像のクライマックスシンクロによる音楽動画コンテンツ創作支援	23
音楽に連動した個性的な手書きアニメーション生成による創作支援	24
音楽理解技術に基づくマイクロタスク回答可能な音楽ゲーム自動生成	25
ひとの視覚的特性を利用したエフェクトの付与による音楽動画鑑賞支援	26
音楽動画に対する印象評価データセットの構築	27
確率的生成モデルに基づく音楽音響信号解析	28
楽譜データに対する音楽言語モデルの教師なし学習	29
階層ベイズ音響・言語統合モデルに基づく多重基本周波数推定	30
準ビート同期 HMM に基づく歌声 F0 軌跡の自動採譜	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

OngaCREST プロジェクトの5年間では、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に取り組み、様々な鑑賞支援技術、創作支援技術、類似度・ありがち度の推定と音楽理解技術を研究開発してきた。次の5年間のOngaACCEL プロジェクトでは、次世代メディアコンテンツ生態系技術の基盤構築と応用展開により、コピー不可能な能動的体験による価値創出を目指していく。

1. OngaCREST プロジェクトの狙い

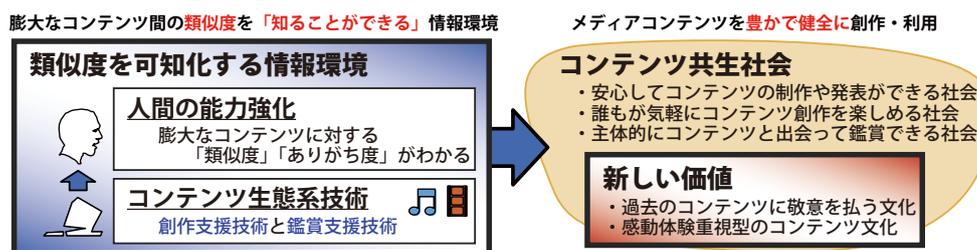
本研究プロジェクト「コンテンツ共生社会のための類似度を可視化する情報環境の実現」（プロジェクト名：OngaCREST（音画 CREST/ おんがくれすと）プロジェクト、科学技術振興機構（JST）戦略的創造研究推進事業 CREST、実施期間：2011年10月～2017年3月、研究代表者：後藤 真孝、研究分担者：森島 繁生、中村 聡史、吉井 和佳）では、図1のように、過去のコンテンツに敬意を払う文化を築きつつ、感動体験重視型の新たなコンテンツ文化が形成される社会の実現に寄与するために、膨大なコンテンツ間の類似度を人々が知ることができる（可視化する）情報環境を実現する研究に取り組んできた。そして、音楽の聴き

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）を対象としている。

2. OngaCREST シンポジウム 2016 で紹介する研究成果

OngaCREST プロジェクトでは、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

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



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

開発を進めてきた [1][2]。2年前の2014年8月23日に開催した公開シンポジウム「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」では、5年間の研究プロジェクトの3年目の中間段階での研究成果を紹介した。

そこで今回の公開シンポジウム「OngaCREST シンポジウム 2016: 音楽情報処理研究が切り拓いた世界」では、その後の2年間の研究成果を中心に紹介する。以下では、本シンポジウムの20件のポスター題名を『』に囲んで概要を説明する。本プロジェクトの前半では鑑賞支援技術の研究開発に重点的に取り組む計画だったため、その成果の多くは前回のシンポジウムで紹介済みだが、その後も様々な研究を実施してきたので、まず、2.1節でその一部を紹介する。次に、本プロジェクトの後半では創作支援技術の研究開発に重点的に取り組む計画だったため、今回のシンポジウムでは2.2節のようにその成果を一番多く紹介する。そして、それらと並行して持続的に取り組んできた類似度・ありがち度の推定と音楽理解技術の研究成果を2.3節で紹介する。

2.1 鑑賞支援技術

鑑賞支援技術に関してプレス発表をした最も代表的な成果は、『Songle & Songle Widget: 楽曲の中身を自動解析する能動的音楽鑑賞サービスとそれに基づく音楽連動制御』と『Songrium: 膨大な音楽コンテンツを俯瞰的に鑑賞できる音楽視聴支援サービス』の二つのWeb上のサービスである。OngaCREST シンポジウム 2014 の時点でも、既に一般公開して実証実験を開始していたが、その後も新機能の研究開発・サービスの改良を進めてきた。「Songle (そんぐる)」（<http://songle.jp>）は2012年8月29日に、「Songrium (そんぐりうむ)」（<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 が単一の楽曲の中身を対象とした技術であるのに対し、Songrium 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。

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 [3]。音楽連動制御のためのブラウザベース開発フレームワーク「Songle Widget」（<http://widget.songle.jp>）を提供したことで、

Songle が他の様々なサービスを可能にするプラットフォームとしてより一層発展してきた [4]。

一方、Songrium は Web マイニング技術と音楽理解技術を用いて、膨大な音楽コンテンツを俯瞰的に鑑賞できるサービスである [5][6]。「生成力ランキング」「コメント分析」「スマートプレーヤ」「歌声リレー再生」のような新機能を次々と実現・提供してきた。特に『Songrium コメント分析: 視聴者の時刻同期コメントに基づく音楽動画検索支援サービス』は、<http://comment.songrium.jp> で利用でき、コメントを利用した新たな音楽動画検索だけでなく、コメントに基づく再生位置の頭出し（シーク）も可能にした点が優れている。これは時刻同期コメントに基づく動画ランキングシステム「SmartVideoRanking」 [7] の研究成果を応用して実現した機能だが、他にも類似アーティストと動画多様性に基づく動画検索システム「ExploratoryVideoSearch」 [8] も実現した。

Songle Widget の活用例には、『音楽理解技術に基づくマイクロタスク回答可能な音楽ゲーム自動生成』 [9] もあり、音楽の中身に連動して自動生成された音楽ゲームをユーザが楽しみながら、選択肢型のマイクロタスクに取り組んで貢献できるようにした。また、『ひとの視覚的特性を利用したエフェクトの付与による音楽動画鑑賞支援』 [10] では、動画周囲や動画中央付近にエフェクト・図形を追加表示することで、音楽動画の視聴中に受ける印象を変化させる新しい取り組みをした。

2.2 創作支援技術

創作支援技術では、音楽コンテンツの自動生成技術を基礎技術として、非専門家に対する創作支援インタフェースを中心に、人間の能力を補いつつ創作の支援をするための研究開発を進めてきた。

創作支援技術に関してプレス発表をした最も代表的な成果は、『TextAlive: 音楽に同期した歌詞アニメーション制作支援サービス』 [11][12] で、Web上のサービス「TextAlive (てきすとあらいぶ)」（<http://textalive.jp>）として既に一般公開して実証実験を開始している。これは2015年9月8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。TextAlive では音楽理解技術とプログラミング環境技術を融合し、ユーザが楽曲を選んで動画の「スタイル」を指定するだけで、音楽理解技術であらかじ

め解析された発声タイミングやサビ情報などを利用して、様々な演出の歌詞アニメーションを制作可能にした。さらに、ユーザ自身もテンプレートやその編集用インタフェースを TextAlive の画面上でプログラミングし、他のユーザと共有できる特長を持つ。

『Songmash: マッシュアップ音楽の創作支援サービス』[13]と『Songroid: 音楽に連動したダンスの分析・生成技術に基づくダンスアニメーション鑑賞サービス』[14][15]では、OngaCREST シンポジウム 2016 に合わせて、Web 上のサービス「Songmash (そんぐまっしゅ)」(<http://songmash.jp>)と「Songroid (そんぐろいど)」(<http://songroid.jp>)の α 版の試験公開を開始した。Songmash は、複数の異なる楽曲を巧みに混ぜ合わせて自然に聞こえる「マッシュアップ音楽」に注目し、音楽理解技術に基づいてマッシュアップ度を算出して自動生成した結果を、自分好みの作品に変えられる創作支援サービスである。Songroid も最終的にはそうした創作支援サービスとなる予定だが、 α 版試験公開の段階では、自動生成されたダンスアニメーションを鑑賞をしながら、ダンス動作印象分析結果に基づいて再生位置の頭出しが可能なサービスとなっている。

以上の TextAlive、Songmash、Songroid は、いずれも最終的には、創作支援技術でありながら創作結果を鑑賞できる鑑賞支援技術にも位置付けられるようになる点が重要であり、これらは音楽発見(推薦)のための新たな手段としても活用できる。例えば、Songmash でマッシュアップ音楽作品を視聴しているときに、重ね合っている楽曲断片に興味を持ってその配信元サイトで音楽を楽しむことができ、音楽と出会う機会を増やすことが可能である。

音楽動画で代表的なダンスアニメーションに関しては、『Dancing Snap Shot System: 個性あるキャラクタ自動生成とダンス音楽動画創作支援システム』が、これまでに研究開発してきたダンス動画生成関連技術の集大成であり、ユーザの顔写真を入力すると三次元顔形状を復元し、本人に似たキャラクタのダンサーを音楽に合わせて踊らせることができる。また、『ユーザの好みを反映した 3D ダンス編集・創作支援』[16]では、適合フィードバックに基づくモーション検索アルゴリズムによって、局所的な振り付けの検索・選択をくり返すだけで好みのダンス制作を可能にする。『楽曲の歌詞情報を考慮した手話ダンス創作支援システ

ム』[17]では、手話とダンスを組み合わせたダンスモーションの生成手法を提案した。さらに、『物理ベースのキャラクタ 2 次動作生成と音楽シンクロキャラクタ生成』[18]では、より躍動感のあるダンス生成に向けて肉揺れアニメーションを実現した。

ダンス以外の、楽曲に連動した映像を対象とした創作支援技術として、『音楽と映像のクライマックスシンクロによる音楽動画コンテンツ創作支援』[19]では、映像とその BGM の楽曲の両者の盛り上がる箇所が一致するように、映像は固定して BGM を自動編集することを可能にした。また、『音楽に連動した個性的な手書きアニメーション生成による創作支援』[20]では、音楽動画中に自分らしさを加えるために、手書きアニメーション生成技術を実現した。

2.3 類似度・ありがち度の推定と音楽理解技術

類似度(「どれくらい似ているか」の度合い)・ありがち度(「どれくらいありふれているか」の度合い)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とその応用』では、5種類の音楽的要素の確率的生成モデルを学習し、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」を推定する枠組みを構築し[21]、さらにそれを改善した[22]。また、音楽プレイリストの可視化・並べ替えインタフェース「PlaylistPlayer」にも応用した[23]。

このような客観的な推定だけでなく、音楽動画から受ける主観的な印象の類似度も適切に扱っていくために、『音楽動画に対する印象評価データセットの構築』[24][25]では、500個の音楽動画に対して、音楽のみ、映像のみ、音楽と映像の組み合わせのそれぞれの場合に対する印象をラベル付けしたデータセットを構築し、公開した。そして、それらの相関を分析したり、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を評価したりした。

より深い音楽理解技術の実現を目指した『確率的生成モデルに基づく音楽音響信号解析』では、音楽の教師なし構造学習のための研究開発を進め、振幅スペクトログラムを音高・音色へ分解するための無限重畳離散全極モデル[26]や、複素スペクトログラムを基底へ分解する無限半正定値テンソル分解[27]を実現し

た。さらに、『楽譜データに対する音楽言語モデルの教師なし学習』[28]では、音符の背後にある音楽言語モデルの学習を目指して、楽譜から音型や変形パターンを学習したり、MIDI演奏のリズム変動をモデル化したり、音符列の構文木をPCFGで学習したりした。他にも、『階層ベイズ音響・言語統合モデルに基づく多重基本周波数推定』[29]では、音高とコードの両者をそれらの依存関係を考慮しつつ教師なしで同時推定する手法を実現し、『準ビート同期HMMに基づく歌声F0軌跡の自動採譜』[30]では、ビート時刻を既知として、連続的な歌声F0軌跡から離散的な音符系列を推定する手法を実現した。

3. OngaCRESTの次のOngaACCELでの挑戦

OngaCRESTプロジェクトの研究成果を最大限に生かしてより一層インパクトの大きな研究開発を進めるために、新たな研究プロジェクト「次世代メディアコンテンツ生態系技術の基盤構築と応用展開」(プロジェクト名: OngaACCEL (音画 ACCEL/ おんがあくせる) プロジェクト、科学技術振興機構 (JST) 戦略的創造研究推進事業 ACCEL、実施期間: 2016年8月～2021年3月、研究代表者: 後藤 真孝、プログラムマネージャー: 伊藤 博之、研究分担者: 森島 繁生、中村 聡史、吉井 和佳、佐々木 渉)を開始した。OngaACCELプロジェクトでは、これまでの成果を発展させて、図2のように、大規模な音楽コンテンツの蓄積に対する分析・合成を可能にする技術を実現することで、鑑賞・創作・協創の三つの側面から人々を支援するための基盤技術を構築する。さらにその応用展開として、鑑賞支援サービス、創作支援サービス、協創支援サービスなどの様々なサービスを実現可能にする「音楽コンテンツ生態系のサービスプラットフォーム」も構築する

ことで、次世代のメディアコンテンツ産業の発展に貢献することを目指していく。

OngaACCELプロジェクトでは、「コピー不可能な能動的体験」による価値創出を可能にする様々な技術を実現していく。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。そこでOngaACCELで取り組む鑑賞支援技術・サービスでは、OngaCRESTでも実現してきた「能動的音楽鑑賞」の能動性によりコピー不可能な体験を創出する。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となる。次に、創作支援技術・サービスでは、何かをクリエイト(創作)する創造的活動によりコピー不可能な体験を創出する。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。さらに、協創支援技術・サービスにより、他者と協力してコンテンツの新たな価値やコミュニティを創り出す行為を支援することで、コピー不可能な体験を創出する。

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

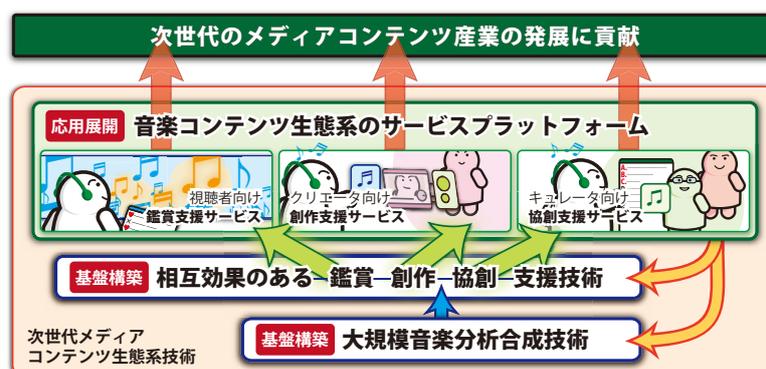


図2: OngaACCELプロジェクトの概要

(コンテンツ数が増える変化)が中心であった。次の段階は、「コピー不可能な能動体験」を情報技術の力で実現する質的な変化(体験の質が変わる変化)であり、それこそがデジタル化の本質である。そのような量的でなく質的な変化に基づくコンテンツの真価は、従来、まだ充分には引き出されていなかった。そこでOngaACCELプロジェクトでは、「コピー不可能な能動的体験」による価値創出が未来社会において本質的であることを認識しながら、多様な創作・鑑賞・協創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く。

参考文献

- [1] 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)
- [2] Masataka Goto: Frontiers of Music Information Research Based on Signal Processing, Proc. IEEE ICSP 2014.
- [3] 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.
- [4] Masataka Goto, Kazuyoshi Yoshii, and Tomoyasu Nakano: Songle Widget: Making Animation and Physical Devices Synchronized with Music Videos on the Web, Proc. IEEE ISM 2015.
- [5] 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.
- [6] Masahiro Hamasaki, Masataka Goto, and Tomoyasu Nakano: Songrium: Browsing and Listening Environment for Music Content Creation Community, Proc. SMC 2015.
- [7] 佃 洗撰, 濱崎 雅弘, 後藤 真孝: SmartVideoRanking: 視聴者の時刻同期コメントに基づく動画ランキングシステム, WebDB Forum 2015.
- [8] Kosetsu Tsukuda and Masataka Goto: ExploratoryVideoSearch: A Music Video Search System Based on Coordinate Terms and Diversification, Proc. IEEE ISM 2015.
- [9] 三輪 聡哉, 中村 聡史: マイクロタスク埋め込み型音楽ゲームの提案, 情処研報 2014-EC-34, 2014.
- [10] 松田 滉平, 中村 聡史: 動画に対する音響的装飾の分析と視覚的装飾を可能とする手法の提案, エンタテインメントコンピューティング 2015.
- [11] 加藤 淳, 中野 倫靖, 後藤 真孝: TextAlive: インタラクティブでプログラマブルな Kinetic Typography 制作環境, WISS 2014.
- [12] Jun Kato, Tomoyasu Nakano, and Masataka Goto: TextAlive: Integrated Design Environment for Kinetic Typography, Proc. ACM CHI 2015.
- [13] Matthew E. P. Davies, Philippe Hamel, Kazuyoshi Yoshii, and Masataka Goto: AutoMashUpper: Automatic Creation of Multi-Song Music Mashups, IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol.22, No.12, pp.1726-1737, 2014.
- [14] Satoru Fukayama and Masataka Goto: Automated Choreography Synthesis Using a Gaussian Process Leveraging Consumer-Generated Dance Motions, Proc. ACE 2014.
- [15] Satoru Fukayama and Masataka Goto: Music Content Driven Automated Choreography with Beat-wise Motion Connectivity Constraints, Proc. SMC 2015.
- [16] Ryo Kakitsuka, Kosetsu Tsukuda, Satoru Fukayama, Naoya Iwamoto, Masataka Goto, and Shigeo Morishima: A Choreographic Authoring System for Character Dance Animation Reflecting a User's Preference, Proc. ACM SCA 2016 (poster).
- [17] Wakana Asahina, Naoya Iwamoto, Hubert P.H. Shum, and Shigeo Morishima: Automatic Dance Generation System Considering Sign Language Information, ACM SIGGRAPH 2016 (poster).
- [18] 岩本 尚也, 森島 繁生: キャラクターの身体構造を考慮した実時間肉揺れ生成手法, 画像電子学会誌, Vol.44, No.3, pp.502-511, 2015.
- [19] Haruki Sato, Tatsunori Hirai, Tomoyasu Nakano, Masataka Goto, and Shigeo Morishima: A Soundtrack Generation System to Synchronize the Climax of a Video Clip with Music, Proc. IEEE ICME 2016.
- [20] 佐藤 剣太, 中村 聡史, 鈴木 正明: 個性豊かな手書きメッセージ生成手法の提案, エンタテインメントコンピューティング 2015.
- [21] Tomoyasu Nakano, Kazuyoshi Yoshii, and Masataka Goto: Musical Similarity and Commonness Estimation Based on Probabilistic Generative Models of Musical Elements, International Journal of Semantic Computing, Vol.10, No.1, pp.27-52, 2016.
- [22] Tomoyasu Nakano, Daichi Mochihashi, Kazuyoshi Yoshii, and Masataka Goto: Musical Typicality: How Many Similar Songs Exist?, Proc. ISMIR 2016.
- [23] Tomoyasu Nakano, Jun Kato, Masahiro Hamasaki, and Masataka Goto: PlaylistPlayer: An Interface Using Multiple Criteria to Change the Playback Order of a Music Playlist, Proc. ACM IUI 2016.
- [24] 大野 直紀, 中村 聡史, 山本 岳洋, 後藤 真孝: 音楽動画への印象評価データセット構築とその特性の調査, 情処研報 2015-MUS-108, 2015.
- [25] 土屋 駿貴, 大野 直紀, 中村 聡史, 山本 岳洋: ソーシャルコメントからの音楽動画印象推定手法の提案, DEIM 2016.
- [26] Kazuyoshi Yoshii, Katsutoshi Itoyama, and Masataka Goto: Infinite Superimposed Discrete All-Pole Modeling for Multipitch Analysis of Wavelet Spectrograms, Proc. ISMIR 2015.
- [27] Kazuyoshi Yoshii, Katsutoshi Itoyama, and Masataka Goto: Student's t Nonnegative Matrix Factorization and Positive Semidefinite Tensor Factorization for Single-Channel Audio Source Separation, Proc. ICASSP 2016.
- [28] Eita Nakamura, Katsutoshi Itoyama, and Kazuyoshi Yoshii: Rhythm Transcription of MIDI Performances Based on Hierarchical Bayesian Modelling of Repetition and Modification of Musical Note Patterns, Proc. EUSIPCO 2016.
- [29] Yuta Ojima, Katsutoshi Itoyama, and Kazuyoshi Yoshii: A Hierarchical Bayesian Model of Chords, Pitches, and Spectrograms for Multipitch Analysis, Proc. ISMIR 2016.
- [30] Ryo Nishikimi, Katsutoshi Itoyama, and Kazuyoshi Yoshii: Musical Note Estimation for FO Trajectories of Singing Voices based on a Bayesian Semi-beat-synchronous HMM, Proc. ISMIR 2016.

 OngaCREST シンポジウム 2016

音楽情報処理研究が 切り拓いた世界

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

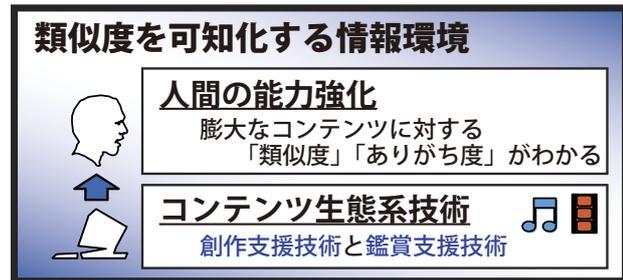


研究分担者：吉井 和佳 (京都大学 講師)

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。

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)

OngaCREST：これまでの成果

プロジェクト開始時点での想定を上回る研究成果を創出

- ・JST 戦略的創造研究推進事業 CREST に採択されなければ困難だった大きな挑戦ができた
- ・学術的に深みのある基礎技術と社会的インパクトのある応用技術をバランス良く研究開発してきた
- ・研究成果を Web 上のサービス等の形で社会に展開しユーザーや産業界の声を聞きながらプロジェクトを展開

鑑賞支援技術

- ・音楽理解技術等を活用した Songle + Songle Widget、Songrium を公開
- ・能動的音楽鑑賞のための様々なインターフェース・機能を実現

創作支援技術

- ・音楽理解技術等を活用した TextAlive、Songmash、Songroid を公開
- ・人間の能力を補いつつ創作の支援をするインターフェース・機能を実現

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類似度・ありがち度の推定を実現して応用
- ・統計的機械学習等に基づく様々な音楽音響信号解析技術を実現

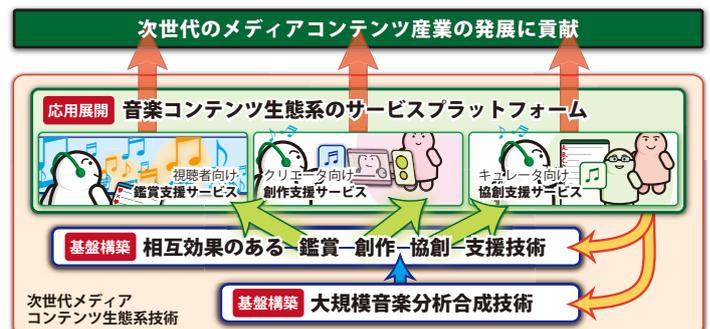
OngaACCEL：これからの挑戦

「次世代メディアコンテンツ生態系技術の基盤構築と応用展開」

- ・5年間の新プロジェクトがJST 戦略的創造研究推進事業 ACCEL に採択
- ・OngaCREST プロジェクトの成果を最大限に生かして2016年8月から開始

コピー不可能な能動的体験による価値創出の実現を狙う

- ・大規模なメディアコンテンツの蓄積に基づいて新たな価値を生む技術
- ・新たなコンテンツを生み出し育む生態系を実現するための技術



OngaCREST シンポジウム 2016 で紹介する研究成果

鑑賞支援技術

Songle & Songle Widget: 楽曲の中身を自動解析する能動的音楽鑑賞サービスとそれに基づく音楽連動制御

後藤 真孝, 川崎 裕太, 井上 隆広, 吉井 和佳, 中野 倫靖

Songrium: 膨大な音楽コンテンツを俯瞰的に鑑賞でき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, 佃 光摂, 深山 覚

Songrium コメント分析: 視聴者の時刻同期コメントに基づく音楽動画検索支援サービス

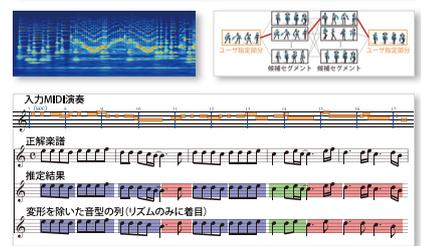
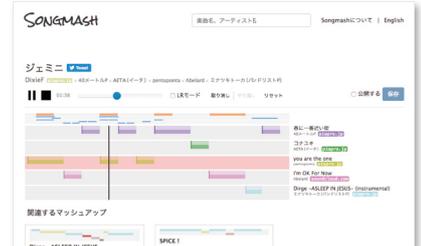
佃 光摂, 石田 啓介, 濱崎 雅弘, 後藤 真孝

音楽理解技術に基づくマイクロタスク回答可能な音楽ゲーム自動生成

中村 聡史, 三輪 聡哉, 十文字 優斗

ひとの視覚的特性を利用したエフェクトの付与による音楽動画鑑賞支援

松井 啓司, 田村 証優紀, 松田 滉平, 福地 翼, 山浦 祐明, 大島 遼, 中村 聡史



創作支援技術

TextAlive: 音楽に同期した歌詞アニメーション制作支援サービス

加藤 淳, 中野 倫靖, 後藤 真孝

Songmash: マッシュアップ音楽の創作支援サービス

後藤 真孝, 川崎 裕太, Matthew Davies

Songroid: 音楽に連動したダンスの分析・生成技術に基づくダンスアニメーション鑑賞サービス

深山 覚, 井上 隆広, 後藤 真孝

Dancing Snap Shot System: 個性あるキャラクタ自動生成とダンス音楽動画創作支援システム

岩本 尚也, 谷田川 達也, 加藤 卓哉, Savkin Pavel, 福原 吉博, 中塚 貴之, 野沢 絵佐, 柿塚 亮, 山口 周吾, 朝比奈 わかな, 森島 繁生

ユーザの好みを反映した 3D ダンス編集・創作支援

柿塚 亮, 佃 光摂, 深山 覚, 岩本 尚也, 後藤 真孝, 森島 繁生

楽曲の歌詞情報を考慮した手話ダンス創作支援システム

朝比奈 わかな, 岩本 尚也, Hubert P.H. Shum, 森島 繁生

物理ベースのキャラクタ 2 次動作生成と音楽シンクロキャラクタ生成

岩本 尚也, Hubert P.H. Shum, Longzhi Yang, 森島 繁生

音楽と映像のクライマックスシンクロによる音楽動画コンテンツ創作支援

佐藤 晴紀, 平井 辰典, 中野 倫靖, 後藤 真孝, 森島 繁生

音楽に連動した個性的な手書きアニメーション生成による創作支援

佐藤 剣太, 新納 真次郎, 久保田 夏美, 斉藤 絢基, 中村 聡史, 鈴木 正明, 小松 孝徳

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とその応用

中野 倫靖, 持橋 大地, 吉井 和佳, 後藤 真孝

音楽動画に対する印象評価データセットの構築

土屋 駿貴, 大野 直紀, 阿部 和樹, 中村 聡史, 山本 岳洋, 後藤 真孝, 濱崎 雅弘

確率的生成モデルに基づく音楽音響信号解析

吉井 和佳, 中村 栄太, 糸山 克寿, 後藤 真孝

楽譜データに対する音楽言語モデルの教師なし学習

中村 栄太, 糸山 克寿, 吉井 和佳

階層ベイズ音響・言語統合モデルに基づく多重基本周波数推定

尾島 優太, 中村 栄太, 糸山 克寿, 吉井 和佳

準ビート同期 HMM に基づく歌声 F0 軌跡の自動採譜

錦見 亮, 中村 栄太, 糸山 克寿, 吉井 和佳

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 & Songle Widget:

楽曲の中身を自動解析する能動的音楽鑑賞サービスとそれに基づく音楽連動制御

後藤 真孝 川崎 裕太 井上 隆広 吉井 和佳*1 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
100万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)
- ・解析誤りを候補選択等により訂正して保存・共有可能なインターフェース
- ・サビ出し機能により楽曲中のサビ区間や繰り返し区間への頭出しが可能

Songle Widget とは

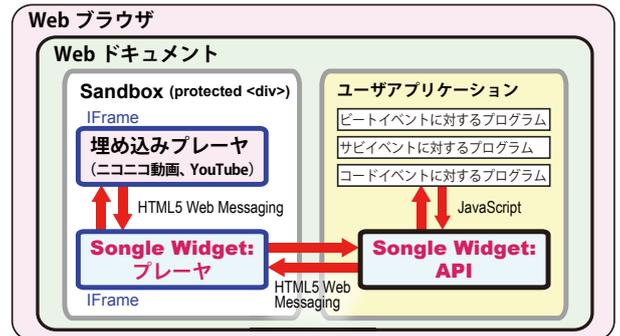
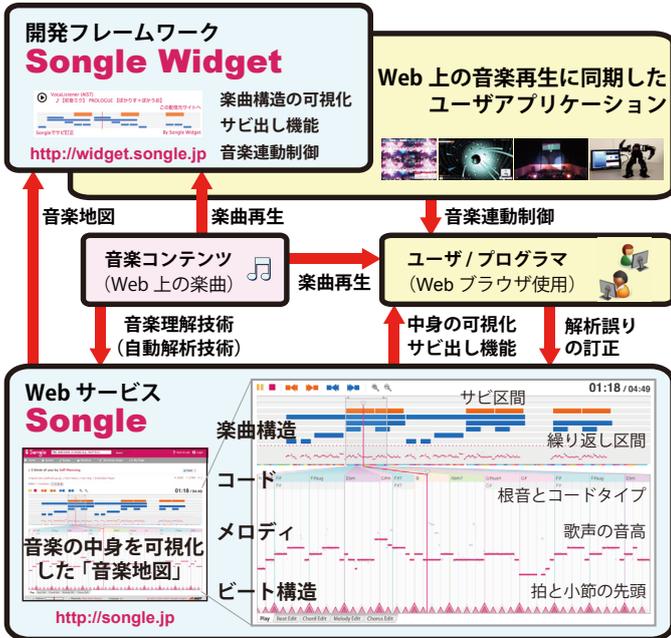
音楽連動制御のためのブラウザベース開発フレームワーク

Songle Widget (<http://widget.songle.jp>)

- ・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
- ・コンピュータグラフィックス、物理デバイス(ロボット)等を Web 上の音楽に同期させて制御
- ・Songle に連動するため、Songle 上で解析誤りを訂正すればより正確に同期

Songle Widget (外部埋め込みプレーヤ) の Web API を提供中

- ・楽曲構造の可視化機能、サビ出し機能、音楽連動制御機能付きのプレーヤ
- ・JavaScript API の詳細ドキュメントと  Vocalist (AST) > [初音ミク] PROLOGUE [ほかりす+ほかうお] この配信元サイトへ サンプルプログラムを公開中 



利用例: Songrium3D (<http://songrium.jp/map/3d/>)

- ・音楽に連動した三次元アニメーションで楽曲を表現する鑑賞支援サービス



利用例: TextAlive (<http://textalive.jp>)

- ・音楽に連動した歌詞アニメーションの創作支援サービス



利用例: Photo x Songle (<http://photo.songle.jp/>)

- ・Web 上の写真で音楽を彩る鑑賞支援サービス

利用例: V-Sido x Songle (アストラテック社「V-Sido OS」と連携)

- ・音楽に合わせてロボットを踊らせる制御システム
- ・音楽地図に基づいて動的に振り付けを
変えながら複数のロボットがダンス可能

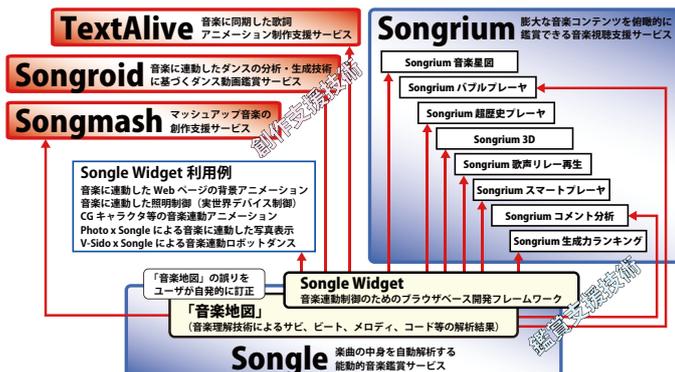


Songle の目指す世界

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

サービスプラットフォーム

Songle がプラットフォームとして様々なサービスを可能に



謝辞: 藤原 弘将氏、Matthias Mauch 氏 (音楽理解技術等)、櫻井 稔 氏 (Web デザイン)

Songrium: 膨大な音楽コンテンツを俯瞰的に鑑賞でき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 佃 洸撰 深山 覚
産業技術総合研究所

Songrium(ソングリウム)とは

音楽視聴支援サービス Songrium (<http://songrium.jp>)

- 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さまざまな可視化インタフェース

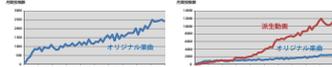


Web-native Music

Webで発表、共有、視聴され、派生作品が生まれる音楽

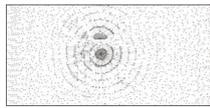
- すべてがWeb上にあるため誰でも聴ける、統計や機械処理に向く
- ニコニコ動画上で広がりを見せる VOCALOID 音楽はその最先端事例

ニコニコ動画上の VOCALOID オリジナル楽曲は 14 万曲以上、その派生動画は 59 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

- 膨大な音楽コンテンツを俯瞰的に鑑賞できる音楽視聴支援サービス Songrium

Songrium の様々な機能

音楽星図

膨大なオリジナル楽曲を、音響特徴量の類似関係をもとに平面配置。マウスによるパン・ズーム操作で14万曲の楽曲を俯瞰できる。惑星ビューによる派生作品の俯瞰や、サビから再生を用いたコンテンツ視聴もできる。

類似度に基づく楽曲の平面配置

音楽音響信号から人が受ける印象と時代性(作品公開日)を組み合わせた類似度尺度で平面を構成し楽曲を配置。

惑星ビュー

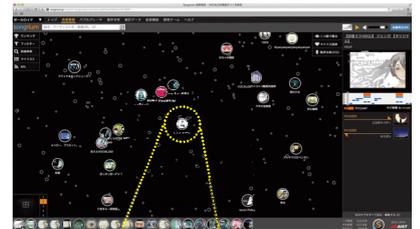
派生動画の様子が俯瞰でき、さらにそこからオリジナル楽曲の特徴も知ることができる。

矢印タグ

音楽コンテンツ間の関係に名前を付ける。新しいソーシャルタギング。

YouTube バージョン

YouTubeでも音楽星図を利用できる。国内外のミュージックビデオ作品を楽しめる。



Songrium3D

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。



超歴史プレーヤ

膨大な音楽コンテンツを、時間軸にそって俯瞰的に鑑賞できる。楽曲や派生作品を投稿日順にクロスフェード再生。タッチ操作に対応したユーザーインタフェースで興味のある時期に移動したり、カテゴリを絞り込んで楽しめる。



新機能その1

生成力ランキング 2016年8月リリース

派生作品が多く作られる=生成力がある楽曲を見つけ出せる新しい楽曲ランキング。どのカテゴリの派生作品が多く生まれたかも見え、派生作品から元作品の特徴が浮かび上がる。



新機能その2

コメント分析 2016年5月リリース

たくさんの動画コンテンツの中から、「かわい」や「高音綺麗」、「イントロ超かっこいい」など思わず作品を見たくなるようなコメントが投稿されているものを探ることができる。



新機能その3

スマートプレーヤ 2016年2月リリース

投稿期間やマイリスト数、ジャンルタグ等の条件で絞り込み、サビだけ連続再生(サビから再生開始・サビが終われば次の曲へ)ができる。新曲が毎日たくさん増えても手軽に試聴。



新機能その4

歌声リレー再生 2015年7月リリース

同曲異唱コンテンツをつなぎ変えて連続再生する「リレー再生」で楽曲を聞ける。一つの楽曲を多様な歌声で楽しんだり、歌声をきっかけに新しい楽曲と出会ったりすることができる。



Songrium コメント分析：

視聴者の時刻同期コメントに基づく音楽動画検索支援サービス

佃 洸 石田 啓介 濱崎 雅弘 後藤 真孝

産業技術総合研究所

Songrium コメント分析とは

時刻同期コメントを利用した音楽動画検索支援サービス

<http://comment.songrium.jp>

- コメントを利用することで、従来よりも直感的に音楽動画を検索できるようにするサービス
- 機械学習による有用なコメントの自動抽出

コメントの有用度推定

コメントを正規化したうえで特徴量を計算

1. クエリ関連度 $P(q, c) \cdot \log \frac{P(q, c)}{P(q)P(c)}$	2. クエリ類似度 $1 - \frac{D_L(q, c_T)}{L_{max}(q, c_T)}$	3. 文字数 かわいい → 4
4. 出現頻度 12 5 33 9 → 928	5. 元コメントの種類数 かわい → [かわいいいいい!!]	6. 形容詞の有無 かわいい → ○
7. 形容動詞の有無 かわいい → ×	8. 投稿時刻のエントロピー	9. サビ区間の出現確率
10. 類似コメント種類数 かわい → [かわええ かわいすぎる かわず]	11. 類似コメ+出現頻度 19 11 86 23 → 2591	12. 類似コメ+エントロピー
13. 類似コメ+サビ出現確率	14. 文字バイグラム かわいい → かわ+わい	本サービスでは特に効果の高かった特徴量 3 と 14 を使用

有用度の推定精度

- 8名の評価者が1万件のコメントに対して有用度を3段階でラベル付け
- 有用度の正解スコアと推定スコアの相関係数 0.753 を達成

今後の展望

- ユーザの検索ログを利用することで検索に有用なコメントの教師データを動的に更新
- ゲーム実況等の音楽動画以外のジャンルへの適用

Songrium コメント分析の機能



コメント推薦

機械学習によって推定されたおすすめのコメが表示される。「更新」ボタンを押すと次々に新しいコメントが推薦される。



コメント検索

画面左上のテキストボックスに文字を入力するとコメント候補のリストが表示される。コメントを選択することで検索結果が表示される。



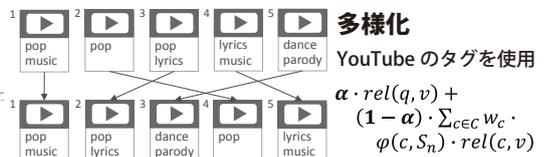
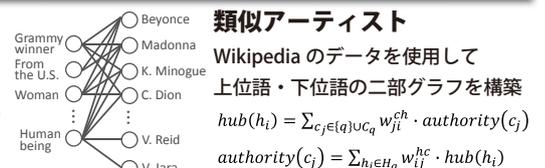
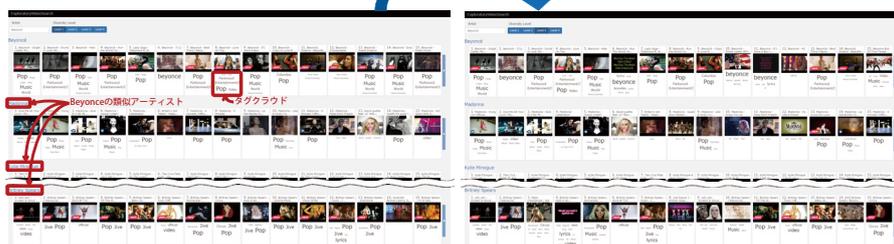
コメントシークバー

動画内でコメントが盛り上がっている時間帯を表す。グラフの山をクリックすることで、盛り上がりシーンへのジャンプが可能。シークバーの下のコメをクリックすると、そのコメントの盛り上がりシーンを順にジャンプすることも可能。



ExploratoryVideoSearch: 類似アーティストと動画多様性に基づく動画検索

さまざまな種類の動画が見られるように検索結果を多様化



TextAlive: 音楽に同期した歌詞アニメーション制作支援サービス

加藤 淳 中野 倫靖 後藤 真孝
産業技術総合研究所

歌詞アニメーションとは

音楽と共に歌詞の内容を魅力的に伝える動画の演出手法

- 音楽と同期して歌詞がアニメーションする **Kinetic Typography** の一種
- 従来、**タイミング調整と効果的な動きの設計**を文字ごとにする必要があった
- 歌詞カードや単なるカラオケ表示と比べて、**制作に膨大な手間が必要**だった

問題 魅力的だが、制作が難しい & 手間が膨大



TextAlive (テキストアライブ) とは

みんなで創作を楽しめる、ウェブブラウザ上で動く歌詞アニメーション制作支援サービス

[ACM CHI2015] Best Paper Honorable Mention Award
[IJCLC 2015], [LIVE 2016] 著者はすべて Kato+

サービス連携



音楽理解技術に基づく歌詞アニメーションの自動生成

プログラミング環境技術に基づく動画のリアルタイム合成・編集



- 音楽理解エンジンで解析したサビやビートなどの情報と、歌詞同期エンジンで解析した歌詞の発声タイミングの情報を用い、歌詞アニメーションを自動生成
- 歌詞発声タイミングの誤りは、ブラウザ上でインタラクティブに訂正できる
- 曲調や歌詞などに合わせて歌詞アニメーションのスタイルを選択して楽しめる
- TextAlive上の歌詞アニメーション(動画)はプログラミング環境技術でブラウザ上のプログラムを動的に変更、実行してリアルタイムに描画される
- 既存の動画共有サイトと異なり、見ている動画は全てブラウザ上で編集できる
- 動画の演出に用いるアルゴリズムを動画を再生しながら編集できるライブプログラミング機能を搭載し、プログラマが機能を拡張できる統合開発環境を実現
- プログラマが作ったテンプレートをユーザが調整できるライブチューニング機能により、プログラマとユーザが協力できるソーシャルな制作支援環境を実現

解決 音楽理解技術とプログラミング環境技術で
手間いらずの、もっと自由な動画制作環境を実現

謝辞: 上記の図では、ピアプロ・キャラクター・ライセンスに基づいてクリプトン・フューチャー・メディア株式会社のキャラクター「初音ミク」を使用しました。

Songmash: マッシュアップ音楽の創作支援サービス

後藤 真孝 川崎 裕太 Matthew Davies

産業技術総合研究所

Songmash (ソングマッシュ) とは

マッシュアップ音楽の創作支援サービス

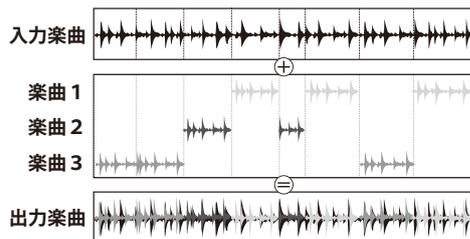
Songmash (<http://songmash.jp>) [α 版を試験公開中]

- 音楽を混ぜ合わせて楽しむことができるサービス
- 複数の異なる楽曲を巧みに混ぜ合わせて自然に聞こえるように音楽を制作する「マッシュアップ」と呼ばれる音楽制作手法を支援
- ピアプロ、SoundCloud に対応
- 8千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)
- 既存のコンテンツ群の力を借りた創作が可能になる



マッシュアップ音楽の創作方法

- 入力楽曲(既存楽曲)に他の楽曲の断片を重ね合わせて出力楽曲を得る
- コード進行が調和してテンポが同期するように楽曲の断片を加工
時間長を変えずに音高変更(ピッチシフト)
音高を変えずに時間伸縮(タイムストレッチ)
- 区間毎に多様な楽曲から断片が次々に選ばれて同時再生される



Songmash の様々な機能

マッシュアップの自動生成機能

- ユーザが好きな楽曲を選ぶと自動的にマッシュアップが生成される

マッシュアップの編集機能(自分好みのマッシュアップの創作支援)

- 自動生成結果で変えたい箇所を楽曲候補から選択して置き換えられる
どの候補がどの時刻に配置可能かは自動的に決定される
- 各楽曲断片のフェードイン・フェードアウトの指定が可能

マッシュアップ作品の公開機能

- 編集したマッシュアップ作品を公開することができる
- 公開されたマッシュアップを誰でも楽しめ、派生作品を制作できる
- URLを知っている人だけがアクセスできる限定公開も可能

Songmash の実現方法

「マッシュアップ度(mashability)」に基づく自動生成技術

- ある入力楽曲を基準として、それに混ぜることが可能な楽曲とそのフレーズの断片を探索し、音の高さとテンポを自動調整して混ぜ合わせる
- 独自のマッシュアップ度(マッシュアップ可能性)の算出方法を提案
ビートとフレーズの解析後に楽曲断片間の類似度を以下に基づき計算
(1) ハーモニーの類似度(調和度)
(2) リズムの類似度
(3) 周波数特性のバランス(スペクトルバランス)
- 元のフレーズ断片の状態ではマッシュアップ度が低くても、移調とテンポ変更によってどのように自動調整すればマッシュアップ度を最大化して自然に聞こえるようになるかを推定

[M. E. P. Davies, P. Hamel, K. Yoshii, and M. Goto: AutoMashUpper: Automatic Creation of Multi-Song Music Mashups, IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol.22, No.12, pp.1726-1737, December 2014.]

Web ブラウザ上でインタラクティブな編集を可能にする技術

- 楽曲配信元サイトから直接音楽再生をする際に、ユーザのブラウザ上でリアルタイムにピッチシフト・タイムストレッチができるエンジンを実装
- マッシュアップ度だけサーバ側で事前計算しておき、それ以外の多様な編集をブラウザ上で実現できるように実装



Songmash の目指す世界

新たな音楽発見(推薦)手段としてのマッシュアップ鑑賞・創作

- 重ね合わせたすべての楽曲断片をリファアーすることで、視聴者が楽曲断片に興味を持った際に、それらの配信元サイトへ移動して音楽との出会いを拡大
- 楽曲断片提供による新たな音楽プロモーション手段への発展可能性を模索

2次創作文化の発展に寄与

- 従来は手作業での膨大な試行錯誤が必要だったマッシュアップ制作を支援
- 自分好みにパーソナライズすることでコピー不可能な能動的な体験を提供

ポピュラリティを確保したまま人間が快適に受容可能な複雑度の限界を押し上げることによる音楽の進化の可能性

- 人々が快適に感じる音楽の複雑度には限界があり、従来は、複雑度を増加すると理解する難易度が高くなりポピュラリティを得ることが困難であった
- マッシュアップは、聴取者の頭の中に既にある音楽の記憶を参照することによって、ポピュラリティを確保したまま、人間が快適に受容可能な複雑度の限界を押し上げられる可能性がある
- 電子楽器以前の個々の楽器音を単位とした音楽制作から発展して、技術の進歩により数小節のループ素材を単位とした音楽制作が可能となったが、マッシュアップではさらに進んで、楽曲を単位・素材とした音楽制作を可能にする点で新しく、一から作曲しているときには作り出そうと思えないような複雑な音響信号に到達することが容易になっている

Songroid:

音楽に連動したダンスの分析・生成技術に基づくダンスアニメーション鑑賞サービス

深山 覚 井上 隆広 後藤 真孝
産業技術総合研究所

Songroid (ソングロイド) とは

音楽に連動した CG キャラクターの

ダンスアニメーション鑑賞サービス

Songroid (<http://songroid.jp>) [α版を試験公開中]

- ・三次元 CG キャラクターが踊るアニメーションを見て楽しめるサービス
能動的音楽鑑賞サービス Songle と連携してダンスを自動生成
キャラクターと背景を Web ブラウザ上でリアルタイムにレンダリング
- ・ダンス分析技術に基づいて好みのダンスを探して楽しめる
「すごい」「かわいい」「カッコいい」動作を探せるシーク機能と検索機能
動作に対してどのようなコメントが付きやすいかを自動推定



ダンス動作の印象を表わすコメントの予測と検索

- ・「すごい」「かわいい」「カッコいい」のコメントが付きそうな動作を推定



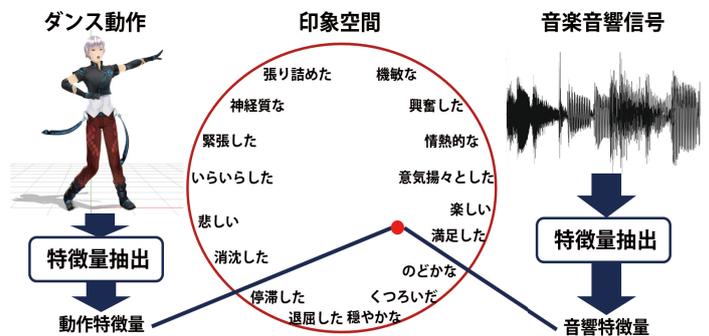
- ・ダンス 1 曲を通じたコメント予測に基づいたダンス動作検索
- 「すごい」「かわいい」「カッコいい」動作で
ソートされた自動生成ダンスを一覧表示



ダンスの分析・生成技術

印象空間を通じたダンス動作の分析と生成

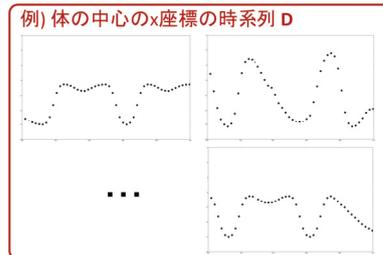
- ・機械学習によってダンス動作と音楽を共通の印象空間に対応付けて分析
- ・与えられた音楽の印象に最も近い印象を与えるダンスを生成



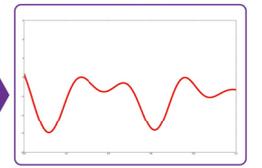
確率モデルで特徴的なダンス動作を機械学習

- ・ガウシアンプロセス回帰に基づいて、似た特徴をもつ複数のダンス動作から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

類似するダンス動作のクラスタ



学習されたダンス語彙



$$P(\mathbf{D}|\mathbf{t}; \sigma, \lambda) = \frac{1}{\sqrt{(2\pi)^{4|t|} |\mathbf{K}|^{|t|}}} \exp\left(-\frac{1}{2} \text{tr}(\mathbf{D}^T \mathbf{K}^{-1} \mathbf{D})\right)$$

ガウシアンプロセスで学習

Songroid の今後

ダンスアニメーション創作支援サービスへの発展

- ・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楽曲の内容に連動し構成をもつダンスを自動生成

※ 上記の図で用いている三次元 CG キャラクターには、OngaCREST プロジェクトが研究目的で制作したオリジナルキャラクターを使用しました。

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とその応用

中野 倫靖 持橋 大地*1 吉井 和佳*2 後藤 真孝

産業技術総合研究所 *1 統計数理研究所 *2 京都大学

研究背景

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

- 音楽がデジタル化されアクセス可能な楽曲が単調増加
- 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
- 安心して楽曲の制作や発表をしにくい社会になりかねない
- 類似度に関する人間の能力には限界がある
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

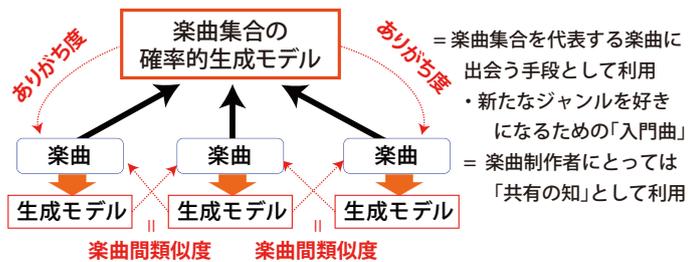
「何が似ているのか」「どれくらいありふれているのか」を知ろうと思ったときに容易に知るための手段の実現

- 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
例) 論文のように引用(再利用)を可視化し喜びを感じられる音楽文化
- 「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榮し、人々を感動させる魅力や完成度の高さ等を重視

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(音響特徴量等)がどういった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



応用

PlaylistPlayer: 音楽プレイリストの可視化・並べ替えインターフェイス [Nakano+, ACM IUI 2016]



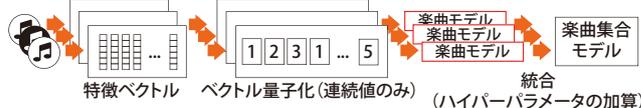
- 使用例)
- リズムが「ありがち」な曲から順に再生したい
 - 前後の曲で音色が類似した順で再生したい
 - ボーカルを男女交互に再生したい

楽曲・楽曲集合モデル

[Nakano+, IJSC 2016]

潜在的ディリクレ配分法(Latent Dirichlet Allocation: LDA)

- ボーカルの歌声: 線形予測メルケプストラム係数(LPMCC)、 $\Delta F0$
- 楽曲中の音色: メル周波数ケプストラム係数(MFCC)、 $\Delta MFCC$ 、 Δ パワー
- リズム: Fluctuation Pattern (FP) ・歌詞: 日本語形態素(原形)、英単語

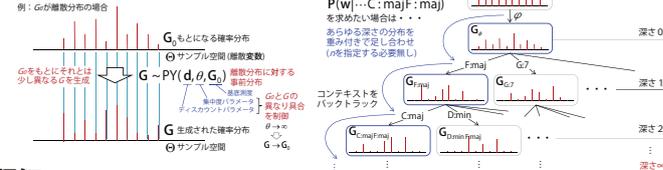


可変長 Pitman-Yor 言語モデル (VPYLM)

- 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)

Pitman-Yor過程

- PY: 確率分布上の確率分布
- ディリクレ過程(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



評価

- 類似度 (similarity): 人間(ミュージシャン1名)の評価結果と比較
 - 日本語楽曲20組を評価(類似度が高いと推定された10組、低い10組)
 - ⇒ 類似度が高い曲の評価と低い曲の評価に有意差 ($p < 5\% \sim 0.1\%$)
- ありがち度 (typicality): 「類似度が高い曲の数」との相関 [Barsalou 1985]
 - 認知心理学分野における「典型性」(Central tendency = 類似した曲の数)
 - 日本語楽曲3,278曲(歌詞に関しては1,996曲分)
 - ⇒ 類似度の高い曲の数と相関(0.677 ~ 0.834)を確認
 - ※英語楽曲415曲を用いた評価も行い、有効性を確認

LDA 楽曲集合モデルの改善

[Nakano+, ISMIR 2016]

改善点1: 情報理論に基づくありがち度の計算

- タイプを共有する系列の集合の確率 = KL 尺度の指数の期待値 → ありがち度
- 楽曲のタイプ: 各楽曲におけるトピック系列のユニグラム分布
- 情報源のタイプ: ディリクレ分布(楽曲集合モデル)に従う多項分布



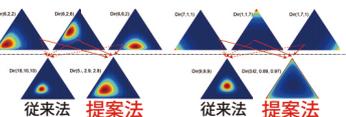
改善点2: 楽曲集合モデル(楽曲モデルの統合)をベイズ推定

- 楽曲集合が各曲のタイプを生成するディリクレ分布を持つとする
- ⇒ 全曲分のタイプ(多項分布の集合)から事前分布(ディリクレ分布)を推定

多項分布の集合

※楽曲トピック分布 = ディリクレ分布 → 期待値

推定されたディリクレ分布



ありがち度推定の評価: ボーカルの男女比率を変えた楽曲集合モデル

- 提案法で推定した各楽曲のありがち度が、楽曲集合の男女比変化を最も反映

Dancing Snap Shot System:

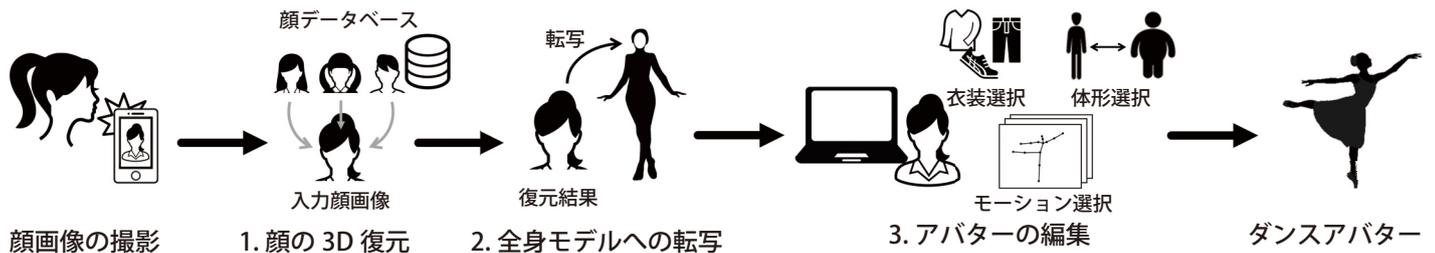
個性あるキャラクタ自動生成とダンス音楽動画創作支援システム

岩本尚也 谷田川達也 加藤卓哉 Savkin Pavel 福原吉博 中塚貴之
野沢 綸佐 柿塚亮 山口周吾 朝比奈わか な 森島 繁生

早稲田大学

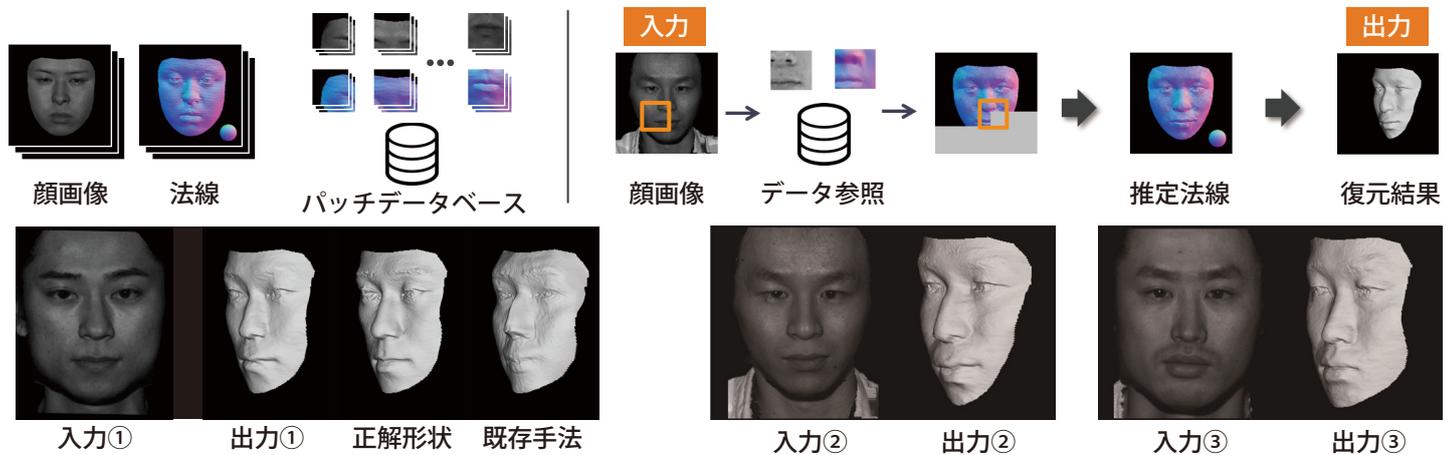
概要

顔写真1枚の入力から本人そっくりの“踊るアバター”を作成・編集できるプラットフォーム



1. 顔の 3D 復元

データベースから似た陰影を持つ箇所の法線を参照することで、3D スキャナと同等精度で顔形状を復元
事前準備



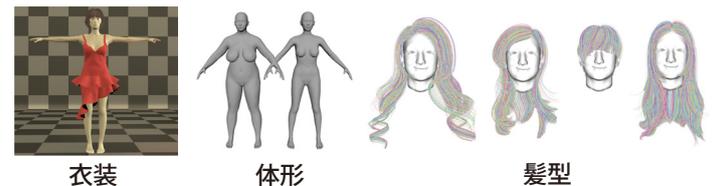
2. 全身モデルへの転写

復元結果を再現するため、全身モデルの顔領域を変形、境界を滑らかに変形



3. アバターの編集

衣装や、体形、髪型などを選ぶと、生成したアバターにマッチした動きを自動生成



今後の展望

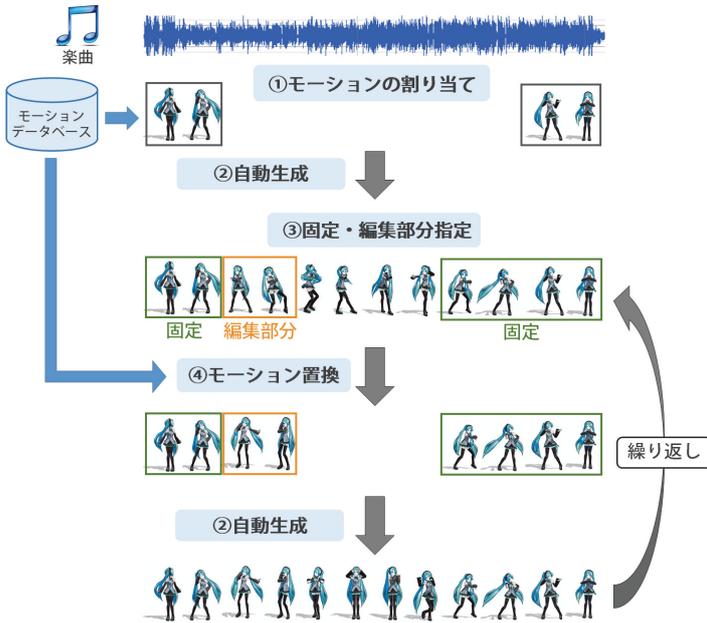
より高精度かつ手軽な撮影とモデリングの実現、インタラクションなどを加えたシステムの汎用化

ユーザの好みを反映した 3D ダンス編集・創作支援

柿塚亮 佃洗撰*1 深山覚*1 岩本尚也 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

ユーザの好みを反映した
キャラクタダンスアニメーション制作システム



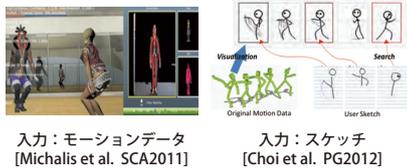
はじめに

提案システムにおける技術的課題

1. ユーザがどのようにデータベースから好みの振付を探すか？
2. ユーザが指定した振付間のダンスをどのように自動生成するか？

関連研究

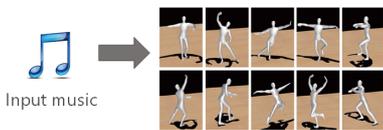
1. モーション検索に関する研究



ユーザの負担が大きい

振付の知識が必要

2. 音楽にシンクロしたダンスの自動生成に関する研究



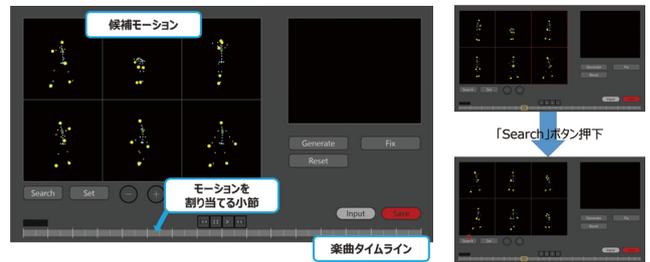
[Shiratori et al. EG2006]

ユーザ指定振付間の
ダンス生成手法が必要

1. ダンスモーション検索

好みの振付の探索方法

- ユーザが指定した楽曲中の任意の部分について、候補の振付となるモーションデータを複数提示し、ユーザは楽曲と合わせてダンスをプレビューしながら、探索的に好みの振付を検索
- ユーザは**適合フィードバック**に基づくモーション検索アルゴリズムにより、提示される候補モーションを変更可能

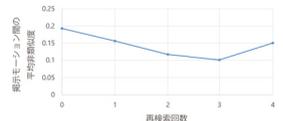


提示される候補モーション

- 好みの振付と出会う可能性を高めるため、検索の初期段階では提示モーション間の非類似度は大きく、検索を繰り返すことでユーザの好みに近いダンスモーションに**探索範囲が絞られていく**
- 編集中の楽曲部分と候補モーションの相性、ユーザがどのようなモーションに高評価を付けたかを同時に考慮

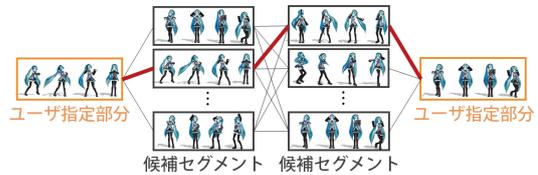
$$m_{n+1} = \underset{m \in \mathcal{L}^n S_n}{\operatorname{argmax}} \{ \rho \cdot \operatorname{rel}(q, m) + (1 - \rho) \cdot \Phi(m, S_n, L_n) \}$$

[Dou et al. WSDM2011]



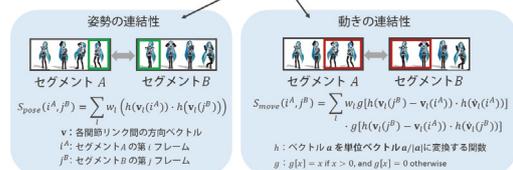
2. ユーザ指定振付間のダンス生成

ダイクストラ法を用いて一連の自然なダンスを生成



- ① **入力楽曲のテンポに対応**するため、1小節の長さを基準にダンスセグメントを正規化し、上図のようにグラフ構造を生成
- ② 各セグメント間で姿勢の連結性と動きの連結性を評価し、コスト関数の総和を最小化するセグメント列を**ダイクストラ法**で算出

$$C(A, B) = 1 / (S_{\text{pose}} + S_{\text{move}})$$



- ③ 選ばれたセグメント列の各セグメント間の姿勢を三次関数補間で滑らかにつなげて一連のダンスを生成

謝辞: 上記の図では、ピアプロ・キャラクター・ライセンスに基づいてクリプトン・フューチャー・メディア株式会社のキャラクター「初音ミク」を使用した。その3Dモデルには、Lat氏によって制作されたMMD用モデルを使用した。

楽曲の歌詞情報を考慮した 手話ダンス創作支援システム

朝比奈 わかな 岩本 尚也 Hubert P.H. Shum*1 森島 繁生
早稲田大学 *Northumbria University

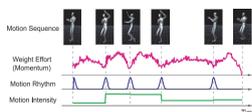
はじめに

3D 生成ツールの普及

近年、MikuMikuDance のような 3D 制作ツールの普及により 3D キャラクタを踊らせるようなダンス動画が多く作成されるようになった。しかし、一つの作品を生成するには多くの時間的労力や知識が必要である。

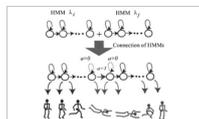
ダンスモーションの自動生成

このような背景から簡単にダンスモーションを自動生成する研究が行われるようになった。



[Shiratori et al. 2006]

楽曲の盛り上がりを考慮



[Takano et al. 2010]

モーションラベルを考慮

ダンスの楽曲の歌詞の意味やダンスに含まれる感情的要素を十分に表現しきれていない

手話ダンスの可能性

一方、手話ダンスは手話と HIPHOP などのダンスを組み合わせたダンスであり、楽曲の歌詞や雰囲気や適切に伝えることができる。



研究目的

音楽の歌詞情報を考慮した手話ダンス自動生成



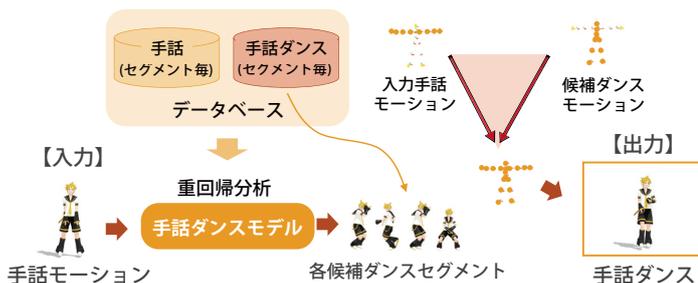
楽曲の歌詞情報 → 手話

楽曲の雰囲気 → ダンスモーション

簡単に意味のあるモーションを付与することができるようにすることが可能。

提案手法

研究の概要



謝辞: 上記の図では、ピアプロ・キャラクター・ライセンスに基づいてクリプトン・フューチャー・メディア株式会社のキャラクター「鏡音レン」を使用した。その 3D モデルには、もぐら氏および Lat 氏によって制作された MMD 用モデルを使用した。

1. 手話ダンスデータの取得

モーションキャプチャシステムを用いることで、手話ダンスモデル構築の際に必要な学習用モーションデータを取得した。

- 手話モーション・手話ダンスモーション(各 10 曲分)
- それら 20 曲分のモーションを 4 カウント毎に分割したセグメント(360 セグメント × 2)



2. 手話ダンスモデルの構築

重回帰分析を用いることで手話ダンスモデルを構築する。

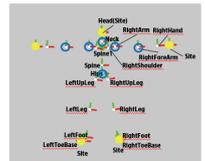
$$SD = a_0 + \sum_{i=1}^N a_i x_i$$

SD: 手話ダンスモーション特徴量
 x_i : 手話モーション特徴量
 a : 偏回帰係数

3. 使用する特徴量

我々は推定の際に用いるモーション特徴量として以下を用いた。

- 6 つの関節のクォータニオン(24 次元)
- 手・腕・ルートなど上半身に関する 6 つの関節



4. 候補手話ダンス決定

$$\gamma = w_1 \times \alpha + w_2 \times (1/\beta)$$

γ が最小のセグメントを最終候補とする

モーション特徴量距離 α : 推定された各セグメントと学習データの特徴量距離
 姿勢類似度 $1/\beta$: 前セグメントとの接続部分の姿勢類似度

※入力楽曲との BPM の差が閾値以上の候補セグメントは除外 ※ $w_1, w_2 = 1$

結果と考察



生成結果から、出力された手話ダンスは入力の手話の動きを強調するような動きが多く見られた(図2)。また、モーションキャプチャで取得した元の手話ダンスの動きと同じような動きが推定・生成された(図3)。

まとめと今後の課題

我々は楽曲の歌詞を考慮した手話ダンス自動生成手法を提案した。これにより、楽曲に合った意味を持つダンスモーションを生成できるようになるため、見ごたえのあるダンスモーションを作成することができる。

今後の課題・・・

- 音響特徴量を使用したモーション推定
- 手話データ入手することで入力は歌詞と音響特徴量のみ
- 主観評価実験による生成結果の評価

物理ベースのキャラクタ 2 次動作生成 と音楽シンクロキャラクタ生成

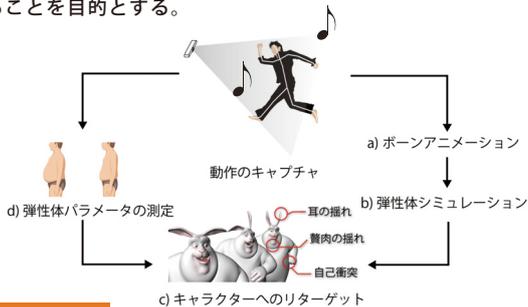
岩本 尚也 Hubert P.H. Shum*¹ Longzy Yang*¹ 森島 繁生
早稲田大学 *¹Northumbria University

アニメーション生成結果

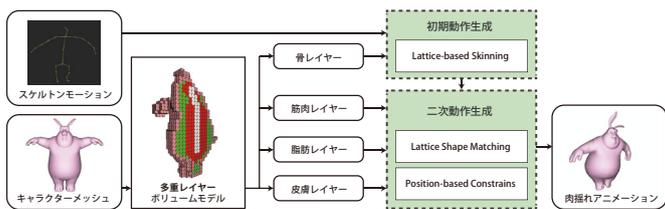


はじめに

CG キャラクタを用いたダンス動画コンテンツは、楽曲に含まれる音楽的、感情的要素を身体動作と表情によって表現できることから、あらゆる世代、地域、ジャンルを超えるエンターテインメントの原動力となりつつある。我々は従来のダンス動画コンテンツでは表現困難であった複雑な物理表現（二次動作）を可能にするシステムを実現することで、より躍動感のある音楽にシンクロしたダンスを表現することを目的とする。



ワークフロー



身体構造の考慮

骨

ユーザが与えた骨からの距離の値以内にある頂点を骨レイヤーとして登録。

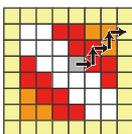
皮

各頂点の隣接位置に頂点が存在しない場合、その頂点を皮膚レイヤーとして登録。

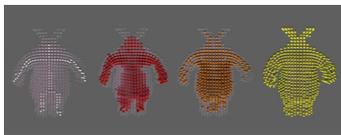
筋

各ボーンから皮膚への距離を計測し、最短ルート上の点に骨からの距離を登録。各ルートの最大距離を1とし、ユーザの指定した [0,1] の値で筋肉、脂肪を登録。

脂



例 骨から皮膚への距離を計算する。この場合、最大距離は5となる。ユーザが [0,1] の値を 0.5 とした時、マンハッタン距離が 2 (2.5) のレイヤーが筋肉、残りが脂肪レイヤーとなる。

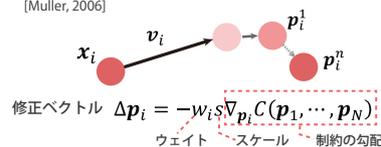


形状を保持する制約

Position based Dynamics における位置修正の制約

先に求めた座標に対し、制約を満たす位置へと修正し、現在位置と修正位置から次のフレームの速度を求める。

[Muller, 2006]



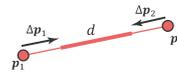
- $p_i = x_i + \Delta t v_i$
- n 回 (任意数) ループ
 - 制約を満たす Δp_i を算出
 - $p_i = p_i + \Delta p_i$
- $v_i = (p_i - x_i) / \Delta t$
- $x_i = p_i$

距離制約 [Muller, 2006]

隣接ボクセルの中心座標の距離を保つ制約を仮定。

$$\text{距離制約 } C_{dist}(p_1, p_2) = |p_1 - p_2| - d$$

自然長

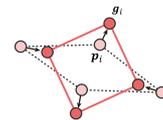


弾性体制約 [Rivers, 2007]

Lattice Shape Matching を用いた弾性体制約を仮定。

$$\text{弾性体制約 } \Delta p_i = \alpha (g_i(t) - p_i(t))$$

硬さ 目標位置 現在位置

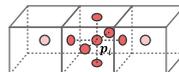


体積保存制約

現在の体積と初期の体積の差がゼロになる制約を仮定。

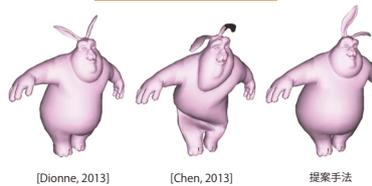
$$\text{体積保存制約 } C_{vol}(p_1, \dots, p_N) = vol(p_i) - V_i$$

頂点 p_i の現在の体積 初期体積 [Takamatsu, 2011]

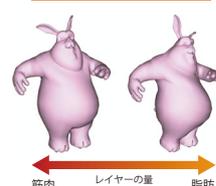


結果

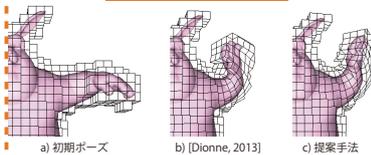
① 既存手法との比較



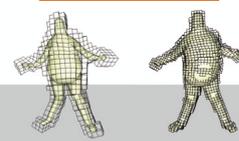
② 身体構造の違い



③ スキニングの比較



④ 解像度の違い

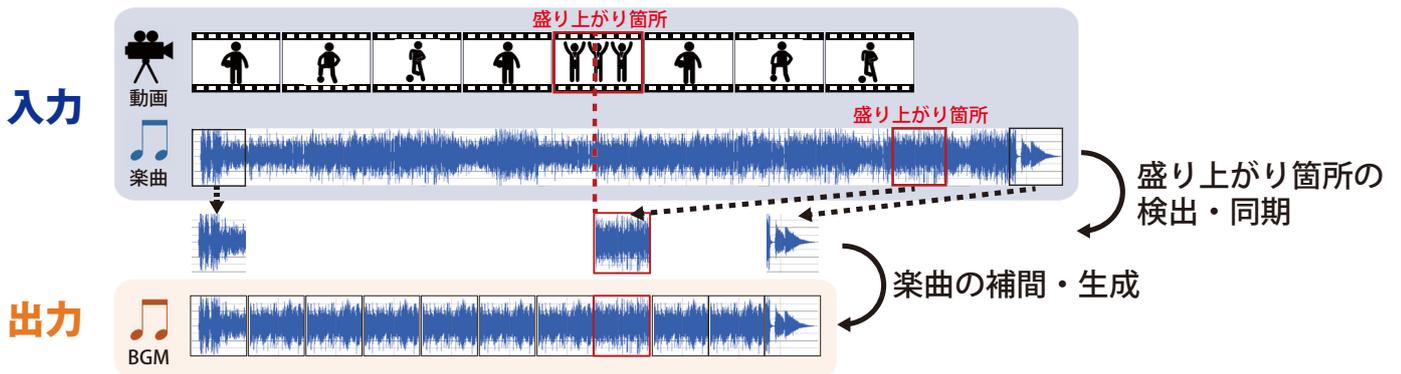


音楽と映像のクライマックスシンクロによる音楽動画コンテンツ創作支援

佐藤 晴紀 平井 辰典*1 中野 倫靖*2 後藤 真孝*2 森島 繁生
 早稲田大学 *1 駒澤大学 *2 産業技術総合研究所

研究目的

クライマックス(盛り上がり箇所)が同期した BGM の自動生成



1. はじめに

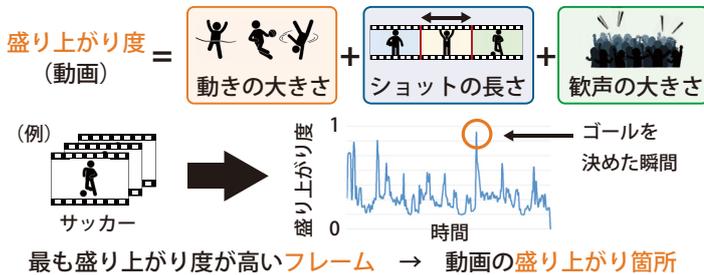
BGM の編集作業には**労力がかかる**



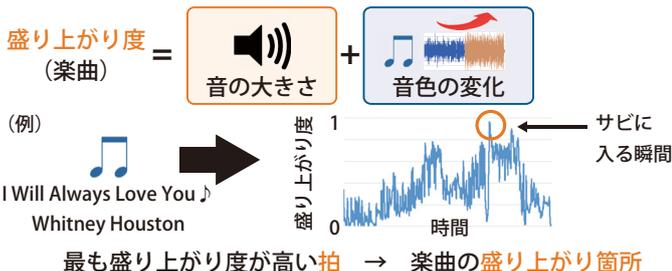
簡単に魅力的な BGM を編集できる技術が必要

2. 提案手法

盛り上がり箇所の検出: 動画 [Hanjalic et al., 2005]

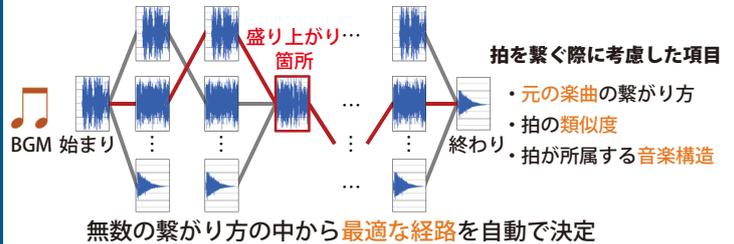


盛り上がり箇所の検出: 楽曲



楽曲の補間・生成 [Wenner et al., 2013]

動的計画法を用いた拍単位での切り貼りにより BGM を生成

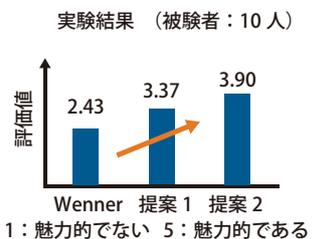


3. 評価実験

実験内容

- 質問: BGM が動画を魅力的にしているか
- 比較手法: 盛り上がりと長さの同期(提案 1)、盛り上がりの同期(提案 2) 長さの同期(Wenner et al., 2013)
- 使用動画: スポーツ、発表会見、映画

動画	付加手法	楽曲
スポーツ	Wenner	I Will Always Love You ♪
	提案 1	Whitney Houston
発表会見	Wenner	UNICORN ♪
	提案 2	Hiroyuki Sawano
映画	Wenner	Totetsunakimiti ♪
	提案 1	Kosuke Yamashita



4. まとめ

- 動画や楽曲の盛り上がり箇所の検出
- 盛り上がり箇所と長さが一致した BGM の自動生成

今後の課題

- 盛り上がり箇所検出の精度の向上
- 歌声がある楽曲への対応

音楽理解技術に基づくマイクロタスク 回答可能な音楽ゲーム自動生成

中村 聡史 三輪 聡哉 十文字 優斗 *1
 明治大学 *1 東京農工大学

本研究の目的

マイクロタスクを楽しくこなせるように！

マイクロタスク(10秒以内で回答可能なタスク)による社会貢献

- ・道に進行方向で回答可能なタスクを提示したり、ダウンロード時にちょっとしたタスクを提示して貢献してもらう
- ・ひとの特性を利用しつつ、より楽しくマイクロタスクをこなせるような場を実現することが重要

何が問題か？

- ・マイクロタスクを埋め込むことが可能な場に現時点で制約がある
- ・マイクロタスク用のコンテンツを用意すると、コンテンツによって飽きが生じてしまう
- ・コンテンツを用意するのも面倒

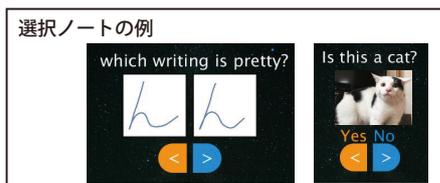
アプローチ

- ・音楽の理解技術に基づき、音楽ゲームを自動生成するとともに、音楽ゲームの中にマイクロタスクを埋め込む
- ・音楽ゲームのボタン操作を、マイクロタスクのタスク回答に利用する！



音楽ゲームにタスクを埋め込む

- ・音楽にあわせてノート(押すボタン・押すタイミング)を設定
- ・ノートの一部をどちらのボタンでもマイクロタスクを選択する
選択ノートに置き換え
- ・ユーザはノートに合わせてボタンを押すだけ
- ・選択ノートの部分(マイクロタスク提示部)では、どちらのボタンを押してもよいが、ユーザは正解と考えるほうを押すのではないか？

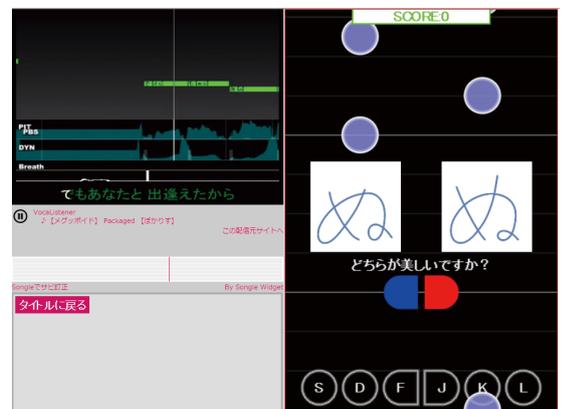


音楽理解からのゲーム自動生成

音楽ごとに音楽ゲームを毎回作るのは面倒！

- ・音楽・動画共有サイトの膨大な音楽に着目
- ・音楽の理解技術に基づき、音楽ゲームを自動生成！

1. 時間 $t=0$ とする
2. t 秒目に音がある場合は 3 へ、ない場合は 6 へ
3. t 秒目がサビなら 4 へ、サビでないなら 5 へ
4. t 秒と $t+\Delta t$ 秒の間にメロディの音に変化している場合はノートを挿入
そうでない場合は 6 へ
5. t 秒がビート開始位置の場合はノートを挿入、そうでない場合は 6 へ
6. t に Δt を加え、2 に戻る
7. 盛り上がりに応じてノートを倍の量に
8. ランダムに選択ノートを適用(周辺を削除)



http://nkmr.io/music_game/

ひとの視覚的特性を利用した エフェクトの付与による音楽動画鑑賞支援

松井 啓司 田村 柁優紀 松田 滉平 福地 翼 山浦 祐明 大島 遼 中村 聡史
明治大学

本研究の目的

音楽動画の鑑賞体験をより豊かなものに！

音楽動画から受ける「印象」はさまざまだが、こうした鑑賞体験をより豊かなものにする

- ・普通の音楽動画をより迫力のあるものにした
- ・ちょっと怖い音楽動画に覚える恐怖感を、より強いものにした

何が問題か？

- ・ディスプレイやスピーカとしてよいものを使うのは多くの制約がある
- ・コンテンツを編集するのは手間がある
- ・オリジナルのコンテンツをダウンロードして編集するのはクリエイターに好まれない場合がある

アプローチ

- ・ディスプレイやスピーカ、コンテンツはそのままエフェクトをコンテンツ・周辺に提示することで解決
- ・視線の特性を効果的に利用(周辺視、中心視)

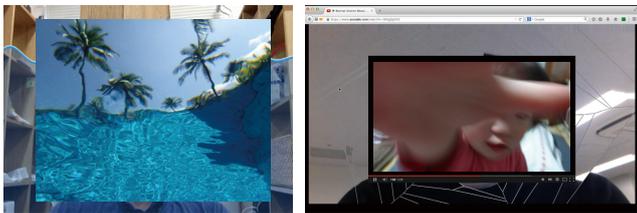
周辺視に対する錯視図形提示

ひとがぼんやりみている周辺視に着目

- ・周辺視部分の視力は弱い
- ・周辺視部分では形などを識別することは困難
- ・周辺視部分から何らかの影響を受ける

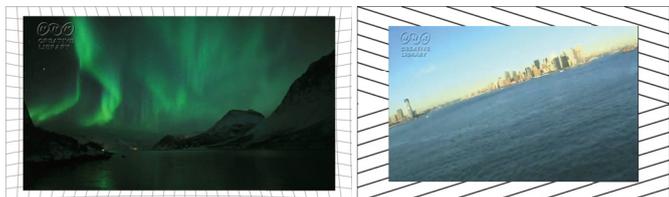
周辺視に対して意味のあるエフェクトを提示

- ・海に潜ったかのようなエフェクトの提示
- ・カメラが割れるようなエフェクトの提示
- ・両側から壁が迫りくるようなエフェクトの提示



周辺視に対して錯視図形を提示

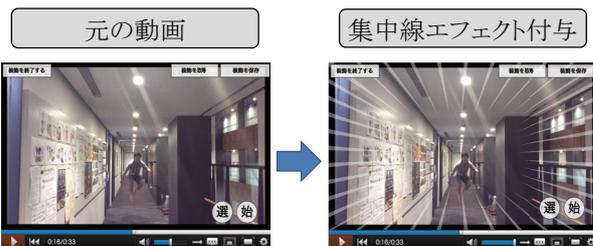
- ・奥行きに関する錯視を音楽に合わせて提示
- ・大きさに関する錯視を音楽に合わせて提示
- ・移動感に関する錯視を音楽に合わせて提示



中心視に対するエフェクト重畳

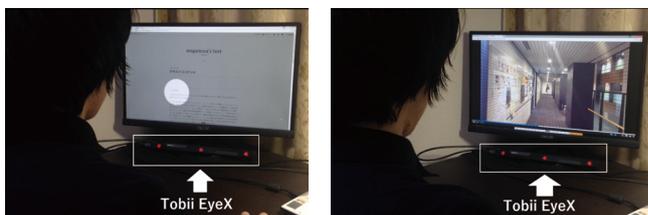
コンテンツ視聴中のユーザの視線を誘導するようなエフェクトを装飾可能とする(Decoby)

- ・音楽動画に集中線やセピア調、砂嵐といったエフェクトを付与可能に
- ・音楽動画の上に透明なレイヤを生成し、その上でエフェクトを提示



コンテンツ視聴中のユーザの視線に追従して
コンテンツ上にエフェクトを重畳

- ・どの部分を見ているか(視線の位置)を視線検出装置にて取得
- ・視線の位置に応じてエフェクトを動的に動かすことで、新たな鑑賞体験を創出



音楽動画に対する 印象評価データセットの構築

土屋 駿貴 大野 直紀 阿部 和樹 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

本研究の目的

印象に基づく音楽動画検索の実現

音楽動画から受ける「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、**爽やかな感じの音楽動画を視聴したい!**
- ・この音楽動画と、**似た印象の音楽動画を視聴したい!**



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**
 - 既存のデータセットは音楽動画全体を扱うものであり問題が多い
 - 音楽と映像それぞれの印象とその組み合わせはどういう印象?



新たなるデータセットの構築

メディア独立での印象評価データセットの必要性

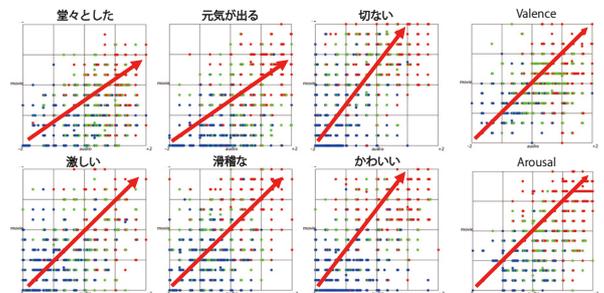
- ・VOCALOID で視聴数の多い 500 個の音楽動画を選定
- ・評価対象はサビ部分に限定
- ・音楽のみ、映像のみ、音楽と映像の組み合わせについてそれぞれ評価
- ・8 印象軸の印象評価 (C1: 堂々とした、C2: 元気が出る、C3: 切ない、C4: 激しい、C5: 滑らかな、C6: かわいい、V: Valence、A: Arousal)
- ・3 人以上が評価

下記 URL にてデータセットを公開中!

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

音楽と映像の印象の組み合わせ

- ・オリジナルのものについては音楽と映像は相関
- ・印象軸によって引っ張られる傾向が異なる



音楽と映像をランダムに組み合わせただけのものは音楽印象に引っ張られる傾向あり

コメントによる印象推定

音楽動画へのコメントからの印象推定 (SVM)

- ・コメントからの映像の印象推定精度は高い
- ・コメントからの音楽の印象推定精度は低い
- ・コメントの形容詞が有効に働く

	C1	C2	C3	C4	C5	C6	V	A	平均				
音楽動画	動形 形	全 名形	名形 動形	全 形副	0.781	0.869	0.713	0.780	0.750	0.856	0.783	0.844	0.797
音楽のみ	名形 全	名形 名形	全2 全2	名形 名形	0.754	0.671	0.612	0.750	0.725	0.787	0.740	0.806	0.730
映像のみ	動形 名形	全 形	名形 形	形副 形副	0.921	0.792	0.752	0.759	0.657	0.829	0.840	0.884	0.804
平均	0.819	0.777	0.692	0.763	0.711	0.824	0.788	0.845	0.777				

主観特徴+物理特徴は?

主観的な印象: コメントからの印象推定

物理的な印象: 音楽特徴、映像特徴による印象推定

コメントベクトル+音楽特徴ベクトル

- ・お互いに補完しあう関係になり、概ね精度は向上

	C1	C2	C3	C4	C5	C6	V	A	平均
主観特徴	0.736	0.769	0.717	0.844	0.649	0.756	0.842	0.784	0.762
音響特徴	0.806	0.644	0.652	0.863	0.711	0.800	0.902	0.707	0.761
主観+音響	0.817	0.722	0.810	0.824	0.667	0.780	0.917	0.744	0.785

+映像特徴については今後追加予定

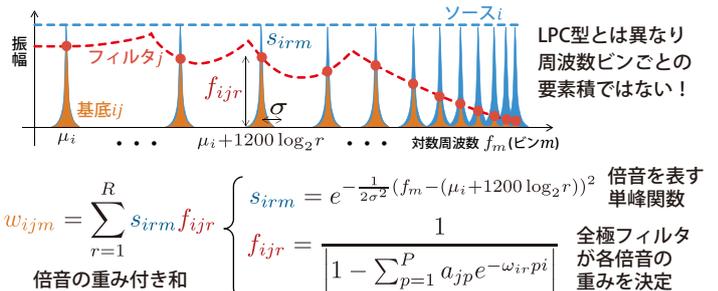
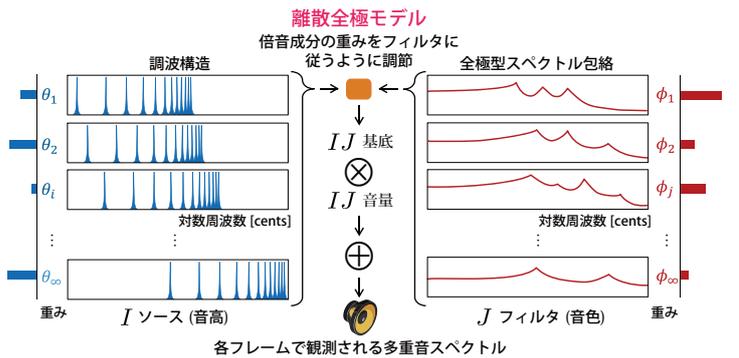
確率的生成モデルに基づく 音楽音響信号解析

吉井和佳 中村栄太 糸山克寿 後藤真孝*1

京都大学 *1 産業技術総合研究所

無限重畳離散全極モデル：振幅スペクトログラムの音高・音色への分解

無限複合自己回帰モデル [Yoshii 2012] と離散全極モデルとの統合による対数周波数領域でのソース・フィルタ分解



観測スペクトログラムに対する確率モデルの尤度

$$X_{mn} \sim \text{Poisson} \left(\sum_{i=1}^{I \rightarrow \infty} \sum_{j=1}^{J \rightarrow \infty} \theta_i \phi_j w_{ijm} h_{ijn} \right)$$

音量に対するガンマ事前分布 (時間的連続性も考慮可能)

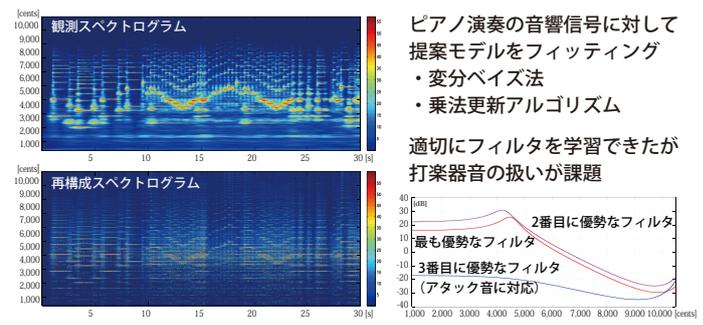
$$h_{ijn} \sim \text{Gamma}(a_H, b_H)$$

スパースな解に誘導する効果

ソース・フィルタの重みに対するガンマ過程 (無限次元の非負ベクトル上の事前分布)

$\theta_i \sim \text{GaP}(\alpha_\theta, 1)$ ほとんどすべての要素が限りなくゼロに近く、

$\phi_j \sim \text{GaP}(\alpha_\phi, 1)$ 高々有限個の要素のみ値をもつように誘導できる



無限半正定値テンソル分解：複素スペクトログラムの基底への分解

PSDTF：半正定値行列 (観測行列) を少数の半正定値行列 (基底行列) の線形結合で近似

確率モデルの尤度関数を複素ガウス分布から裾の広い複素t分布に変更することで初期値依存性が低く高精度な分解が可能

基底スペクトル w_1, w_2, w_3

行列データ：パワースペクトログラム (複素スペクトログラムの絶対値の二乗)

$$X \approx WH$$

時間方向の音量 h_1, h_2, h_3

基底共分散行列 (複素エルミート半正定値行列) W_1, W_2, W_3

周波数領域での分解

ガンマ過程に基づく基底数の無限化も可能

観測行列 (非負値ベクトル群) $X = [x_1, \dots, x_N] \in \mathbb{R}^{M \times N}$

基底行列 (非負値ベクトル群) $W = [w_1, \dots, w_K] \in \mathbb{R}^{M \times K}$

音量行列 (非負値ベクトル群) $H = [h_1, \dots, h_K]^T \in \mathbb{R}^{K \times N}$

ベクトル単位の分解 $x_n \approx \sum_{k=1}^K h_{kn} w_k \stackrel{\text{def}}{=} y_n$

観測テンソル (半正定値行列群) $\mathcal{X} = [X_1, \dots, X_N] \in \mathbb{C}^{M \times M \times N}$

基底テンソル (半正定値行列群) $\mathcal{W} = [W_1, \dots, W_K] \in \mathbb{C}^{M \times M \times K}$

音量行列 (非負値ベクトル群) $H = [h_1, \dots, h_K]^T \in \mathbb{R}^{K \times N}$

行列単位の分解 $X_n \approx \sum_{k=1}^K h_{kn} W_k \stackrel{\text{def}}{=} Y_n$

複素スペクトル s_n

$$X_n = s_n s_n^H$$

対角成分はパワースペクトル x_n

複素t分布の自由度が無数のとき複素ガウス分布に一致

→ 従来のPSDTFに一致

多変量複素t分布に基づく対数尤度 $p(s_n | Y_n) = \mathcal{T}_\nu(s_n | Y_n)$

自由度パラメータ

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

$X_{10}, X_{130}, X_{250}, X_{370}, X_{490}, X_{610}, X_{730}$

時間方向の音量 h_1, h_2, h_3

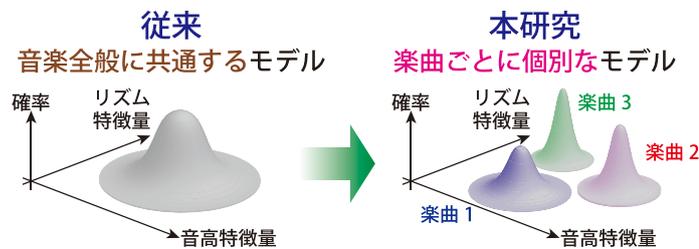
楽譜データに対する 音楽言語モデルの教師なし学習

中村 栄太 糸山 克寿 吉井 和佳
京都大学

音楽言語モデルの教師なし学習

各楽曲に特有な文法を記述するモデルの構築

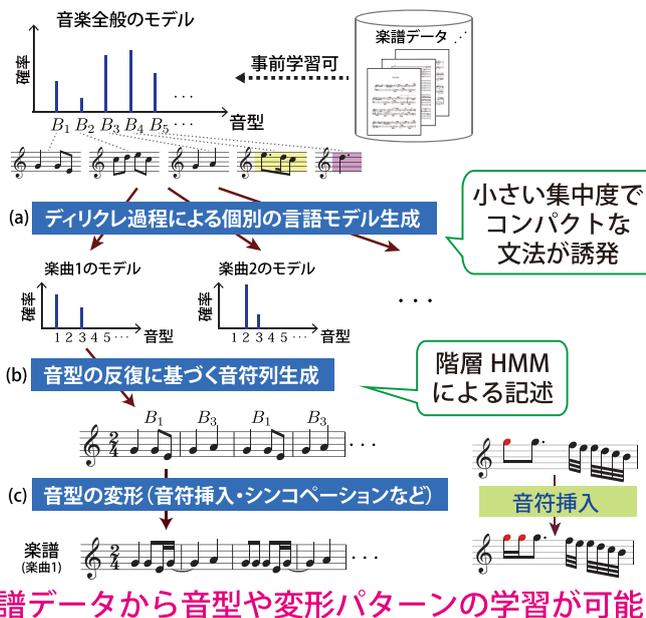
- ジャンル・スタイル・作曲者により異なる多様な音楽文法を精密に表現
- 多くの楽曲では、反復構造により現れる音符のパターンは少数 → コンパクトな音楽文法の学習
- 音楽の背後にある構造・規則をデータから学習



音型の反復と変形の階層ベイズモデル

3ステップからなる楽曲生成過程のモデル

- 楽曲で用いる (少数の) 音型の選択
- 音型の貼り合わせ (反復) による音符列の生成
- 音型の変形による最終的な音符列の生成



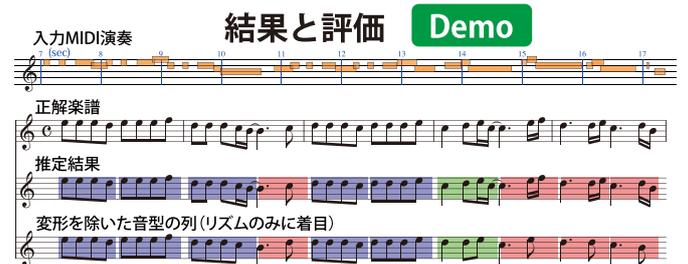
リズム採譜への応用

目標: MIDI 演奏から楽譜 (リズム) を推定

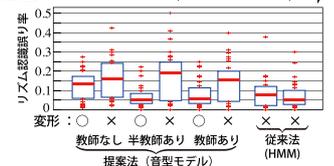
- テンポ変動や動作ノイズにより不正確なリズムから正しいリズムを復元
 - 音型モデルを用いて背後にある文法を同時推論
- 動的システムによる演奏のリズム変動モデル

基本式: 音長 $d_n = (\text{テンポ } v_n) \times (\text{音価 } r_n)$
にテンポ変動と動作ノイズを確率的に導入

$$v_n = v_{n-1} + \epsilon_v, \quad d_n = v_n r_n + \epsilon_t$$

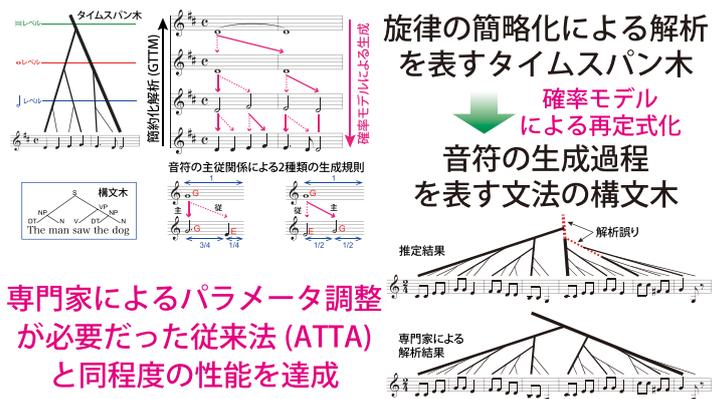


従来法 (HMM) より
認識精度も向上
(92.1% → 93.4%)



音楽文法構文木の学習

音楽理論 GTTM で現れる音符列の構文木を
統計モデル (PCFG) を用いて学習



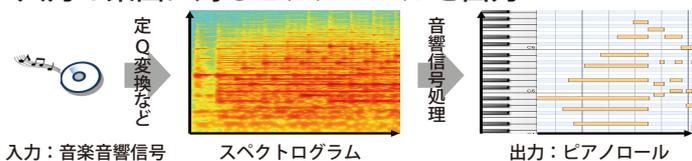
階層ベイズ音響・言語統合モデルに基づく多重基本周波数推定

尾島 優太 中村 栄太 糸山 克寿 吉井 和佳

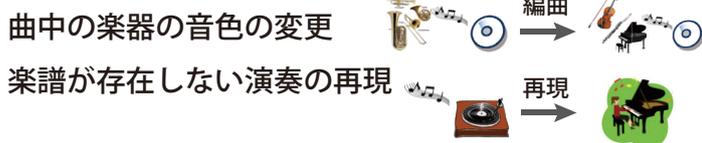
京都大学

多重基本周波数推定とは

音楽音響信号から各音の音の高さを推定
 入力楽曲に対しピアノロールを出力

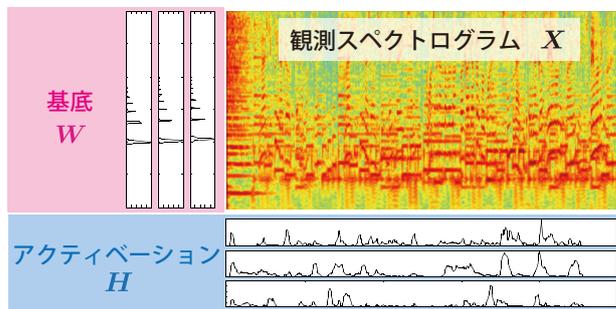


こんなことができるように



非負値行列因子分解

入力音響信号から得られたスペクトログラムを
基底 と **アクティベーション** の積として近似
 各音のスペクトル 各音の時間的音量変化



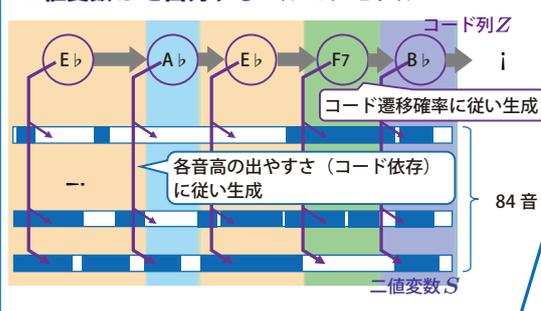
音響モデルと言語モデル

キーアイデア ピアノロールの背後にはコード進行が存在

コード列, コードの構造を同時推定することで音高推定精度の向上を図る

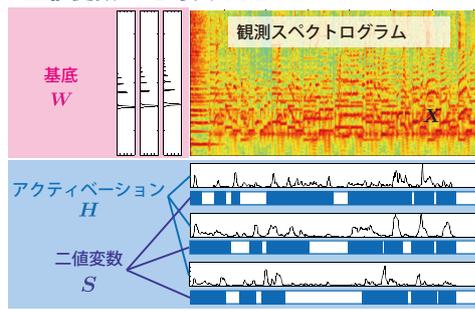
言語モデル (隠れマルコフモデル)

コード列 Z を隠れ変数に持ち
 二値変数 S を出力するマルコフモデル



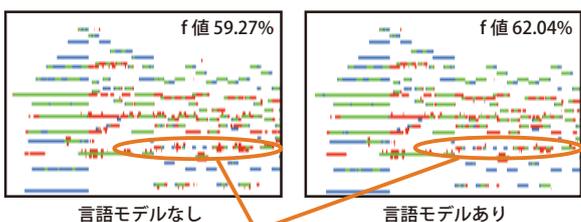
音響モデル (Beta-process NMF)

ピアノロールを表現する
 二値変数 S を導入したNMF



推定結果

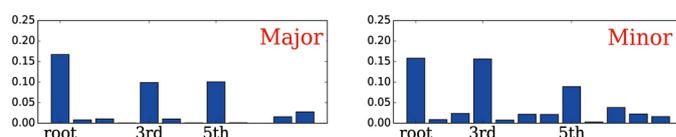
言語モデルの統合により **不必要な音が抑制**



中低音域での挿入誤りが減少

緑：正解
 赤：挿入誤り
 青：脱落誤り

音響信号のみに基づき、出力確率として
Major / Minor に対応したコード構造を獲得



→ 音楽文法としてのコードの構造を
 音響信号から直接学習可能

準ビート同期 HMM に基づく 歌声 F0 軌跡の自動採譜

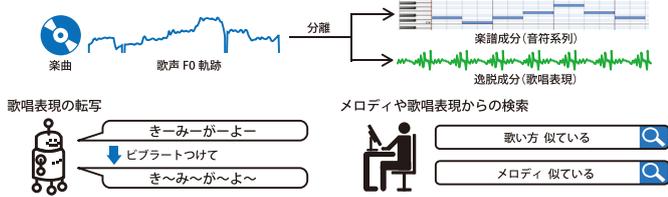
錦見亮 中村 栄太 糸山 克寿 吉井 和佳

京都大学

研究背景

楽曲の歌声の解析

歌声を楽譜成分(音符系列)と逸脱成分(歌唱表現)とに分離
歌唱表現の転写や楽曲の検索に応用可能



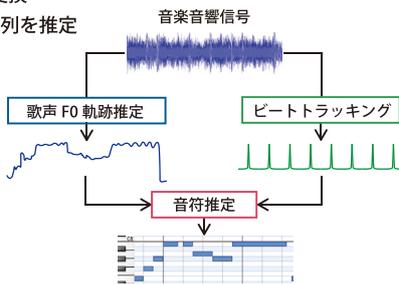
音符推定問題

音楽音響信号から楽譜(音符の系列)を推定する問題

連続的なから離散的な信号への変換
歌声 F0 軌跡からメロディ 音符系列を推定

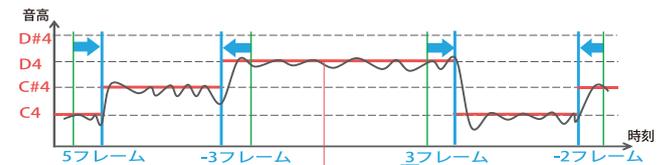
問題設定

入力
音楽音高信号から推定された
10ms 単位の歌声 F0 軌跡
16 分音符単位のビート時刻
出力
16 分音符単位での音符の音高



歌声の生成過程の定式化

音高と時間変動 | HMM の潜在変数で表現

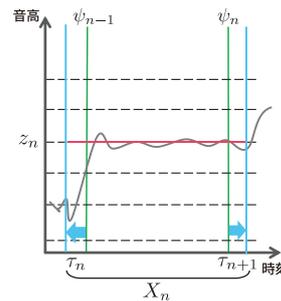


$$p(z_n | z_{n-1}, A) = \text{Categorical}(z_n | \mathbf{a}_{z_{n-1}})$$

$$p(\tau_n | \rho) = \text{Categorical}(\tau_n | \rho)$$

周波数変動 | HMM の出力確率で表現

尺度パラメータを $\Delta F0$ に比例させる



$$p(X_n | z_n, \tau_{n-1}, \tau_n) = \frac{1}{(\psi_n + \tau_n) - (\psi_{n-1} + \tau_{n-1})}$$

$$\left\{ \prod_{x \in X_n} p(x | z_n) \right\}$$

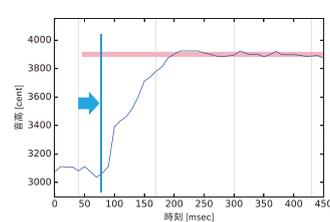
$$p(x_t | z_n = k) = \text{Cauchy}(x_t | \mu_k, \lambda_t)$$

$\lambda_t = c|x_t - x_{t-1}| + d$ (尺度パラメータ $\Delta F0$ に比例)

歌声 F0 軌跡の変動と生成過程

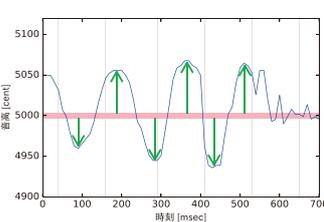
時間変動

ビート時刻とオンセット時刻のずれ



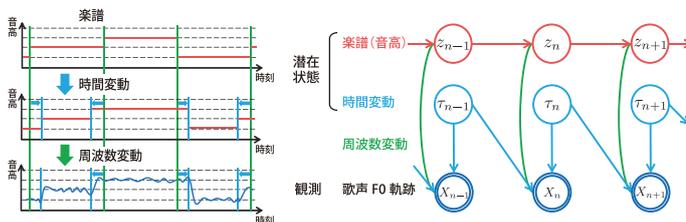
周波数変動

楽譜の音高と歌声の音高のずれ



歌声 F0 軌跡の生成過程と HMM によるモデル化

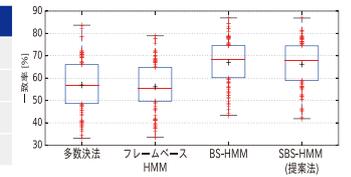
楽譜に時間・周波数変動が付与されて歌声 F0 軌跡が生成される過程
楽譜と時間変動は潜在状態, 観測歌声 F0 軌跡は出力確率で表現



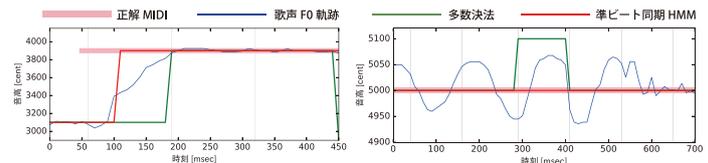
実験結果

音高推定精度

モデル	一致率(±標準誤差)
多数決法	56.9 (±1.14)
フレームベース	56.1 (±1.09)
ビート同期	67.0 (±1.00)
準ビート同期	66.3 (±0.99)

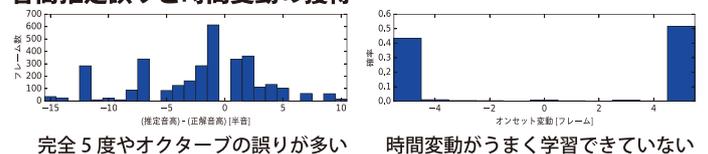


時間・周波数変動に対する結果



変化する大きなビブラートに頑健

音高推定誤りと時間変動の獲得





OngaCREST プロジェクト研究参加者一覧

後藤グループ

後藤 真孝	(産業技術総合研究所)	2011年10月～2017年3月
濱崎 雅弘	(産業技術総合研究所)	2011年10月～2017年3月
中野 倫晴	(産業技術総合研究所)	2011年10月～2017年3月
深山 覚	(産業技術総合研究所)	2013年4月～2017年3月
加藤 淳	(産業技術総合研究所)	2014年4月～2017年3月
佃 洸撰	(産業技術総合研究所)	2015年4月～2017年3月
尾形 正泰	(産業技術総合研究所)	2016年4月～2017年3月
Jordan Smith	(産業技術総合研究所)	2014年11月～2017年3月
川崎 裕太	(産業技術総合研究所)	2011年12月～2017年3月
石田 啓介	(産業技術総合研究所)	2012年4月～2017年3月
井上 隆広	(産業技術総合研究所)	2012年11月～2017年3月

田中 一大	(産業技術総合研究所)	2012年4月～2017年3月
戸田 智基	(奈良先端科学技術大学院大学)	2012年4月～2017年3月
吉井 和佳	(産業技術総合研究所)	2011年10月～2014年3月
持橋 大地	(統計数理研究所)	2012年4月～2014年3月
藤原 弘将	(産業技術総合研究所)	2011年10月～2012年8月
栗原 一貴	(津田塾大学)	2011年10月～2015年3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
Matt McVicar	(産業技術総合研究所)	2013年9月～2014年10月
Graham Percival	(産業技術総合研究所)	2014年6月～2015年5月

森島グループ

森島 繁生	(早稲田大学)	2011年10月～2017年3月
岩本 尚也	(早稲田大学)	2014年2月～2017年3月
福里 司	(早稲田大学)	2014年4月～2017年3月
加藤 卓哉	(早稲田大学)	2015年6月～2017年3月
赤堀 渉	(早稲田大学)	2014年7月～2017年3月
朝比奈 わかな	(早稲田大学)	2014年7月～2017年3月
佐藤 晴紀	(早稲田大学)	2014年7月～2017年3月
成田 史弥	(早稲田大学)	2015年10月～2017年3月
中塚 貴之	(早稲田大学)	2015年6月～2017年3月
福原 吉博	(早稲田大学)	2015年10月～2017年3月
柿塚 亮	(早稲田大学)	2015年6月～2017年3月
古川 翔一	(早稲田大学)	2016年4月～2017年3月
野沢 繪佐	(早稲田大学)	2016年4月～2017年3月
Savkin Pavel	(早稲田大学)	2016年5月～2017年3月

山口 周悟	(早稲田大学)	2016年5月～2017年3月
持田 恵佑	(早稲田大学)	2016年5月～2017年3月
小澤 禎裕	(早稲田大学)	2016年5月～2017年3月
松田 聡子	(早稲田大学)	2013年10月～2017年3月
前島 謙宣	(早稲田大学)	2011年10月～2014年5月
大矢 隼士	(早稲田大学)	2012年4月～2015年3月
濱崎 皓介	(早稲田大学)	2012年4月～2014年3月
佐々木 将人	(早稲田大学)	2012年4月～2015年3月
増田 太郎	(早稲田大学)	2012年4月～2015年3月
岡田 成美	(早稲田大学)	2014年4月～2015年3月
平井 辰典	(早稲田大学)	2011年10月～2016年3月
張 卓鵬	(早稲田大学)	2014年7月～2016年3月
古澤 知英	(早稲田大学)	2014年7月～2016年3月

中村グループ

中村 聡史	(明治大学)	2011年10月～2017年3月
宮下 芳明	(明治大学)	2012年10月～2017年3月
渡邊 恵太	(明治大学)	2013年4月～2017年3月
大島 遼	(明治大学)	2013年11月～2017年3月
加藤 邦拓	(明治大学)	2015年11月～2017年3月
高橋 治輝	(明治大学)	2015年11月～2017年3月
若林 裕太	(明治大学)	2015年11月～2017年3月
秋山 耀	(明治大学)	2016年4月～2017年3月
佐藤 剣太	(明治大学)	2015年6月～2017年3月
新納 真次郎	(明治大学)	2015年6月～2017年3月
牧 良樹	(明治大学)	2015年6月～2017年3月
松田 滉平	(明治大学)	2015年6月～2017年3月
松井 啓司	(明治大学)	2015年6月～2017年3月
田村 柁優紀	(明治大学)	2015年6月～2017年3月
大野 直紀	(明治大学)	2015年6月～2017年3月
土屋 駿貴	(明治大学)	2015年8月～2017年3月
前島 紘希	(明治大学)	2015年8月～2017年3月

田島 一樹	(明治大学)	2016年4月～2017年3月
白鳥 裕士	(明治大学)	2016年4月～2017年3月
神山 拓史	(明治大学)	2016年4月～2017年3月
薄羽 大樹	(明治大学)	2016年4月～2017年3月
土井 麻由佳	(明治大学)	2016年4月～2017年3月
鳥山 らいか	(明治大学)	2016年4月～2017年3月
宮代 理弘	(明治大学)	2016年4月～2017年3月
斉藤 絢基	(明治大学)	2016年6月～2017年3月
樋川 一幸	(明治大学)	2016年6月～2017年3月
久保田 夏美	(明治大学)	2016年6月～2017年3月
山浦 祐明	(明治大学)	2016年6月～2017年3月
阿部 和樹	(明治大学)	2016年6月～2017年3月
福地 翼	(明治大学)	2016年6月～2017年3月
山中 祥太	(明治大学)	2016年4月～2017年3月
山本 岳洋	(京都大学)	2011年10月～2017年3月
佃 洸撰	(京都大学)	2011年10月～2015年3月
湯村 翼	(明治大学)	2013年12月～2015年3月

吉井グループ

吉井 和佳	(京都大学)	2014年4月～2017年3月
糸山 克寿	(京都大学)	2014年4月～2017年3月
中村 栄太	(京都大学)	2015年8月～2017年3月
大喜多 美里	(京都大学)	2014年8月～2017年3月
福田 翼	(京都大学)	2015年4月～2017年3月
尾島 優太	(京都大学)	2015年8月～2017年3月
錦見 亮	(京都大学)	2015年8月～2017年3月

津島 啓晃	(京都大学)	2016年4月～2017年3月
和田 雄介	(京都大学)	2016年4月～2017年3月
持橋 大地	(統計数理研究所)	2014年4月～2017年3月
中村 友彦	(東京大学)	2014年4月～2016年3月
池宮 由菜	(京都大学)	2014年4月～2015年3月
丸尾 智志	(京都大学)	2014年8月～2016年3月
土橋 彩香	(京都大学)	2015年4月～2016年3月

OngaCRESTシンポジウム2016 (<http://ongacrest.jp/symposium2016>)
「音楽情報処理研究が切り拓いた世界」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 2016年8月27日

代表者 後藤 真孝

編集者 笠井 志麻

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

国立研究開発法人 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
