

OngaACCEL

シンポジウム2020

音楽情報処理研究が切り拓く
未来の音楽体験

予稿集

開催概要

開催名	OngaACCELシンポジウム2020 「音楽情報処理研究が切り拓く未来の音楽体験」
開催日	2020年9月12日（土） 13:30-16:30（13:15 配信開始予定）
開催形式	オンライン開催（YouTube Live によるライブ動画配信）
主催	科学技術振興機構 (JST) OngaACCEL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」 科学技術振興機構(JST) 戦略的創造研究推進事業 ACCEL 2016年度採択研究課題 研究代表者 後藤 真孝（産業技術総合研究所 首席研究員） プログラム マネージャー 伊藤 博之（科学技術振興機構 ACCELプログラムマネージャー） 研究分担者 吉井 和佳（京都大学 准教授） 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 教授） 研究分担者 佐々木 渉（クリプトン・フューチャー・メディア(株) チームマネージャー）

プログラム

13:15	動画配信開始
13:30-13:35	「プログラムマネージャーよりご挨拶」 伊藤 博之（科学技術振興機構 ACCELプログラムマネージャー）
13:35-13:45	「研究代表者によるプロジェクト紹介」 後藤 真孝（研究代表者 / 産業技術総合研究所 首席研究員）
13:45-14:25	「後藤グループに関する研究成果の紹介」 後藤 真孝（研究代表者 / 産業技術総合研究所 首席研究員）
14:25-14:50	「吉井グループに関する研究成果の紹介」 吉井 和佳（研究分担者 / 京都大学 准教授）
14:50-15:00	休憩
15:00-15:25	「森島グループに関する研究成果の紹介」 森島 繁生（研究分担者 / 早稲田大学 教授）
15:25-15:50	「中村グループに関する研究成果の紹介」 中村 聡史（研究分担者 / 明治大学 教授）
15:50-16:05	「佐々木グループに関する研究成果の紹介」 佐々木 渉（研究分担者 / クリプトン・フューチャー・メディア(株) チームマネージャー）
16:05-16:25	「プログラムマネージャーによる総括」 伊藤 博之（科学技術振興機構 ACCELプログラムマネージャー）
16:25-16:30	「研究代表者よりご挨拶」 後藤 真孝（研究代表者 / 産業技術総合研究所 首席研究員）



OngaACCEL シンポジウム2020

音楽情報処理研究が切り拓く
未来の音楽体験

予稿集

はじめに

OngaACCEL シンポジウム 2020「音楽情報処理研究が切り拓く未来の音楽体験」をご視聴いただき、どうもありがとうございます。我々の「OngaACCEL（音画 ACCEL/ おんがあくせる）プロジェクト」では、音楽体験の未来を切り拓く技術開発によって、音楽の楽しみ方がより能動的で豊かになるように、鑑賞・創作・協創の3つの側面から人々を支援できるサービスプラットフォームの構築を目指しています。2016年8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、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。

本シンポジウムはオンラインで開催され、3時間のライブ動画配信により、OngaACCEL プロジェクトの研究代表者（後藤 真孝）とプログラスマネージャー（伊藤 博之）、4名の研究分担者（吉井 和佳、森島 繁生、中村 聡史、佐々木 渉）が研究成果のデモンストレーションを交えながら講演します。また、講演を補足するために、本予稿集には主要な研究成果に関するポスター集も掲載しました。是非併せてお楽しみいただければ幸いです。

我々の研究活動は、2016年に研究プロジェクト「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」（OngaACCEL プロジェクト）が科学技術振興機構（JST）戦略的創造研究推進事業 ACCEL に採択され、多くの方々を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。しかも、このOngaACCEL プロジェクトは、その前に5年間、JST 戦略的創造研究推進事業 CREST に採択されて2011年から実施した研究プロジェクト「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」（OngaCREST プロジェクト）の研究成果が土台となっています。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く未来の音楽体験をより一層豊かにして、学術的・産業的・社会的・文化的に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



プログラスマネージャー
伊藤 博之

研究代表者 略歴

後藤 真孝（産業技術総合研究所 首席研究員）

1998年早稲田大学大学院 理工学研究科 博士後期課程修了。博士(工学)。現在、国立研究開発法人 産業技術総合研究所 人間情報インタラクション研究部門 首席研究員 兼 メディアコンテンツ生態系プロジェクトユニット代表。JST ACT-「情報と未来」研究総括、日本学術会議 連携会員等を兼任。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、FIT船井業績賞、科学技術分野の文部科学大臣表彰 若手科学者賞、星雲賞【ノンフィクション部門】等、51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。未来のメディアコンテンツ体験をより豊かにす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28年間推進。

プログラスマネージャー 略歴

伊藤 博之（科学技術振興機構 ACCELプログラスマネージャー）

北海道大学に勤務の後、1995年7月札幌市内にてクリプトン・フューチャー・メディア株式会社を設立。世界各国に100数社の提携先を持ち、3000万件以上のサウンドコンテンツは世界でも最大級。DTMソフトウェア、音楽配信アグリゲーター、3DCG技術など、音を発想源としたサービス構築・技術開発を日々進めている。「初音ミク」の開発会社としても知られている。北海学園大学経済卒。NoMaps実行委員長、日本ネットクリエイター協会理事、Independent Music Coalition Japan (IMCJ)理事、北海道情報大学客員教授も兼任。2016年より科学技術振興機構ACCELプログラスマネージャーとしてOngaACCELプロジェクトに参加。主な受賞歴として、藍綬褒章受章(2013年秋の受勲)、FIT船井業績賞(2016年)。

目次

はじめに	2
目次	3
OngaACCEL プロジェクトの狙いと研究成果の紹介	4
Songle: 楽曲の中身を自動解析する能動的音楽鑑賞サービス	10
Songle Sync: 大規模音楽連動制御プラットフォームとその活用事例	11
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	12
TextAlive: 音楽に同期したリリックビデオ制作支援サービスとその活用事例	13
Lyric Jumper: 歌詞トピック解析技術に基づく歌詞探索サービス	14
音楽音響信号中の様々な要素を推定する音楽解析技術	15
音楽推薦・検索に基づく音楽鑑賞支援	16
身体モーションの情報処理と制作支援技術	17
CRNN-HSMM ハイブリッドモデルに基づく歌声採譜	18
変分オートエンコーダーに基づく半教師あり深層コード推定	19
事前学習済み言語モデルを用いた正則化に基づくドラム採譜	20
LSTM-HSMM ハイブリッドモデルに基づく音楽構造解析	21
深層多重音検出に基づくピアノ採譜	22
フォトリアルなキャラクターのダンスおよび歌唱動作生成システム	23
ダンス動画創作支援: 1枚画像からの着衣全身3次元モデリング技術	24
ダンス動画創作支援: 動画からの3次元ダンスモーション自動推定技術	25
歌唱動画創作支援: 1枚画像からの顔の3次元モデリングおよび表情合成技術	26
歌唱動画創作支援: 音楽からの個性ある歌唱アニメーション生成技術	27
音楽理解技術に基づくリズムを利用した体験型コンテンツの創出	28
音楽を聴く行為に着目した音楽キュレーション技術	29
文字の数式化と融合による未来の歌詞表現の実現	30
ダンス理解のための基礎的研究と応用のためのデータベース構築	31
Kiite: 楽曲と出会う音楽発掘サービス	32
OngaACCEL プロジェクト研究参加者一覧	33

OngaACCEL 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」 の狙いと研究成果の紹介

後藤 真孝

産業技術総合研究所

OngaACCEL プロジェクトでは、音楽体験の未来を切り拓く技術開発によって、音楽の楽しみ方がより能動的で豊かになるように、鑑賞・創作・協創の三つの側面から人々を支援できるサービスプラットフォームの構築を目指している。これまでに基礎から応用まで幅広い研究成果を創出し、学術論文として発表するだけでなく、視聴者やクリエイターが利用できるウェブサービス群を開発して公開してきた。さらに、ライブコンサートやイベント等での実証実験を通じて技術を洗練させてきた。今後もさらなる研究開発により、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. OngaACCEL プロジェクトの狙い

本研究プロジェクト「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: 吉井 和佳、森島 繁生、中村 聡史、佐々木 渉) では、図 1 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してきた。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」も構築し

てきた。これらを通じて、学術的・産業的・社会的・文化的にさまざまな貢献をし、次世代のメディアコンテンツ産業の発展にも貢献する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。

この OngaACCEL プロジェクトは、その前に 5 年間取り組んだ研究プロジェクト「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) の研究成果が土台となっている。OngaCREST では、音楽の音響信号から構成要素を推定する音楽理解技術

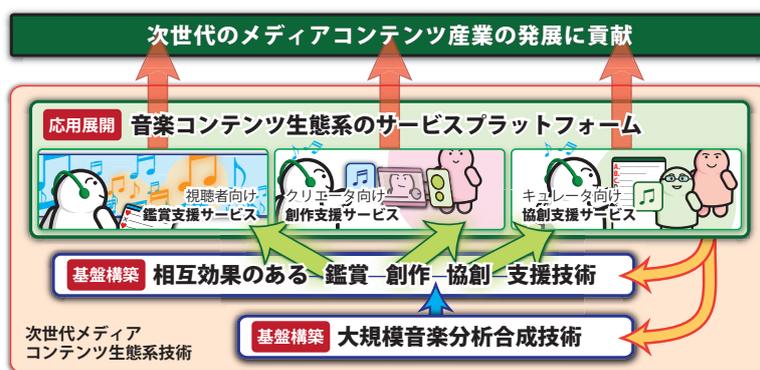


図 1: OngaACCEL プロジェクトの概要

によってデジタル音楽の真価を引き出し、人々がより能動的に音楽を楽しめるようにする技術の研究開発に取り組んだ。そうした研究成果を活かして、ACCEL プロジェクトでは大規模なメディアコンテンツから価値を創出するためのプラットフォームを構築している。そして、音楽の中身を自動解析する技術により、新たな価値を生み出し貢献することを目指している。

2. OngaACCEL シンポジウム 2020 で紹介する研究成果

OngaACCEL プロジェクトでは、コンテンツの膨大な蓄積が土壌となってコンテンツの豊かな生態系を生み出し、新たな価値を創出するための技術を「次世代メディアコンテンツ生態系技術」と名付け、音楽の楽しみ方がより能動的で豊かになるための研究開発を進めてきた。今回の公開シンポジウム「OngaACCEL シンポジウム 2020: 音楽情報処理研究が切り拓く未来の音楽体験」では、OngaCREST プロジェクトの成果も土台にしつつ、2016 年 8 月以降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 23 件のポスター題名を『』に囲んで概要を説明する。鑑賞・創作・協創の三つの側面から人々を支援することに取り組んできたため、2.1 節では鑑賞支援、2.2 節では創作支援、2.3 節では協創支援のための技術・サービスに関する研究成果を紹介する。

2.1 鑑賞支援技術・サービス

鑑賞支援では、音楽理解技術により大規模な楽曲を分析し、好みの音楽をより能動的で多様に楽しむことができるようにするための研究開発を中心に進めてきた。

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 Sync: 大規模音楽連動制御プラットフォームとその活用事例』である。この「Songle Sync (ソングルシンク)」(<https://api.songle.jp/sync>) [1] は、産業技術総合研究所(産総研)と科学技術振興機構(JST)が共同で 2017 年 8 月 2 日にプレス発表をした。これは、産総研の音楽理解技術と新たに開発した大規模音楽連動制御技術を融合させた、音楽に連動した演出を容易に利用・開発できるプラットフォームである。一般公開をして様々な実証実験に取り組みながら、機能と性能を改良して

きている。

この Songle Sync や、後述する Songrium、Text Alive、Kiite 等の様々なウェブサービスの基盤となっているプラットフォームが、『Songle: 楽曲の中身を自動解析する能動的音楽鑑賞サービス』である。この「Songle (ソングル)」(<https://songle.jp>) は、音楽の中身を自動解析する技術により、新たな価値を生み出して貢献するための中核となるプラットフォーム的サービスであり、かつ、音楽の中身を可視化することで人々が音楽をより深く理解して能動的に楽しめるようにするサービスでもある。Songle の解析結果を外部から利用するための API を提供しており、それにより上記のような様々なサービスを実現可能にしている。

Songle は、『音楽理解技術に基づくリズムを利用した体験型コンテンツの創出』でも活用されており、「高齢者向け音楽体験システム」では音楽に合わせて運動を促すことを目的に、専用デバイスを踏んだりスマートデバイスをタップしたりしてリズムを楽しむ音楽ゲームを、Songle による解析結果に基づいて実現した。さらに「JusTune」(<https://justune.net>) でも、マイクロタスクを取り込んだ音楽ゲームを Songle の API で実現した。

他にプレス発表をした代表的な成果は、『Lyric Jumper: 歌詞トピック解析技術に基づく歌詞探索サービス』である。この「Lyric Jumper (リリックジャンパー)」(<https://lyric-jumper.petitlyrics.com>) [2] は、株式会社シンクパワーと産総研、JST が共同で 2017 年 2 月 21 日にプレス発表をした。これは、シンクパワーの「プチリリ」が配信する歌詞データの一部(約 18 万曲)を、産総研が開発した歌詞トピック解析技術を用いて自動解析し、歌詞に出てくる単語の出現の仕方から、楽曲ごとのトピックを自動推定することで実現した歌詞探索ツールである。

鑑賞支援のためのウェブサービスとしては、「Songrium (ソングリウム)」(<https://songrium.jp>) も一般公開しており、『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』では、新たに派生作品の創作を引き起こした要因を推定して可視化する技術 [3][4] を実現した。

『音楽推薦・検索に基づく音楽鑑賞支援』では、異種クエリの混合が可能な音楽コンテンツ探索技術「Query-by-Blending」 [5]、作業用 BGM 推薦システム「FocusMusicRecommender」 [6]、ユーザ属性(国

籍・年齢・性別)の匿名性を守るために音楽再生履歴をカモフラージュする技術「Listener Anonymizer」[7]、コンテンツと推薦理由を同時に多様化する推薦技術「DualDiv」[8]、消費済みコンテンツに対する推薦理由の提示フレームワーク「ER for Repeat Consumption」[9]、創作行動ログが消費行動支援に反映される推薦技術「ABCPreC」[10]等の、多様な鑑賞支援技術を実現した。

こうした多様な鑑賞支援を可能にするための基礎技術として、音楽の中身を自動解析する技術は重要であり、メロディに対応する歌声の楽譜を推定する『CRNN-HSMM ハイブリッドモデルに基づく歌声採譜』[11]、コードを推定する『変分オートエンコーダーに基づく半教師あり深層コード推定』[12]、ドラム譜を推定する『事前学習済み言語モデルを用いた正則化に基づくドラム採譜』[13]、音楽構造を推定する『LSTM-HSMM ハイブリッドモデルに基づく音楽構造解析』[14][15]、ピアノ楽曲から楽譜を採譜する『深層多重音検出に基づくピアノ採譜』[16]を実現した。他にも、『音楽音響信号中の様々な要素を推定する音楽解析技術』では、音楽のビート推定[17]や音楽印象推定、楽器音識別[18]、歌声分離[19]、ループ素材の分解・再構成[20]等を実現した。

2.2 創作支援技術・サービス

創作支援では、コンテンツの生成技術を基礎技術として、人間の能力を補いつつ、好みに合ったコンテンツを自在に生み出せるようにするための研究開発を中心に進めてきた。楽曲の創作支援にも取り組んだが、動画共有サービス等を中心に楽曲を発表する際に映像を伴うことも多く、その音楽動画を創作する支援の需要が高いため、以下ではその研究成果を紹介する。

創作支援技術に関してウェブサービスとして公開している代表的な成果は、『TextAlive: 音楽に同期したリリックビデオ制作支援サービスとその活用事例』[21]である。この「TextAlive(テキストアライブ)」(<https://textalive.jp>)では、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、リリックビデオ(歌詞アニメーション)を制作可能にした。様々な実証実験に取り組みながら、機能を改良してきている。

こうした歌詞の表現に関しては、文字の動きだけでなく、フォントも含めた表現を支援できると有用

であり、『文字の数式化と融合による未来の歌詞表現の実現』ではその基礎研究として、音楽印象に基づくフォント生成[22]や、数式化された文字への演算によるアニメーション表現[23]等を実現した。

CGキャラクターが歌ったり踊ったりする音楽動画の創作支援では、フォトリアル(実写風)なキャラクターに焦点をあて、全身の写真1枚のみからその人物のダンス動画を、顔の写真1枚のみからその人物の歌唱動画を自動生成できる総合的なシステムを『フォトリアルなキャラクターのダンスおよび歌唱動作生成システム』で実現した。ダンス動画に関しては、『ダンス動画創作支援:1枚画像からの着衣全身3次元モデリング技術』[24][25]で写真から3次元復元してその人物の全身モデルを生成して、『ダンス動画創作支援:動画からの3次元ダンスモーション自動推定技術』で動画中のダンスモーションを推定して踊らせることが可能になった。歌唱動画に関しては、『歌唱動画創作支援:1枚画像からの顔の3次元モデリングおよび表情合成技術』[26]で写真からその人物の顔モデルを生成して、『歌唱動画創作支援:音楽からの個性ある歌唱アニメーション生成技術』[27]で歌声に合う歌唱動作を生成して歌わせることが可能になった。

動画中のダンスモーションを推定すると誤りが生じるが、その誤りを人手で修正する作業を支援するインタフェース[28]を『ダンス理解のための基礎的研究と応用のためのデータベース構築』で実現した。さらに、ダンスにおいて人はどこで自分を判断しており、機械はそれを判定可能かを明らかにした[29]。

他にも『身体モーションの情報処理と制作支援技術』では、ダンス情報処理の研究として、ダンスの類似性に基づくダンス楽曲検索インタフェース「Query-by-Dancing」[30]、グループダンスパフォーマンス動画作成インタフェース「DanceUnisoner」[31]、ダンス情報処理研究のためのストリートダンス動画データベース「AIST Dance Video Database」[32]、モーション間を任意のタイミングで滑らかに遷移させる手法[33]、身体モーション制作システム「OptiMo」[34]等を実現した。

2.3 協創支援技術・サービス

協創支援では、他の人たちにコンテンツの魅力伝える行為に着目し、音楽コンテンツの価値を増幅させていくようにするための研究開発を中心に進め

てきた。

協創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、『Kiite: 楽曲と出会える音楽発掘サービス』である。この「Kiite(キイテ)」(<https://kiite.jp>)は、クリプトン・フューチャー・メディア株式会社と産総研、JST が共同で 2019 年 8 月 30 日にプレス発表をした。これは、歌声合成技術による歌唱を含む膨大な楽曲（動画共有サービスによってインターネット上で公開されている約 32 万曲の音楽動画）を、「軽快」「激しい」などの自動分析された音楽の印象で視聴者が絞り込み、自動検出されたサビ区間を効率よく試聴することで、次々と新たな楽曲に出会うことができるサービスである。そうして発掘した好みの楽曲をプレイリストにボタン一押しで手軽に登録すると、お気に入りの楽曲が繰り返し鑑賞しやすくなるとともに、他の視聴者とも共有できる。また、音楽推薦エンジンが各視聴者の再生履歴やプレイリスト、楽曲の音響信号の自動解析結果などに基づいて日々自動生成する「お勧め楽曲のプレイリスト」を聴けば、幅広い音楽を楽しめる。さらに、視聴者自身が音楽推薦エンジンをカスタマイズして複数保存でき、気分に応じて使い分けられる高度な機能も備えている。同じ曲を同じ瞬間に楽しめる音楽発掘カフェ「Kiite Cafe (キイテカフェ)」(<https://cafe.kiite.jp>)も開発して公開した。

他にも『音楽を聴く行為に着目した音楽キュレーション技術』では、質問応答型音楽キュレーションサービス「mureq」(<https://mureq.com>)、音楽を庭に植え、育て、聴いてもらうサービス「otoniwa」、分岐型音楽キュレーションサービス「reco.mu」(<https://reco.mu>)^[35]、グループ型音楽鑑賞アプリ「Uniotto」(<https://uniotto.org>)^[36]のように、様々なアプローチから協創支援技術を実現した。

3. OngaACCEL プロジェクトが切り拓く未来の音楽体験

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きた。コンテンツの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 では鑑賞支援、創作支援、協創支援の三つの側面で、そうした体験を創出する研究開発に取り組んできた。まず鑑賞支援では、音楽再生時の可視化等により理解が深まる体験や、音楽に連動した総合的な体験、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が、コピー不可能な体験となる。次に創作支援では、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となる。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。クリエイターのみが得られる体験はコピー不可能である。さらに協創支援では、他の人たちと一緒に音楽を鑑賞する一体感や、自分の好きな曲を他の人たちが気に入る瞬間を見る喜びのような、他の人たちと一緒に創る音楽体験が、コピー不可能な体験となる。

コンテンツのデジタル化が進み膨大に蓄積されてきたが、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（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的体験」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。そのような量的だけでなく質的な変化に基づくコンテンツの真価は、まだ充分には引き出されていなかった。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、コンテンツの膨大な蓄積が土壌となって新たなコンテンツ生態系を生み出し、価値を創出するための「次世代メディアコンテンツ生態系技術」の研究開発に取り組んできた。今後も、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、文化的に貢献することを目指していく。

参考文献

- [1] Jun Kato, Masa Ogata, Takahiro Inoue, and Masataka Goto: Songle Sync: A Large-Scale Web-based Platform for Controlling Various Devices in Synchronization with Music, Proc. ACM Multimedia 2018.
- [2] Kosetsu Tsukuda, Keisuke Ishida, and Masataka Goto: Lyric Jumper: A Lyrics-Based Music Exploratory Web Service by Modeling Lyrics Generative Process, Proc. ISMIR 2017.
- [3] Kosetsu Tsukuda, Keisuke Ishida, Masahiro Hamasaki, and Masataka Goto: Songrium Derivation Factor Analysis: A Web Service for Browsing Derivation Factors by

- Modeling N-th Order Derivative Creation, *IEICE Trans. on Information and Systems*, Vol.E101-D, No.4, pp.1096-1106, 2018.
- [4] Kosetsu Tsukuda, Masahiro Hamasaki, and Masataka Goto: Modeling N-th Order Derivative Creation Based on Content Attractiveness and Time-Dependent Popularity, *IEICE Trans. on Information and Systems*, Vol.E103-D, No.5, 969-981, 2020.
- [5] Kento Watanabe and Masataka Goto: Query-by-Blending: A Music Exploration System Blending Latent Vector Representations of Lyric Word, Song Audio, and Artist, *Proc. ISMIR 2019*.
- [6] Hiromu Yakura, Tomoyasu Nakano, and Masataka Goto: FocusMusicRecommender: A System for Recommending Music to Listen to While Working, *Proc. ACM IUI 2018*.
- [7] Kosetsu Tsukuda, Satoru Fukayama, and Masataka Goto: Listener Anonymizer: Camouflaging Play Logs to Preserve User's Demographic Anonymity, *Proc. ISMIR 2018*.
- [8] Kosetsu Tsukuda and Masataka Goto: DualDiv: Diversifying Items and Explanation Styles in Explainable Hybrid Recommendation, *Proc. ACM RecSys 2019*.
- [9] Kosetsu Tsukuda and Masataka Goto, Explainable Recommendation for Repeat Consumption, *Proc. ACM RecSys 2020*.
- [10] Kosetsu Tsukuda, Satoru Fukayama, and Masataka Goto: ABCPRec: Adaptively Bridging Consumer and Producer Roles for User-Generated Content Recommendation, *Proc. ACM SIGIR 2019*.
- [11] Ryo Nishikimi, Eita Nakamura, Masataka Goto, Katsutoshi Itoyama, and Kazuyoshi Yoshii: Bayesian Singing Transcription Based on a Hierarchical Generative Model of Keys, Musical Notes, and F0 Trajectories, *IEEE/ACM Trans. on ASLP*, Vol.28, pp.1678-1691, 2020.
- [12] 呉 益明, Tristan Carsault, 中村 栄太, 吉井 和佳: 音楽音響信号に対するラベル・テキストチャ分離型変分自己符号化器を用いた半教師ありコード推定, *情報処理学会論文誌*, Vol.2019-MUS-124, 2019.
- [13] 石塚 峻斗, 錦見 亮, 中村 栄太, 吉井 和佳: 事前学習済み言語モデルによる正則化を用いた深層ニューラルネットワークに基づくドラム採譜, *情報処理学会論文誌*, Vol.2020-MUS-128, 2020.
- [14] Go Shibata, Ryo Nishikimi, and Kazuyoshi Yoshii: Music Structure Analysis Based on an LSTM-HSMM Hybrid Model, *Proc. ISMIR 2020*.
- [15] 柴田 剛, 錦見 亮, 中村 栄太, 吉井 和佳: 同質性・反復性・規則性を考慮した階層隠れセミマルコフモデルに基づく統計的音楽構造解析, *情報処理学会論文誌*, Vol.61, No.4, pp.757-767, 2020.
- [16] 柴田 健太郎, 中村 栄太, 錦見 亮, 吉井 和佳: 深層多重音検出を用いた音響信号から楽譜へのピアノ採譜, *情報処理学会論文誌*, Vol.2019-MUS-125, 2019.
- [17] Tian Cheng, Satoru Fukayama, and Masataka Goto: Convolving Gaussian Kernels for RNN-based Beat Tracking, *Proc. EUSIPCO 2018*.
- [18] Takumi Takahashi, Satoru Fukayama, and Masataka Goto: Instrudiver: A Music Visualization System Based on Automatically Recognized Instrumentation, *Proc. ISMIR 2018*.
- [19] Tomoyasu Nakano, Kazuyoshi Yoshii, Yiming Wu, Ryo Nishikimi, Kin Wah Edward Lin, and Masataka Goto: Joint Singing Pitch Estimation and Voice Separation Based on a Neural Harmonic Structure Renderer, *Proc. IEEE WASPAA 2019*.
- [20] Jordan B. L. Smith, Yuta Kawasaki, and Masataka Goto: Unmixer: An Interface for Extracting and Remixing Loops, *Proc. ISMIR 2019*.
- [21] Jun Kato and Masataka Goto: User-Generated Variables: Streamlined Interaction Design for Feature Requests and Implementations, *Companion to Programming 2017*, Article No.28, 2017.
- [22] Kosuke Nonaka, Junki Saito, and Satoshi Nakamura: Music Video Clip Impression Emphasis Method by Font Fusion Synchronized with Music, *Proc. ICEC-JCSG 2019*.
- [23] 佐藤 剣太, 中村 聡史, 鈴木 正明: 印象増幅のための手書き文字へのエフェクト付与手法, *情報処理学会論文誌*, Vol.59, No.9, pp.1761-1773, 2018.
- [24] Shunsuke Saito, Zeng Huang, Ryota Natsume, Shigeo Morishima, Angjoo Kanazawa, and Hao Li: PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization, *Proc. ICCV 2019*.
- [25] Ryota Natsume, Shunsuke Saito, Zeng Huang, Weikai Chen, Chongyang Ma, Hao Li, and Shigeo Morishima: SiCloPe: Silhouette-Based Clothed People, *Proc. CVPR 2019*.
- [26] Shugo Yamaguchi, Shunsuke Saito, Koki Nagano, Yajie Zhao, Weikai Chen, Shigeo Morishima, and Hao Li: High-Fidelity Facial Reflectance and Geometry Inference from an Unconstrained Image, *ACM Trans. on Graphics*, Vol.37, Issue 4, No.162, 2018.
- [27] 加藤 卓哉, 深山 覚, 中野 倫靖, 後藤 真孝, 森島 繁生: 歌声と楽曲構造を入力とした歌唱時の表情アニメーション自動生成手法, *画像電子学会誌*, Vol.48, No.2, pp.303-314, 2019.
- [28] 又吉 康綱, 小山 裕己, 深山 覚, 後藤 真孝, 中村 聡史: 画像の類似度を用いたダンス動画モーション訂正手法, *WISS 2019 論文集*, 2019.
- [29] Saeka Furuichi, Kazuki Abe, and Satoshi Nakamura: The Possibility of Personality Extraction Using Skeletal Information in Hip-hop Dance by Human or Machine, *Proc. INTERACT 2019*.
- [30] Shuhei Tsuchida, Satoru Fukayama, and Masataka Goto: Query-by-Dancing: A Dance Music Retrieval System Based on Body-Motion Similarity, *Proc. MMM 2019*.
- [31] 土田 修平, 深山 覚, 後藤 真孝: DanceUnisoner: グループダンスパフォーマンス動画作成インタフェース, *WISS 2019 論文集*, 2019.
- [32] Shuhei Tsuchida, Satoru Fukayama, Masahiro Hamasaki, and Masataka Goto: AIST Dance Video Database: Multi-genre, Multi-dancer, and Multi-camera Database for Dance Information Processing, *Proc. ISMIR 2019*.
- [33] Yuki Koyama and Masataka Goto: Precomputed Optimal One-Hop Motion Transition for Responsive Character Animation, *The Visual Computer*, Vol.35, No.6-8, pp.1131-1142, 2019.
- [34] Yuki Koyama and Masataka Goto: OptiMo: Optimization-Guided Motion Editing for Keyframe Character Animation, *Proc. ACM CHI 2018*.
- [35] 野中 滉介, 中村 聡史: 未知の音楽に誘導することを目的とした分岐型人力音楽推薦手法の提案, *情報処理学会論文誌*, Vol.2020-HCI-187, 2020.
- [36] 田村 柁優紀, 新納 真次郎, 白鳥 裕士, 田島 一樹, 中村 聡史: Uniott: グループ型音楽鑑賞手法の提案と実装, *GN Workshop 2017 論文集*, 2017.



OngaACCEL シンポジウム2020

音楽情報処理研究が切り拓く
未来の音楽体験

ポスター集

Songle: 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 川崎 裕太 井上 隆広 Tian Cheng 高橋 卓見 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

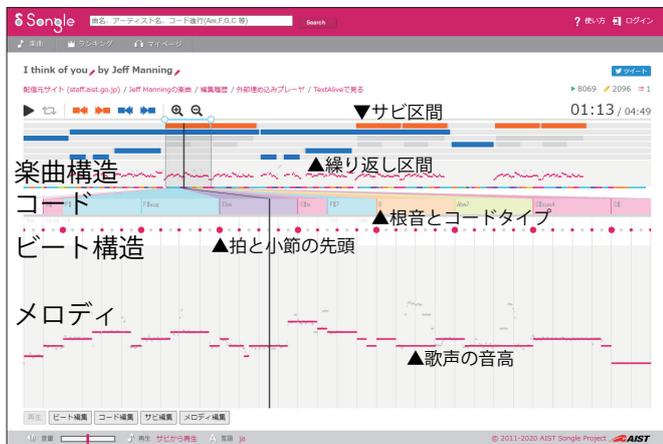
能動的音楽鑑賞サービス Songle (https://songle.jp)

- 音楽の中身を自動解析する技術により、新たな価値を生み出して貢献するための中核となるプラットフォーム的サービス
- 音楽の中身を可視化することで人々が音楽をより深く理解して能動的に楽しめるようにするサービス



- ニコニコ動画、YouTube の音楽動画、ピアプロ、MP3 に対応
180 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)

- 代表的な音楽的要素を自動解析して音楽の中身を可視化した「音楽地図」を表示
楽曲構造(サビ区間と繰り返し区間)
ビート構造(拍と小節の先頭)
コード(根音とコードタイプ)
メロディ(歌声の音高)



- 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

音楽地図訂正インターフェース

音楽地図の誤りを訂正して保存・共有可能なインターフェース

- 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できる

モバイル版 Songle 音楽地図訂正インターフェース

- 2020 年 7 月に公開
- スマートフォン上でサビ区間とビート構造を訂正可能
- モバイル版 Kiite のサビ出し機能で解析誤りに気づいたユーザがその場でサビ区間を訂正できる



Songle API

Songle の解析結果を外部から利用するため API を提供

- 音楽連動制御のためのウェブベースの開発フレームワーク
- 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
- 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 API

- 解析結果に基づく音楽連動制御を手軽に利用できる外部埋め込みプレーヤ

Songle Sync API

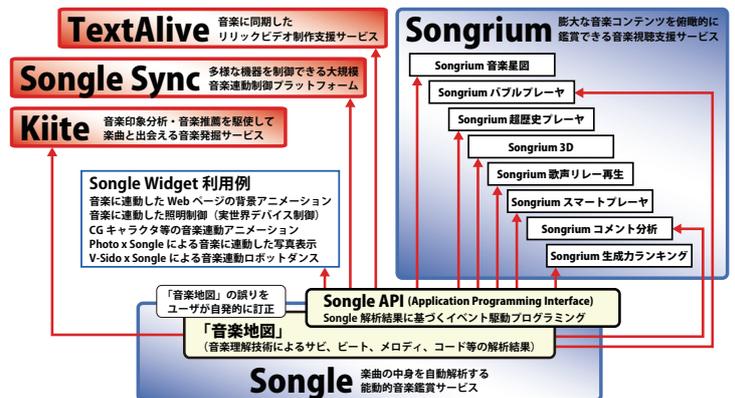
- 解析結果に基づく大規模な音楽連動制御を可能にするプラットフォーム Songle Sync のための API (Songle Sync アプリ開発者向けキットも提供)

Songle REST API

- 解析結果を JSON や XML 形式で取得して利用することができる RESTful API

プラットフォームの役割

Songle がプラットフォームとして様々なサービスを実現可能に



Songle プライベート楽曲機能

非公開状態で楽曲をアップロードして自動解析・訂正する機能

- 2019 年 7 月に公開
- 従来はインターネット上で公開されている楽曲しか扱えなかったのに対し、自分だけが閲覧することができる状態で任意の楽曲を解析できる機能
- TextAlive 等からも「Songle プライベート楽曲機能」を利用可能

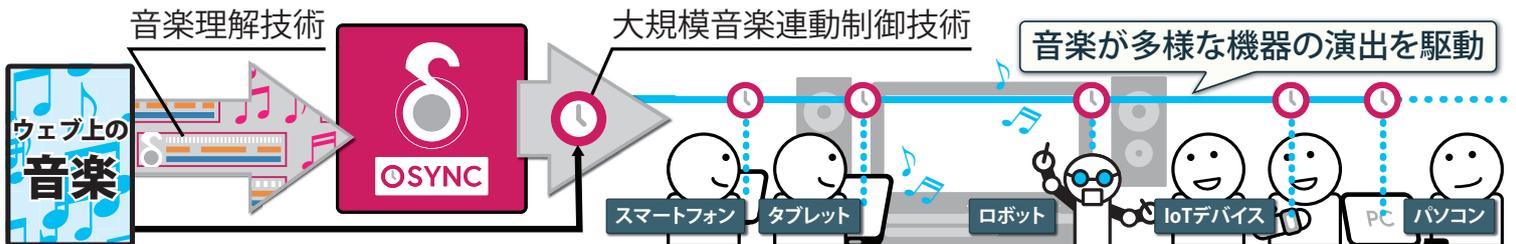


Songle Sync: 大規模音楽連動制御プラットフォームとその活用事例

加藤 淳 井上 隆広 後藤 真孝 石田 啓介 渡邊 研斗
濱崎 雅弘 笠井 志麻 川崎 裕太 尾形 正泰

産業技術総合研究所

Songle Sync (<https://api.songle.jp/sync>) とは



ウェブ上の音楽の再生に合わせて多種多様な機器を同時制御することで、一体感のある演出ができる大規模音楽連動制御プラットフォーム

実証実験

回線種別 (Wi-Fi・携帯)、回線状況 (ライブハウス・展示場・屋外) を変えて応用範囲 (DJ ステージ、待ち時間演出、花火との同期など) を検証



1 マジカルミライ 2017 keisei× ミクダヨー DJ ステージ - Songle Sync コラボ - 2017/9/2



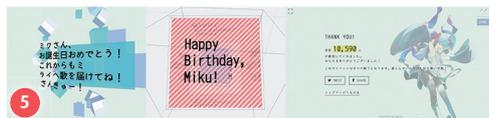
4 マジカルミライ 2018/2019 初音ミクバースデーメッセージ企画 - Songle Sync コラボ - 各年 8/31



5 LIGHT UP NIPPON HOKKAIDO feat. 初音ミク モエレ沼 芸術花火 2019/8/25, 2019/9/7
花火大会 2 件で花火と音楽に同期した演出を表示



2 SNOW MIKU LIVE! 2018 開演前演出 2018/2/9-2/10



5 Digital Stars presents 初音ミクバースデーメッセージ企画 - Songle Sync コラボ - 2020/8/31



3,4 北海道どこでも盆踊り with Domingo 2020/8/8-8/16
遠隔地の盆踊りを同期

Songle Sync アプリ開発者向けキット

- JavaScript でプログラミングできれば Songle Sync を活用したアプリケーションの開発が可能キット
- 2017/8/2 プレスリリースと同時に、実証実験として無償で配布開始

API (Application Programming Interface)

Songleの解析結果に基づくイベント駆動プログラミング

```
player.on("play", listener);
player.on("beatEnter", listener);
player.on("chorusSectionEnter", listener);
...
```

多様な機器を容易に連動させられるJavaScriptライブラリ

```
let player = new require("songle-api").Player();
```



チュートリアル



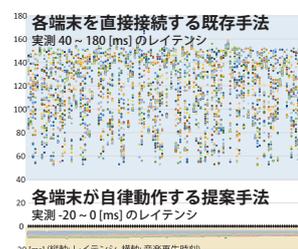
<https://tutorial.songle.jp/sync>

サンプルプログラム

- ウェブブラウザ向け (携帯端末・パソコン)
- IoTデバイス向け
- ロボット向け

<https://github.com/SongleJp/songle-api>

性能評価



- 既存手法と比較してレイテンシ大幅減
- 動画配信と比較して通信量大幅減
- Wi-Fi/4G LTE 実機で実用範囲のジッタ

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 佃 洸撰 中野 倫靖
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 (https://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

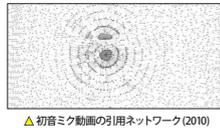
Web で発表、共有、視聴され、派生作品が生まれる音楽

- 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 **21 万曲以上**、そこから生まれた派生動画は **82 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



派生要因分析

派生作品の創作を引き起こした要因を推定して可視化

- オリジナル作品の魅力・オリジナル作品の人気・派生作品の人気の 3 種類の要因を考慮
- クリエイターの膨大な作品履歴から各要因の各クリエイターに対する影響の大きさを確率的に推定する手法を開発

パラメータ推定による派生要因推定のための基礎技術

- 派生コンテンツの全データが観測される確率が最大になるように確率的 EM アルゴリズムを用いてパラメータ推定

$$P(D, Z|O, C, \lambda, \beta, a, b) = \int P(D, Z|O, C, \alpha, \omega, \sigma, \theta, \lambda) P(\alpha|a, b) P(\omega|a, b) P(\sigma|a, b) P(\theta|\beta) d\alpha d\omega d\sigma d\theta$$

D: 全オリジナル作品の全派生作品, Z: 各派生作品の創作を引き起こした要因
O: オリジナル作品のランキング入りイベント, C: 派生作品のランキング入りイベント

Songrium 派生要因分析 α 版



派生作品の詳細

派生作品ツリー：各派生作品がオリジナル作品またはいずれかの派生作品の影響を受けて創作された様子を可視化。
人気影響度推移グラフ：派生作品の創作に対してコンテンツの人気が与えた影響の大きさの1ヶ月ごとの推移を可視化。
クリエイター特性マップ：派生作品を創作した各クリエイターが3種類の各要因を日頃どの程度重視しているかを可視化。

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

Songrium 音楽星図では、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 21 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に基づく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。



オリジナル作品のマッピング

提案手法で推定されるオリジナル作品の性質に応じてオリジナル作品を二次元空間にマッピング。クリエイターは自分がどのような曲の派生作品を作りたいかに応じて「定番曲」「チャレンジ曲」「玄人受け曲」「流行候補曲」といった新しい観点からオリジナル作品を探ることが可能。



クリエイターの詳細

クリエイターがこれまでに創作した各派生作品に関して「派生要因となった可能性の高い動画」および「派生影響を与えた動画」が表示される。オリジナル作品から影響を受けている派生作品や、影響を与えた派生作品といった情報の俯瞰が可能。



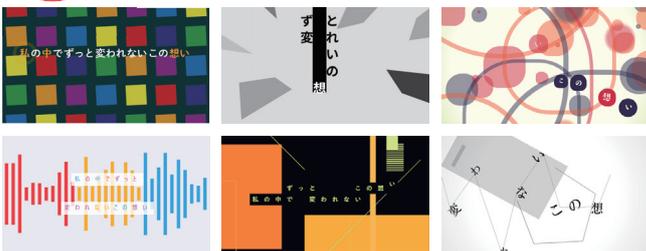
TextAlive: 音楽に同期したリリックビデオ制作支援サービスとその活用事例

加藤 淳 後藤 真孝 中野 倫靖 川崎 裕太

産業技術総合研究所

TextAlive (<https://textalive.jp>) でできること

1 楽曲+歌詞= リリックビデオを自動生成



楽曲と歌詞を指定すれば、**歌詞がタイミングよくアニメーションするリリックビデオ**を自動生成して誰でも楽しめる
多様な演出スタイルから曲に合うものを選び、フォントや配色、動きなどのパラメタを調整して好みの動画を作る

2 Web ブラウザ上で自由に編集



動画制作者は、直感的なユーザインタフェースで演出を詳しく思い通りに変更できる

既存の作品も自由に編集して、N次創作を手軽に楽しめる

3 プログラミングで表現力を拡張



プログラマは、演出内容を決めるアルゴリズムをアニメーション再生中に編集できる

TextAliveの機能を強化して、ユーザ全員の演出力を高められる

多様なユーザのための創作支援技術

2015/9/8 プレスリリース以降 33 回(2020/8/15 時点)バージョンアップを重ね、視聴者、楽曲・動画制作者など**多様なユーザのための創作支援技術を継続的に開発**

プログラマとユーザ間のコミュニケーション支援技術

- プログラマの指定したパラメタをユーザが自由に調整可能 (Live Tuning)
- 新たなパラメタをユーザが変数宣言として提案可能 (User-Generated Variables)

Songle プライベート楽曲などのクリエイター支援機能

- MP3 ファイルと歌詞入力で外部非公開のリリックビデオを制作可能 (2019/8/2)
- 歌詞タイミングを VSQX ファイルからインポートすることも可能 (2019/7/30)
- 楽曲・歌詞・イラストのライセンスが許せば、レンダリングサーバで動画をレンダリングして MP4・WebM 形式でダウンロード可能 (2016/10/14)

Songle Sync 連携機能

- 大規模音楽連動プラットフォーム Songle Sync と連携し、TextAlive 上のリリックビデオはすべて、QR コードにより多数の端末で同時に楽しめる (2019/8/8)

イベントなどでのアウトリーチ活動

- TextAlive を多くの方々に利用いただき研究に活かすため、解説小冊子の制作、技術系同人誌即売会での無償頒布、SNS アカウント運用など多彩な広報活動を展開
- 様々なイベントで TextAlive による動画制作方法を解説した**講座を計 6 回開講**



初音ミク マジカルミライ 2019 TextAlive 講座 頒布物・盛況の会場・鼎談の様子

商業ライブなどでの活用事例



No Maps 2016 IDOL DIVERSITY ベルリン少女ハート
アイドルライブ背景演出「asthma」「ROOM 24-7」

SNOW MIKU LIVE! 2018
背景映像演出「愛の詩」



SNOW MIKU LIVE! 2018
背景映像演出「四角い地球を丸くする」

初音ミク マジカルミライ 2018
背景映像演出「METEOR」

これまでの実証実験等一覧

- 2016/10/14-10/16 No Maps 2016 Trade Show 出展
No Maps 2016 IDOL DIVERSITY ライブ演出 2 曲に利用
- 2018/2/9-2/10 SNOW MIKU LIVE! 2018 ライブ演出 2 曲に利用
- 2018/8/25-8/26, 8/31-9/2 初音ミク マジカルミライ 2018 大阪 / 東京
ライブ演出 1 曲に利用、TextAlive 講座 3 回開講
- 2019/2/9-2/10 SNOW MIKU LIVE! 2019 ライブ演出 2 曲に利用
- 2019/6/17 ジュリアナの楽リ ライブ演出 2 曲に利用
- 2019/8/9-8/11, 8/30-9/1 初音ミク マジカルミライ 2019 大阪 / 東京
ライブ演出 3 曲に利用、TextAlive 講座 2 回開講
- 2019/11/13-11/15 Inter BEE 2019 株式会社モリサワブースにデモ出展
- 2020/8/15 ニコニコネット超会議 2020 夏 JNCA 主催 紹介番組放映

Lyric Jumper: 歌詞トピック解析技術に基づく歌詞探索サービス

佃 洸撰 石田 啓介 後藤 真孝
産業技術総合研究所

Lyric Jumper とは

様々な歌詞やアーティストに出会える新しい歌詞探索サービス

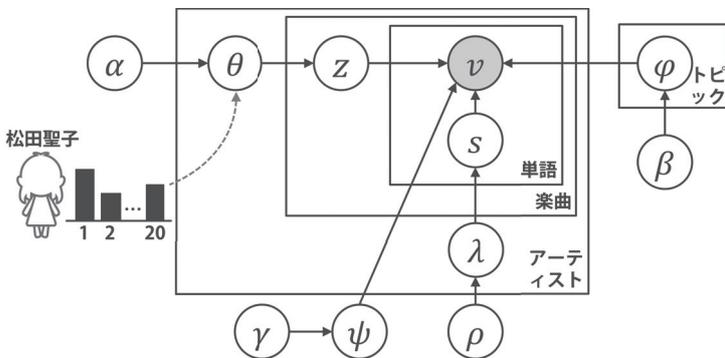
<https://lyric-jumper.petitlyrics.com>

- 歌詞のトピックを可視化
- アーティストごとの歌詞の傾向に基づく歌詞探索
- 株式会社シンクパワーと共同開発
- 同社が運営する歌詞配信サービス「ブチリリ」の日本語歌詞 18 万曲を使用

歌詞トピック推定

提案モデル: 歌詞データ全体での単語の出現傾向を利用し傾向の差異が端的に表現されるようトピックを決定

- アーティスト・楽曲・単語という 3 階層の構造を考慮
- アーティストごとにトピックの分布 θ を持つ
- 楽曲ごとに一つのトピック z を持つ

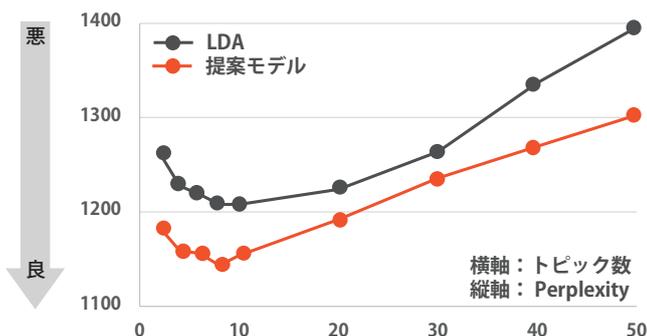


歌詞トピックが k である確率

$$P(z_{ar} = k | D, X, Z_{\setminus ar}, \alpha, \beta, \gamma, \rho) \propto \frac{R_{ak} \lambda_{ar} + \alpha}{R_a - 1 + \alpha k} \times \frac{\Gamma(N_{k \setminus ar} + \beta | V|)}{\Gamma(N_{k \setminus ar} + N_{ar} + \beta | V|)} \times \prod_{v \in V} \frac{\Gamma(N_{kv \setminus ar} + N_{arv} + \beta)}{\Gamma(N_{kv \setminus ar} + \beta)}$$

評価実験

- 登録歌詞数の上位 1000 アーティストの 93,716 件の歌詞を使用



歌詞のトピック推定のデファクトスタンダードである LDA よりも提案モデルが高精度であることを示した

謝辞: 株式会社シンクパワー (Lyric Jumper の共同運用・歌詞データ提供)

Lyric Jumper の様々な機能



歌詞フレーズ発掘機能
フレーズ発掘ボタンを押すとアーティストの様々な歌詞の中からトピックと関連が深いフレーズを抜き出して次々と表示

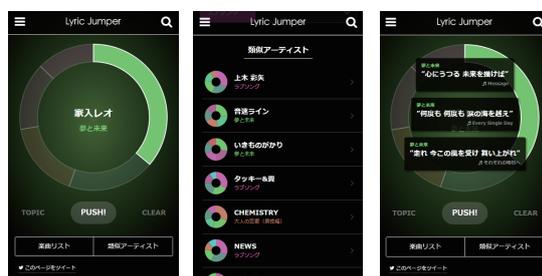


歌詞強調表示機能

トピックを参考に曲名を選択すると歌詞を表示する際にそのトピックと特に関連深いフレーズを強調表示

アーティストフィルタ機能

トピックを選択するとそのトピックに分類された持ち歌の比率の高さと曲数の多さに基づいて関連アーティストを表示



スマートフォン版



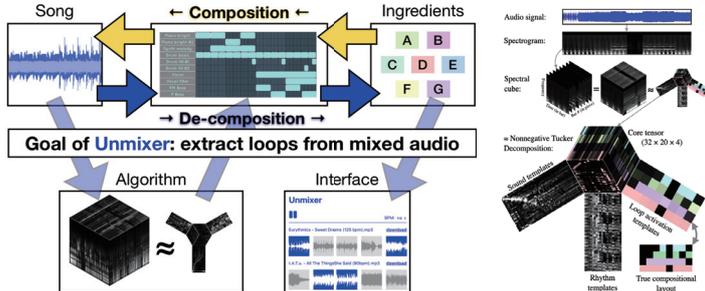
音楽音響信号中の様々な要素を推定する音楽解析技術

深山 覚 Tian Cheng 中野 倫靖 Jordan B. L. Smith 高橋 卓見 後藤 真孝
 川崎 裕太 吉井 和佳*1 呉 益明*1 錦見 亮*1 Kin Wah Edward Lin
 産業技術総合研究所 *1 京都大学

音楽のリミックス

Unmixer：音楽音響信号をループ素材に分解・再構成

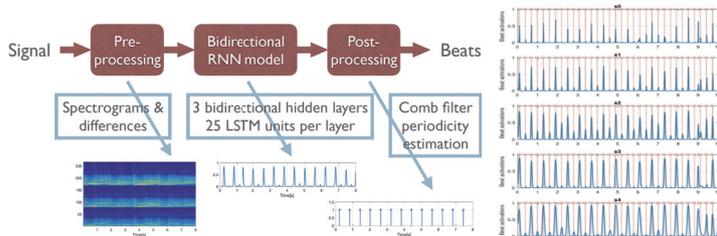
- ループ素材を組み合わせる創作過程を完成した音楽音響信号から逆推定
- 小節単位スペクトログラムに非負値タッカー分解を適用しループ素材を抽出



音楽のビート推定

ガウスクERNELを用いたビート推定器の高性能化

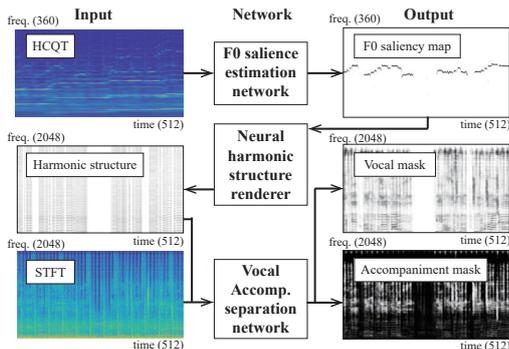
- ビート位置にガウスクERNELを畳み込んで深層学習による推定器を学習
- 学習データ中の正解・不正解データの不均衡を緩和しビート推定性能改善



音楽からの歌声分離

深層学習に基づく歌声調波構造モデルによる歌声分離

- 音楽の歌声音高推定と歌声分離を同時学習する深層ニューラルネットワーク
- 微分可能な Neural Harmonic Structure Renderer でネットワークを接続

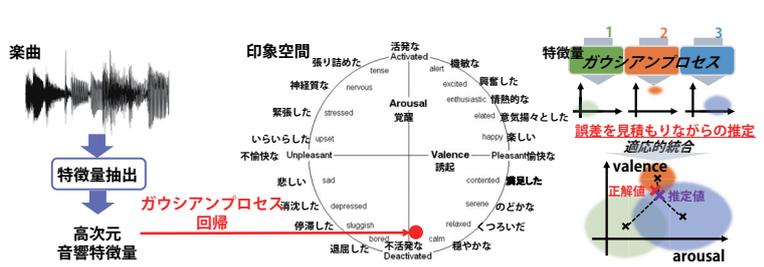


- F0 軌跡が高調波構造へ変換されて歌声の物理的意味の解釈を手助け

音楽印象推定

音楽を聞いて感じる印象を音楽音響信号から自動推定

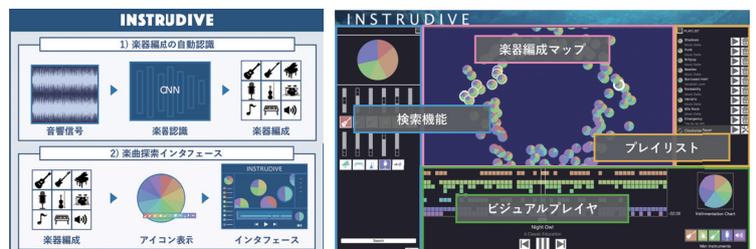
- 特徴量ごとの推定誤差をガウス過程回帰によって見積もって印象を推定
- 見積もった誤差の小さい特徴量に重みをつけて従来法より高性能で推定



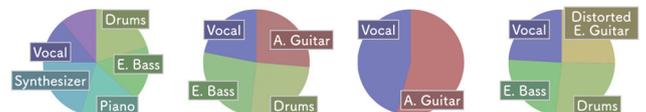
楽器音識別

INSTRUDIVE：楽器編成の自動認識に基づく楽曲探索システム

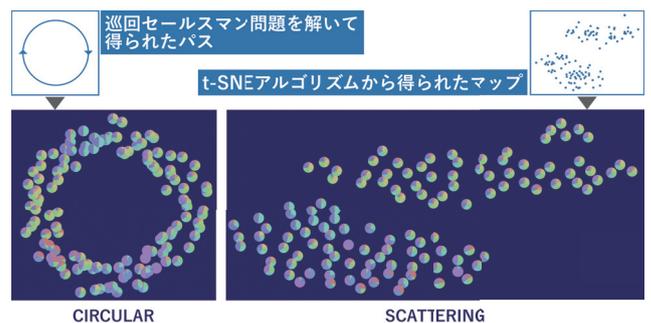
- 楽器編成を深層学習で自動認識：周波数・時間方向に個別にフィルタを設計
- 楽曲を楽器編成で探索可能：ジャンル横断型の新しい楽曲探索・鑑賞を実現



- 楽曲中の楽器が使われる割合をパイチャートで可視化



- 楽曲の聴き方(次々変化 or 同質)に応じた 2 種類の楽曲探索空間



音楽推薦・検索に基づく音楽鑑賞支援

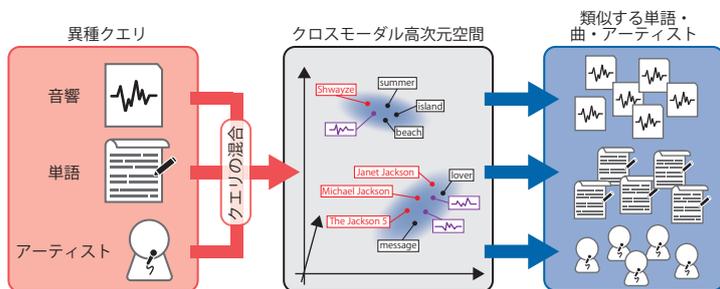
佃 洸撰 渡邊 研斗 矢倉 大夢 後藤 真孝 深山 覚 中野 倫靖

産業技術総合研究所

Query-by-Blending

異種クエリの混合が可能な音楽コンテンツ探索技術

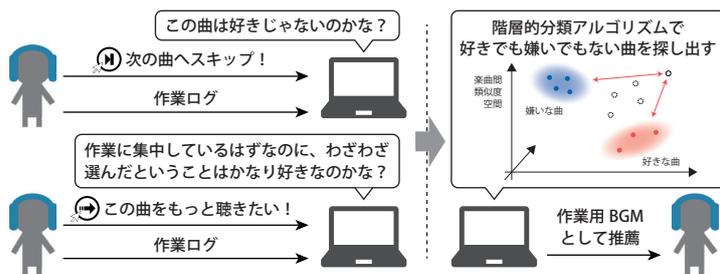
- 類似した音響 / 単語 / メタタグが近傍に配置された高次元空間を教師なし学習
- 音響 / 単語 / メタタグをクエリとして混ぜた新しい音楽コンテンツ探索が可能
- 言語処理分野の word2vec 等で用いられる分布仮説を音楽情報処理分野へ応用



FocusMusicRecommender

PC 作業中の集中を助けるための作業用 BGM 推薦システム

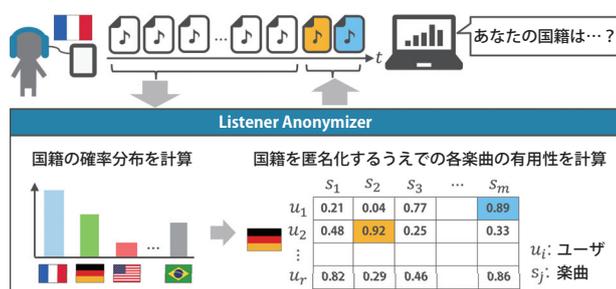
- 集中を保つには嫌いな曲だけでなく、好きな曲の聴取も避けるべきとされている
- 一方で、好きでも嫌いでもない微妙な曲を自ら見つけ出すのは難しい
- 作業中のフィードバックや楽曲間類似度から、好きでも嫌いでもなさそうなそれぞれに適した作業用 BGM を自動で見つけ出すシステムを提案



Listener Anonymizer

ユーザ属性(国籍・年齢・性別)の匿名性を守るために音楽再生履歴をカモフラージュする技術

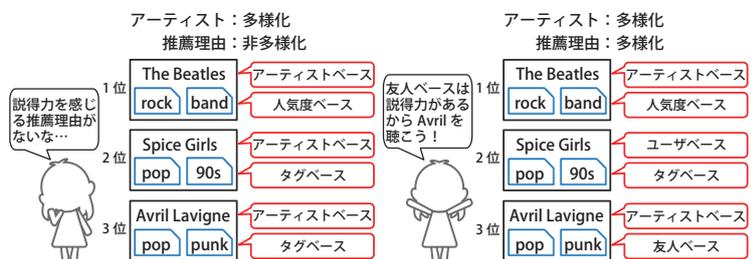
- できるだけ少ない曲数で効率的に匿名化可能な楽曲を選択・推薦
- 100 万件の再生ログを用いた機械学習における特徴選択技術により実現



DualDiv

コンテンツと推薦理由を同時に多様化する推薦技術

- コンテンツの多様化に加えて推薦理由も多様化された推薦が可能
- 推薦リスト中のアイテムがユーザに選択される確率を高める仕組みを実現
- 貪欲法によって実時間で多様化された推薦理由を選択する手法を提案



ER for Repeat Consumption

消費済みコンテンツに対する推薦理由の提示フレームワーク

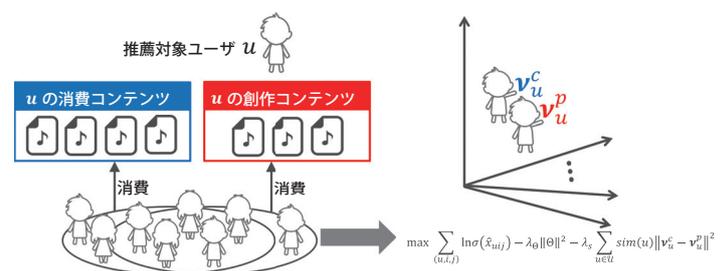
- ユーザが過去に消費したコンテンツを再度理由付きで推薦することでコンテンツの寿命を延ばすことが目的
- 推薦理由に影響を与える要因として Personal・Social・Item の 3 種類を提案
- 622 人を対象としたアンケート調査により提案フレームワークの有用性を検証

要因	スタイル	以下の理由で、過去に聴いた Taylor Swift の Shake It Off を推薦しました:
Personal factor	P-first	初めて Shake It Off を聴いてからちょうど 5 年が経過したからです
	P-last	最後に Shake It Off を聴いてからちょうど 3 年が経過したからです
	P-together	先程聴いた Lady Gaga の Applause と一緒に Shake It Off をよく聴いていたからです
Social factor	P-total	次に Shake It Off を聴くと合計再生回数が 100 回になるからです
	S-total	全ユーザの Shake It Off の総再生回数が 100 万回に到達したからです
Item factor	S-unique	Shake It Off を聴いたユニークユーザ数が 10 万人に到達したからです
	S-favorite	Shake It Off をお気に入り登録したユーザ数が 1 万人に到達したからです
Item factor	I-release	Shake It Off が発売されてからちょうど 5 年が経過したからです
	I-live	Taylor Swift のライブで今日 Shake It Off が歌われたからです

ABCPrac

創作行動ログが消費行動支援に反映される UGC 推薦技術

- ユーザの消費情報に加えて創作情報も考慮した推薦が可能
- コンテンツを創作することで消費コンテンツの推薦精度が上がる仕組みを実現
- ユーザの消費者ベクトルと創作者ベクトルを適応的に学習する手法を提案



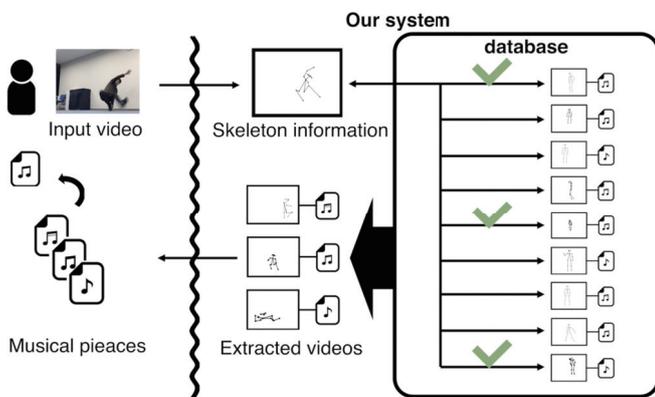
身体モーションの情報処理と制作支援技術

土田 修平 小山 裕己 深山 覚 後藤 真孝 濱崎 雅弘
産業技術総合研究所

Query-by-Dancing

ダンサーの姿勢の類似度に基づくダンス音楽検索

- ダンサーの姿勢時系列をクエリとしてダンス動画中のダンス音楽を検索
- 音楽ジャンルに依らず姿勢から直接的に音楽を検索し音楽鑑賞・ダンス創作可能
- 姿勢時系列の連続性とダンスのカテゴリの違いに敏感な姿勢特徴量を考案
- 姿勢情報を活用した新たな音楽鑑賞・ダンス創作支援を実現



AIST Dance Video Database

世界初のストリートダンス動画の大規模データベース

- 複数ダンスジャンル・複数ダンサー・複数カメラによるダンス動画を収録
- ダンス動作とダンス音楽の分析・生成をするダンス情報処理研究に有用
- 6種類のカメラ配置に加え1種類の移動カメラ配置による動画を収録
- オリジナルダンス動作・オリジナルダンス音楽を収録し研究目的に限り無償で <https://aistdancedb.ongaaccel.jp> より入手可能

🎬 動画 13,940 本 (ソロダンス 1,510 本 / グループダンス 108 本)

🎵 オリジナル楽曲 60 曲 (1 ダンスジャンルにつき 6 曲)

👤 ストリートのダンスジャンル 10 種類 (break, pop, lock, waack, middle hip-hop, krump, LA-style hip-hop, house, street jazz, ballet jazz)

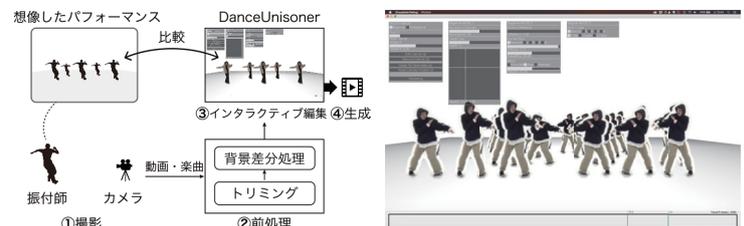
👤 プロダンサー 35 人 (男性 20 人 女性 15 人)

Category	Basic Dance	Advanced Dance	Group Dance	Moving Camera	Showcase	Cypher	Battle
Videos	10,798 (10x3x10x4x9-2)	1,890 (10x3x7x9)	900 (10x1x10x9)	300 (10x(2x3+1x4)x3)	27 (1x3x9)	10 (1x2x5)	15 (3x1x5)
Screenshot							
Genres	10	10	10	10			
Dancers / Groups	3 dancers	3 dancers	1 group (3 dancers)	3 dancers	3 groups (10 dancers)	1 group (10 dancers)	3 groups (2 dancers)
Dances	10 dances x 4 variations	7 dances	10 dances	3 or 4 dances	3 dances	2 sets	1 set
Duration	16 beats	64 beats	64 beats	64 beats	96 beats	0 min	5 min
Cameras							

DanceUnisoner

グループダンスパフォーマンス動画作成インターフェース

- 動画編集加工により想像が困難なグループダンスの検討・確認・共有を可能に
- 1人のダンス動画を複製・加工して多数配置しグループダンスを可視化
- 背景差分処理を施した動画を重量的に配置し音響指紋を利用して音楽と同期再生
- 集約された少ないパラメタの直感的な操作によりインタラクティブな編集加工



Optimal One-Hop Motion Transition

ターゲットシナリオ

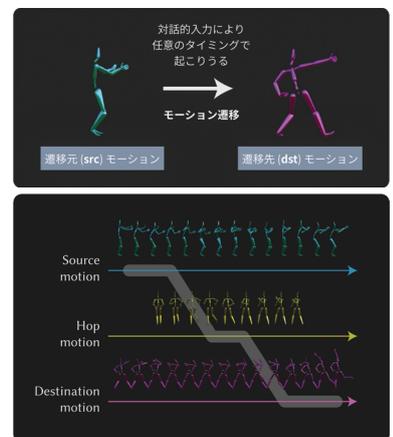
身体モーションデータを切り替えることで対話的にアバターを動作させる

既存アプローチの問題点

モーションデータを遷移させる際に、慣性を無視した不自然な動作や足の滑りなどが発生してしまう

提案：One-Hop Transition

二つのモーションデータを遷移させる際に**最適な中間モーションを計算して挿入**することで、全体として滑らかで自然なモーションを実現する



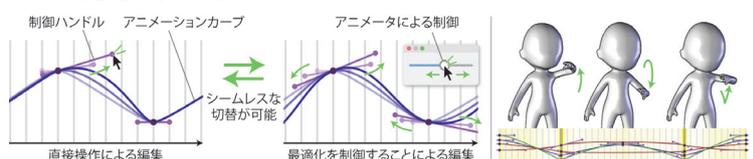
OptiMo

目的：最適化計算の活用

キーフレームに基づく身体モーション制作時に、**アニメーターが最適化計算を活用するために有効なシステム設計**を模索する

提案：身体モーション制作システム OptiMo

透明性 (何が起きているか分かる)・**制御可能性** (プロセスに能動的に介入できる)・**編集可能性** (得られた結果を容易に編集できる) の3つのシステム設計指針に基づき、**目的関数に関連する値を可視化する機能**や、**最適化の振る舞いを対話的に制御する機能**を搭載している



CRNN-HSMM ハイブリッドモデルに基づく歌声採譜

錦見 亮 中村 栄太 後藤 真孝*1 吉井 和佳

京都大学 *1 産業技術総合研究所

研究背景

歌声採譜

音楽音響信号から歌声が担うメロディの楽譜を推定

従来研究 | ピアノロール形式の推定

- 各音符を「半音単位の音高」と「**発音時刻**と**継続長**」で表現
- 発音時刻と継続長は**量子化されていない**



本研究 | 楽譜形式の推定

- 各音符を「半音単位の音高」と「**発音位置**と**継続長**」で表現
- 発音位置と音価は**タイタム**(e.g. 16 分音符)単位で表現



キーアイデア

深層モデルと確率モデルの統合

両モデルのデメリットを相互補完

深層モデル | 注意機構モデルに基づく歌声採譜

- Pros** | 音響スペクトログラムを直接扱える柔軟な表現力
- Cons** | 学習困難, 楽譜の整合性・妥当性の無保証



確率モデル | 隠れセミマルコフモデルに基づく歌声採譜

- Pros** | 楽譜の整合性・妥当性
- Cons** | 同音連打や休符が推定不能



提案手法

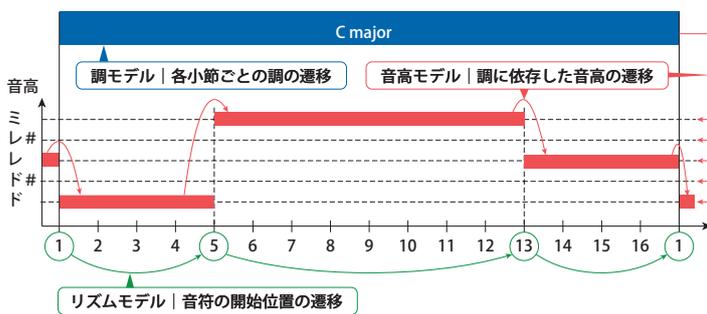
CRNN-HSMM ハイブリッドモデル

畳み込み再帰型ニューラルネットワーク (CRNN) に基づく音響モデルと隠れセミマルコフモデル (HSMM) に基づく楽譜モデル統合モデル



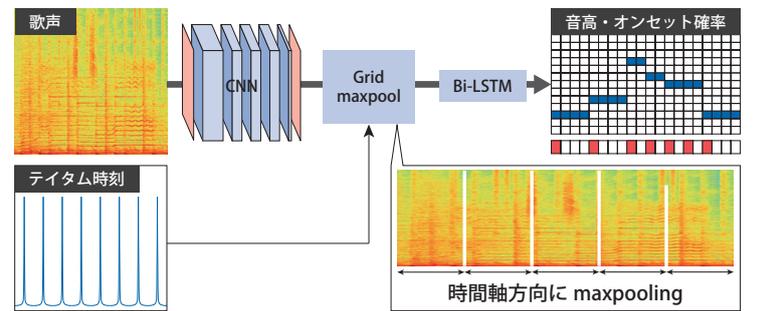
HSMM 楽譜モデル

調・音高・リズムの生成過程を表現



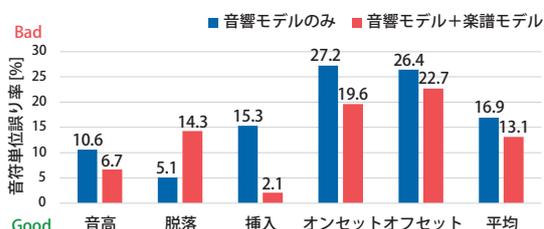
CRNN 音響モデル

音符列とスペクトログラムの関係を表現



実験結果

音響モデル (CRNN) に楽譜モデル (HSMM) を統合することにより精度が向上



変分オートエンコーダーに基づく 半教師あり深層コード推定

呉 益明

Tristan Carsault ^{*1}

中村 栄太

吉井 和佳

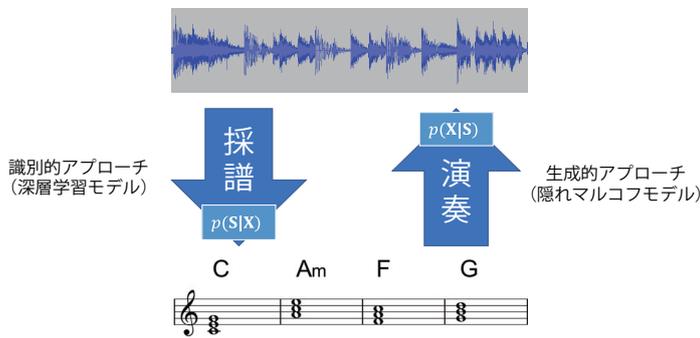
京都大学

^{*1}IRCAM

従来の研究

従来の研究

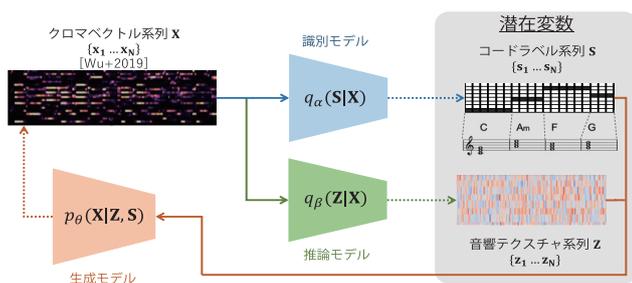
- 統計的機械学習手法に基づく
- 学習機構の定式化は識別的、生成的アプローチに大別される



生成モデルに基づく学習機構

変分オートエンコーダー (VAE) に基づく定式化

- コードラベル系列 (離散値変数) と潜在特徴系列 (連続値変数) を潜在変数とし、音響特徴量 (クロマベクトル) を確率的に生成する深層生成モデルを定義
- 深層コード推定モデルは、生成モデルの潜在変数事後分布の変分近似と位置付け



生成・認識過程の同時学習

$$\mathcal{L}_X(\theta, \alpha, \beta) \triangleq \mathbb{E}_{q_\alpha(S|X)q_\beta(Z|X,S)} [\log p_\theta(X|S,Z)] - KL[q_\beta(Z|X,S) || p(Z)]$$

$$+ \mathbb{E}_{q_\alpha(S|X)} [\log p_\theta(S)] - \mathbb{E}_{q_\alpha(S|X)} [\log q_\alpha(S|X)]$$

S、Zの事前分布による推定モデルの正則化

$$\mathcal{L}'_{X,S}(\theta, \beta) \triangleq \mathbb{E}_{q_\beta(Z|X,S)} [\log p_\theta(X|S,Z)] - KL[q_\beta(Z|X,S) || p(Z)]$$

$$\mathcal{L}_{X,S}(\theta, \alpha, \beta) \triangleq \mathcal{L}'_{X,S}(\theta, \beta) + \log q_\alpha(S|X)$$

推定モデルの教師あり学習

$$\text{目標関数} \Rightarrow \mathcal{L}(\theta, \alpha, \beta) \triangleq \sum_X \mathcal{L}_X(\theta, \alpha, \beta) + \sum_{X \text{ with } S} \mathcal{L}_{X,S}(\theta, \alpha, \beta)$$

マルコフ型言語モデルの統合

- 一次マルコフモデルを言語モデル (変数 S の事前分布) とし、コードラベル系列のラベル連続性を反映させる

- 遷移確率の設定: 自己遷移確率は 0.9、コード間遷移確率は一様

$$p_\phi(S) = p(s_1) \prod_{n=2}^N p(s_n | s_{n-1})$$

$$= \prod_{k=1}^K \phi_k^{s_{1k}} \prod_{n=2}^N \prod_{k'=1}^K \phi_{k'/k}^{s_{n-1,k'} s_{nk}}$$

- 推定分布を与えられたとき、言語モデル尤度の期待値は解析的に求めることができる

$$\gamma(s_1) = \log p(s_1),$$

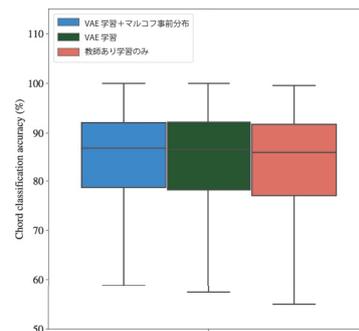
$$\gamma(s_n) = \sum_{s_{n-1}} q_\alpha(s_{n-1} | \mathbf{X}) (\gamma(s_{n-1}) + \log p_\phi(s_n | s_{n-1}))$$

$$\mathbb{E}_{q_\alpha(S|\mathbf{X})} [\log p_\phi(S)] = \sum_{S_N} q_\alpha(s_N | \mathbf{X}) \gamma(s_N).$$

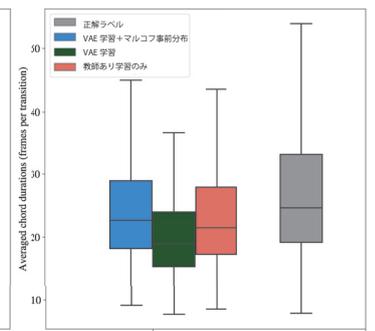
実験結果

正則化によるコード推定性能の向上

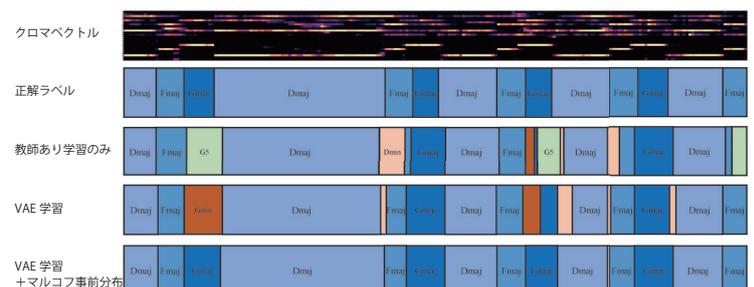
コード推定正解率の改善



マルコフ型言語モデルによるコードラベル連続性維持の効果



コード推定結果の例



事前学習済み言語モデルを用いた 正則化に基づくドラム採譜

石塚 峻斗 錦見 亮 中村 栄太 吉井 和佳

京都大学

研究背景

自動音楽採譜 (automatic music transcription; AMT)

ポピュラー音楽から楽譜を自動で書き起こす



応用場面

1. 音楽の創作活動支援



2. 耳コピの補助



自動ドラム採譜

ドラムの特徴

- ポピュラー楽曲の**音楽的な構造**を支える
- 複数楽器のうち **BD・SD・HH** の主要3パートに注目
- 音の高さ・長さより**オンセット時刻**が重要

問題設定

入力

- 音楽音響信号から得られるメルスペクトログラム
- 予め既存手法で推定されたビート時刻 ($F=96.4\%$)

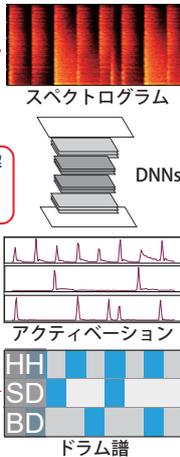
出力

- 16分音符単位のオンセット時刻

仮定

- 各楽器のオンセット時刻はティタム時刻に沿う

音楽のスパース性から非負値行列因子分解が利用されてきた。最近では **CNN** や **RNN** によって高い精度を達成している



本研究ではティタムを16分音符に設定

音楽言語モデル

スキップ型 Bi-gram

16ティタム前からの遷移を学習・推定

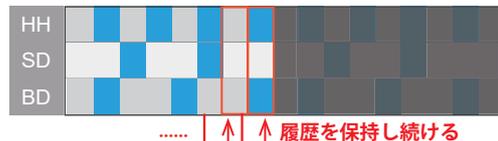


16ティタム前からの遷移

多くのポピュラー音楽は **4/4 拍子** で構成されることが多いという仮定に基づいている

GRU (gated recurrent unit)

各ティタム以前からの遷移を直接学習・推定



..... ↑ ↑ 履歴を保持し続ける

履歴を常に保持し続けるためスキップ型 Bi-gram と比べて **強力な予測性能** を有する

正則化に基づく学習法

Point 1 :

ティタム単位のドラム採譜

Point 2 :

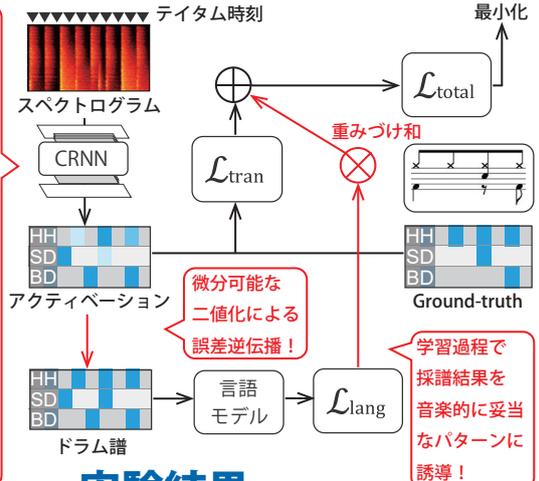
学習時における採譜出力の**正則化**

Point 3 :

言語モデルの設計 (Bigram or GRU)

Point 4 :

微分可能な二値化に基づく誤差逆伝播



実験結果

ビートの種類	推定ビート			正解ビート		
	F	P	R	F	P	R
評価値 (%)						
CRNN (従来法)	70.8	77.4	65.9	71.0	77.6	66.1
CRNN (提案法)	78.9	86.3	73.1	79.3	86.7	73.3
+Bi-gram	81.4	84.7	79.1	80.8	83.7	78.8
+GRU	81.6	84.0	80.2	81.1	83.2	79.7

- 従来の最高精度より約 **8% 向上!**
- 正則化により精度が約 **2% 向上!**
- Bi-gram よりも **GRU** の方が向上幅が大きい
- ビートの推定精度が良くない場合は更なる検討が必要

学習時における損失の正例に対する重み, 言語モデルの重み, RNN の層数・次元数のハイパーパラメータは検証データに対して最適化を行った

デモ



- 正則化が音楽的に**不自然な**ドラムパターンを効果的に防いでいることを確認
- Fill-in などの**少し複雑な**ドラムパターンを単純化しすぎてしまう傾向
- クラッシュシンバル**等と主要3パートの楽器と混同してしまう傾向
- 今後は RNN 部や言語モデルを self-attention に置き換えてドラムパターンのより**グローバルな構造**捉えることや**複数楽器**への対応が考えられる

LSTM-HSMM ハイブリッドモデルに基づく音楽構造解析

柴田剛 錦見亮 中村栄太 吉井和佳

京都大学

研究背景

音楽構造解析

- 楽曲の**セクション構造**の認識
意味上のまとまり(イントロ・Aメロ・サビなど)
- 試聴用音源の生成や楽曲再生時のナビゲーションに応用可能

セクションの持つ重要な性質

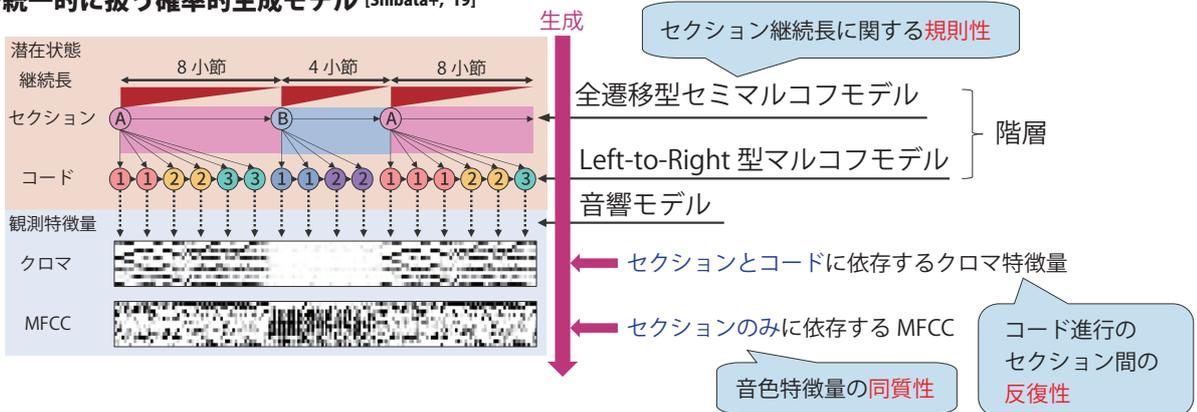
- セクション内での音響特性の**同質性**
- 同じラベルのセクション間での**反復性**
- セクション継続長の**規則性**

階層隠れセミマルコフモデル

同質性・反復性・規則性を統一的に扱う確率的生成モデル [Shibata+, '19]

- 楽曲の**セグメンテーション・クラスタリング**を行う
- イントロ・サビなどの**ラベル**はわからない!

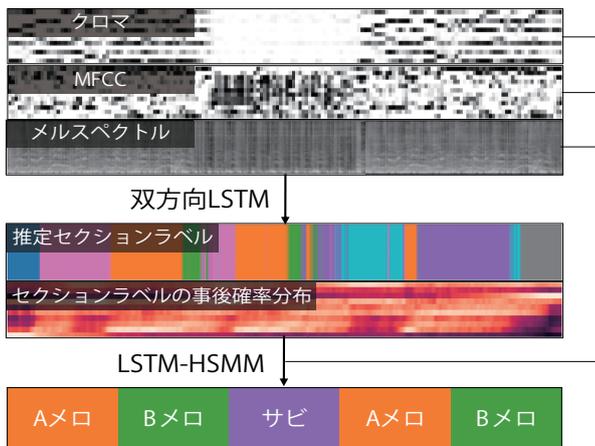
DNNによるラベル推定と統合



LSTM-HSMM ハイブリッドモデル

セグメンテーション・ラベリングを同時に処理

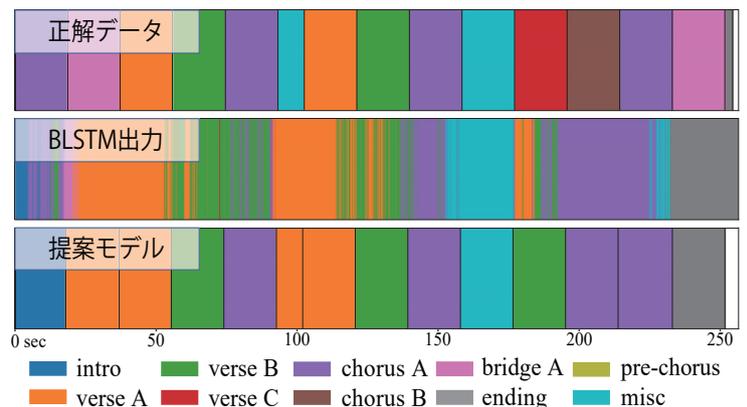
ラベリングを行う**双方向 LSTM** と階層隠れセミマルコフモデル (HSMM) の統合



- 双方向 LSTM
フレーム単位のメルスペクトルからセクションラベルを推定
- 階層 HSMM
ラベルの事後確率分布を利用してメルスペクトルの生成過程も表現
双方向 LSTM の結果を用いて初期化することで悪い初期値を避ける

実験結果

結果例



手法	セグメンテーション		クラスタリング	ラベリング
	$F_{0.5}$ (%)	$F_{3.0}$ (%)	F_{pair} (%)	精度 (%)
GS3	52.3	73.5	54.2	n/a
SUG2	25.8	73.7	37.3	n/a
FK2	30.0	65.7	63.4	n/a
Paulus'09	n/a	63.0	63.7	34.4
提案モデル	43.3	66.5	54.6	45.3

従来手法
より向上!

chorus (サビ) 始まりの曲や verse B/C の認識などが課題

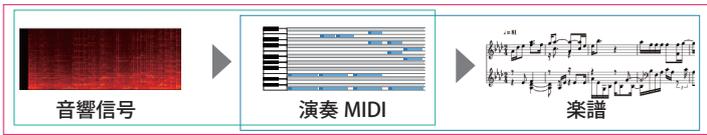
深層多重音検出に基づくピアノ採譜

柴田 健太郎 中村 栄太 錦見 亮 吉井 和佳

京都大学

背景

問題設定：音響信号からピアノ楽譜を推定



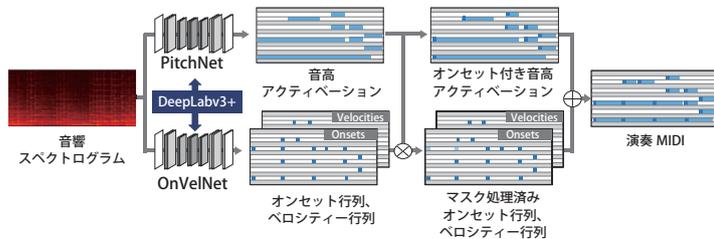
本研究の目標は五線譜で表現された楽譜の推定

- 応用：音楽演奏の手助け、自動作曲システム、楽曲検索など
- ほとんどの従来研究では「採譜 = 演奏 MIDI 推定」
- DNN による演奏 MIDI 推定手法の発展 [Hawthorne+ 18, Wu+ 19 など]
- 統計モデルによる演奏 MIDI から楽譜推定手法の発展 [Nakamura+ 18 など]

システム構成



提案法 1 多重音検出モデル



DeepLabv3+ を拡張

- CNN 約 80 層からなる巨大 ResNet
- オンセット・ベロシティー (音強) も推定可能に
- 同音連打を表現可能

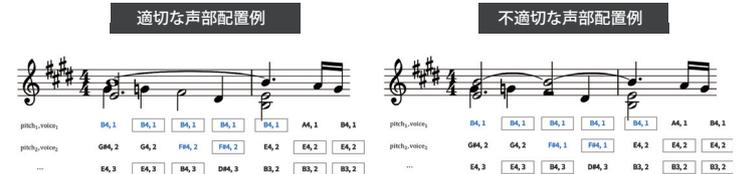
損失関数

- 音高アクティベーション、オンセット行列：バイナリクロスエントロピー
- ベロシティー：平均二乗誤差

提案法 2 声部推定

コスト関数に基づく系列最適化 (動的計画法)

- 演奏上読みやすい楽譜には正しい声部分離が不可欠
- 声部配置を満たすべき制約をコスト関数として定義
 - 同じ声部ラベルを持つ継続音と非継続音にコストを与える
 - 同じ声部で時間が一時的にオーバーラップする音符にコストを与える etc.
- 最適な声部系列をビタビアルゴリズムで推定



結果

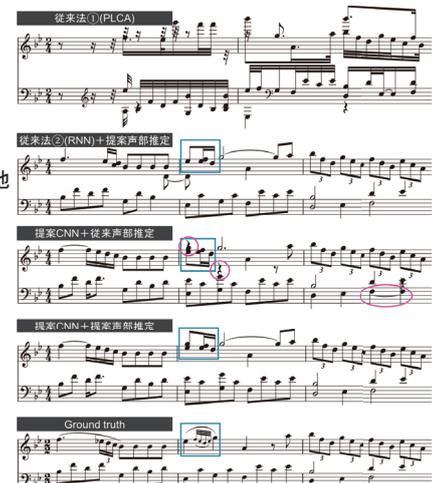
定量評価

- 楽譜推定精度を評価
 - 編集距離に基づく誤り率と多角的な F 値を算出
- 従来法 (PLCA) から提案手法 (DNN 音高検出) により全体的に大幅に精度向上
 - 音高誤り、余分音符、不足音符が激減
 - 声部分離法により声部推定精度が改善

手法	編集距離に基づく誤り率 (低い方が良い)						多層的な尺度 MV2H (F値) (高い方が良い)					
	音高	リズム	オンセット	オフセット	平均	平均	音高	声部	拍節	音強	キー	平均
従来多重音検出 + 提案声部分離	2.92	28.1	11.4	19.4	41.8	20.7	68.2	63.3	31.1	86.7	57.2	61.3
提案多重音検出 + 従来声部分離	0.74	8.19	5.83	12.9	28.6	11.3	83.9	65.2	43.8	91.7	71.1	71.1
提案声部分離 + 提案声部分離	0.74	8.11	5.75	12.8	27.9	11.0	84.3	67.5	43.6	91.4	71.1	71.6

ピアノ採譜のこれから

- 現状
 - 音高・リズムは既に高精度
 - 見にくい休符・タイが減少
 - 音価、拍節構造に改善の余地
- 残された課題
 - 装飾音、演奏表現等の推定
 - トリル、アルペジオ
 - 前打音、後打音
 - スラー、ペダルイベント



フォトリアルなキャラクタのダンスおよび歌唱動作生成システム

岩本尚也 加藤卓哉*1 夏目亮太 山口周悟 岩瀬翔平 島村遼 森島繁生
早稲田大学 *1ExaWizards Inc.

システム概要

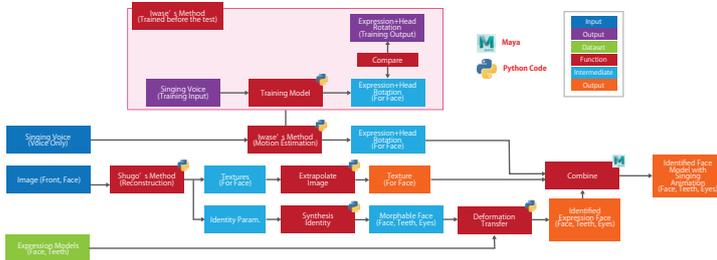
音楽動画創作支援を目的として、画像1枚のみから着衣全身および顔面の3次元モデリングを行い、音楽にシンクロしてダンスおよび歌唱アニメーション動画を自動生成するシステムの開発を行った。歌唱は音楽解析、ダンスは動画解析に基づいて生成される。また、顔および全身のモデルを統合した半自動キャラクタ生成パイプラインを実現した。



歌唱キャラクタ自動生成フロー

1枚の任意方向から撮影された顔画像を入力すると、自動的に音楽にシンクロした歌唱動画が出力される。なお、顔と歌唱生成の2つの独自の研究成果に基く。

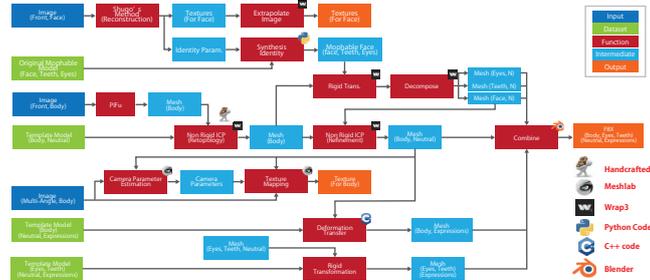
Singing Character Synthesis Pipeline



キャラクタ生成パイプライン

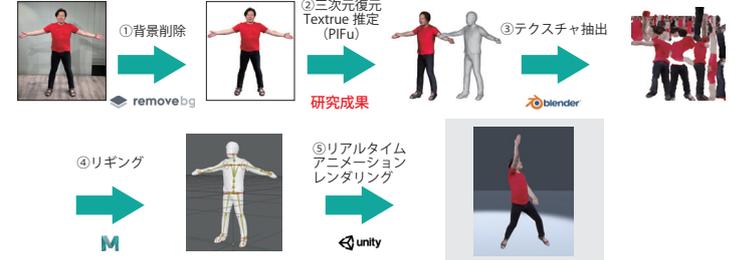
正面画像を含む複数枚 (3枚) の全身画像から、フォトリアルキャラクタを生成するためのパイプラインを示す。全身および顔の後処理方法を改良し、これらを統合してより実写性を高めたキャラクタを生成することが可能となった。ただし、現時点で1か所のみランドマークを手動で指定する必要がある (Retopology 処理)。

Full Body Model Synthesis Pipeline



ダンスキャラクタ自動生成フロー

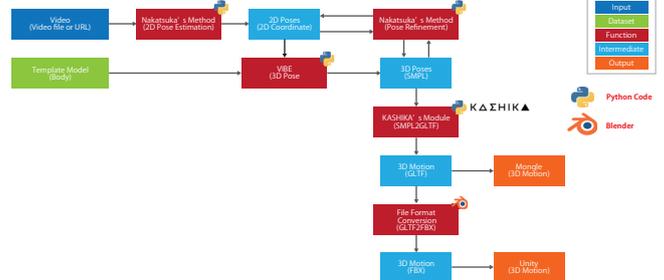
1枚の着衣全身画像を入力すると、自動的に音楽にシンクロしたダンス動画が出力される。なお②PIFuに基づいて生成されたキャラクタは、Unity上でリアルタイムCG合成される。



ダンスポーズ推定システム

ダンス動画を入力として、3次元ポーズを推定する。ポーズは GLTF および FBX で出力される。

Dance Pose Estimation Pipeline



まとめ・今後の課題

1枚のみの画像から歌唱やダンス動作可能なフォトリアルキャラクタ生成技術を開発し、システムとして実装した。通常は長い時間と膨大な人的コストを必要とするキャラクタ生成プロセスを自動化することを可能とした。今後も Uncanny Valley を脱するための研究開発を進め、省力化に大幅に貢献しつつ、プロダクションレベルに到達できる技術開発を推進していく。

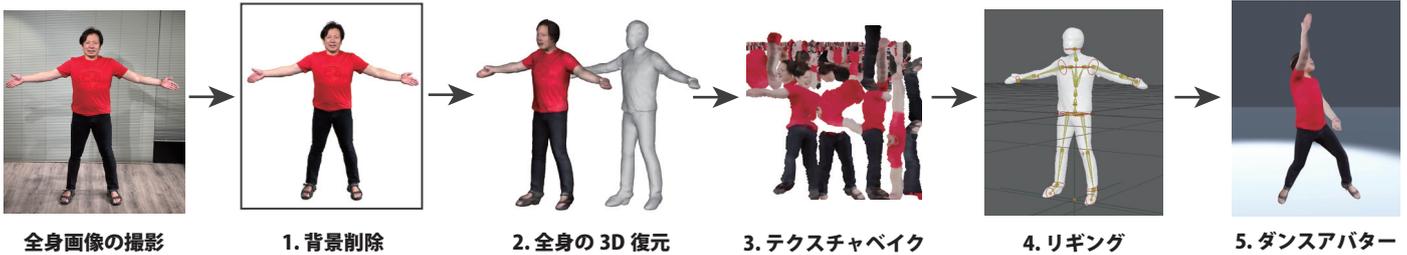
ダンス動画創作支援：1枚画像からの 着衣全身3次元モデリング技術

夏目 亮太 齊藤 隼介*1 森島 繁生

早稲田大学 *1 南カリフォルニア大学

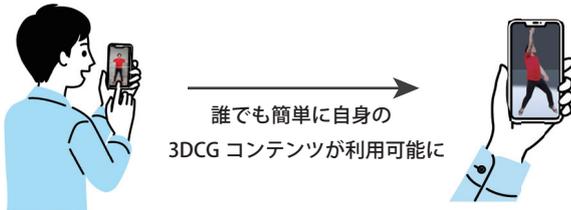
システム概要

全身画像1枚の入力から本人そっくりの“踊るアバター”を作成できるシステム



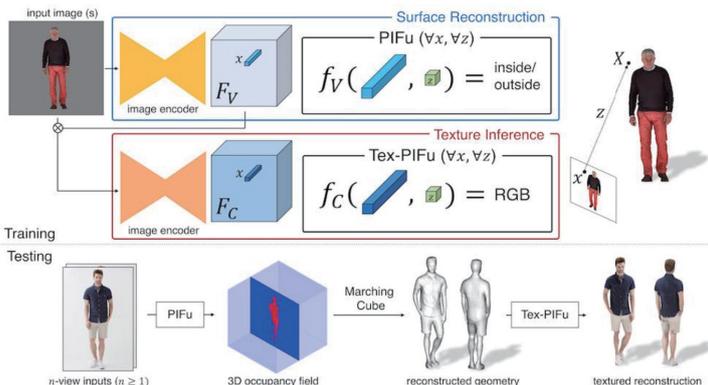
全身の3D 復元

近年、拡張現実や仮想現実などの没入型の映像体験が流行している。それらの映像に必要な3DCGコンテンツを制作するためには、現実世界の3次元物体をデジタル化する必要がある。しかし、着衣人物の3DCG化の場合は、何十台ものカメラを用いた緻密なスキャン装置や、非常に複雑な長時間の撮影作業が必要となる。そこで我々は、着衣人物の写真一枚から、深層学習を用いて全身の3D形状を復元するシステムを実現することで、誰でも気軽に自身の3DCGコンテンツを用いた映像体験が可能となるシステムを開発した。

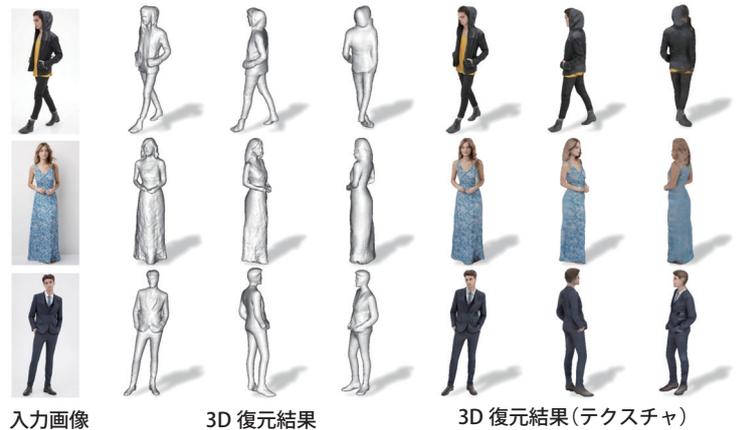


提案手法

- Pixel-Aligned Implicit Function (PIFu) という新しい3D情報の表現方法を提案し効率良く3D情報を扱うことにより、人の全身の高精度な3D復元を実現
- 提案手法では、3D形状だけでなく、画像中には見えない背面のテクスチャも、高精度での復元を実現



3D 復元結果



ダンスアバターの作成

PIFuで3D復元した全身のオブジェクトに対し、自動でリギングを行うことにより、任意のモーションを付与した映像を作成可能



今後の展望

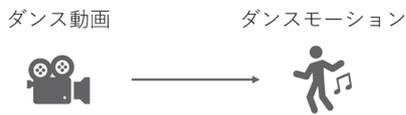
実サービスへの応用を考えると、より高解像度な3D形状とテクスチャの復元が求められる。また、今後の課題としては、動画中の人物に対して、時系列的に安定した復元の実現が考えられる。

ダンス動画創作支援：動画からの3次元ダンスモーション自動推定技術

中塚 貴之 吉井 和佳*¹ 小山 裕己*² 深山 覚*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

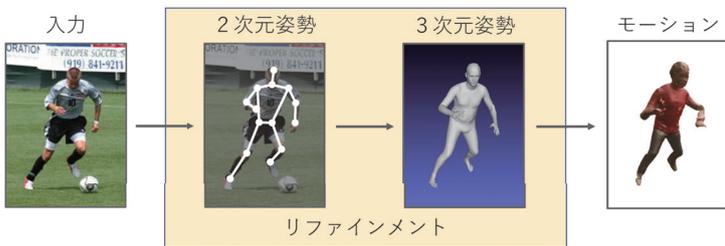
はじめに

バーチャルキャラクタを用いたダンス動画コンテンツは、「踊ってみた動画」等て人気のコンテンツの一つである。近年では、バーチャルユーチューバーやバーチャルアイドルといったように、活躍の場をインターネットに求めるクリエイターが増えている。バーチャルライブでもダンスパフォーマンスのキャプチャシステムが重要だが、ダンスモーションデータの多くは、専用の装置での撮影、またはクリエイターの手作業によって作成されており、コストが高い。我々は単眼カメラの動画から、バーチャルキャラクタに適用可能な3次元ダンスモーションを取得するシステムを実現した。

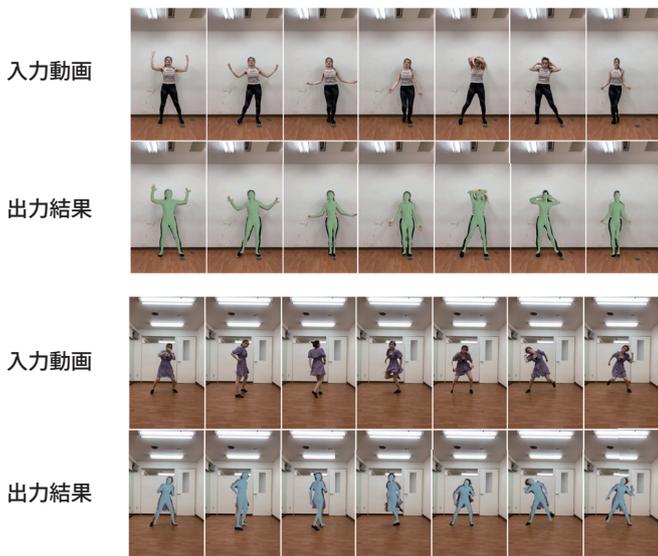


ワークフロー

- 動画ファイル、もしくはニコニコ動画やYouTubeといった動画共有サービスの動画URLを入力とし、3次元ダンスモーションを出力
- 対象人物の2次元姿勢推定結果と3次元姿勢推定結果をリファインメントし、ダンスモーションとして破綻の少ないモーションを推定



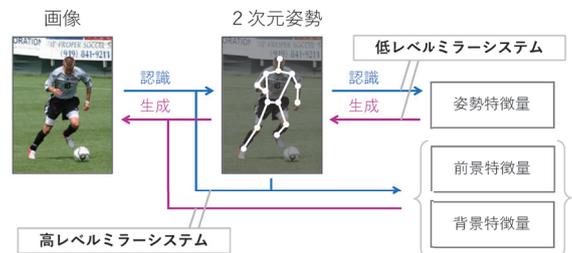
モーション推定結果



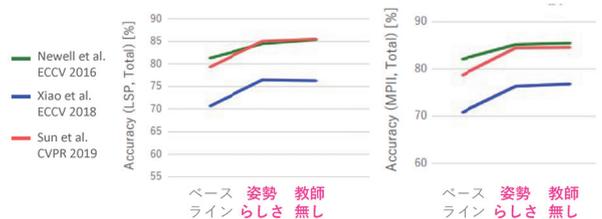
提案手法

2次元姿勢推定

- 画像の認識 - 生成モデル統合による半教師有り2次元姿勢推定
- 生成モデルを導入するメリット
 - 認識モデル、生成モデルの双方から2次元姿勢推定結果の姿勢らしさを多面的に評価することで認識モデルの推定性能を改善
 - 教師無し画像を学習データとして活用



- ベースライン手法に対する姿勢推定性能の改善
 - 生成モデルを導入し、姿勢らしさを考慮することで性能を改善
 - 教師無し画像を学習データに加えることでさらに性能を改善



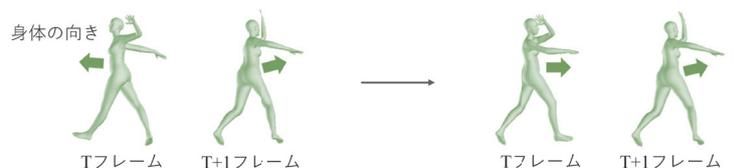
実験条件：ベンチマークデータセットの60%を教師有り、40%を教師無し画像として利用

3次元姿勢推定 [Kocabas et al., 2020]

- ゲート付き回帰型ユニットと敵対的生成ネットワークを組み合わせた手法

姿勢のリファインメント

- 姿勢推定結果に対し関節の位置や身体の向きの変動量が非常に大きいフレームを特定し、前後の推定結果を用いて滑らかに補間



今後の展望

- 2次元姿勢推定で提案した認識 - 生成モデルを統合する枠組みを3次元姿勢推定および時間方向に拡張することで、より高品質な3次元ダンスモーションの推定
- 障害物による遮蔽や自己遮蔽に対しても頑健なモーション推定

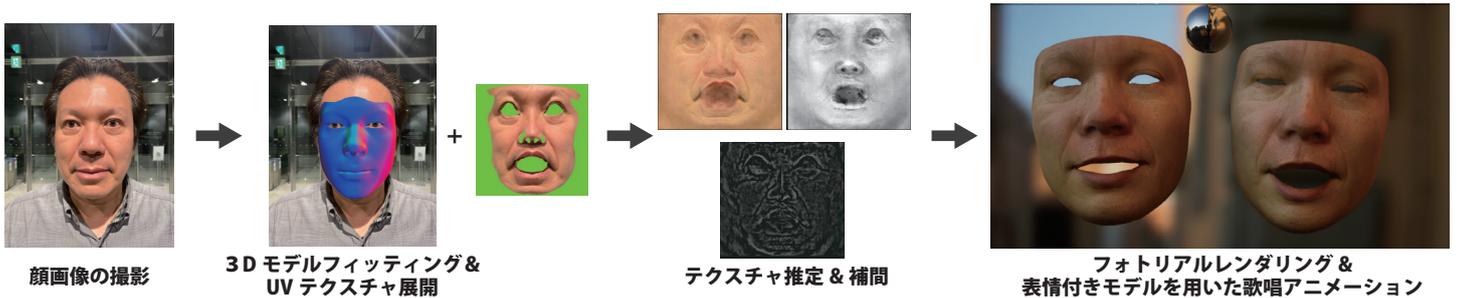
歌唱動画創作支援：1枚画像からの顔の3次元モデリングおよび表情合成技術

山口周悟 齊藤隼介*1 森島繁生

早稲田大学 *1 南カリフォルニア大学

システム概要

顔画像1枚の入力からフォトリアルな顔モデルを作成できるシステム

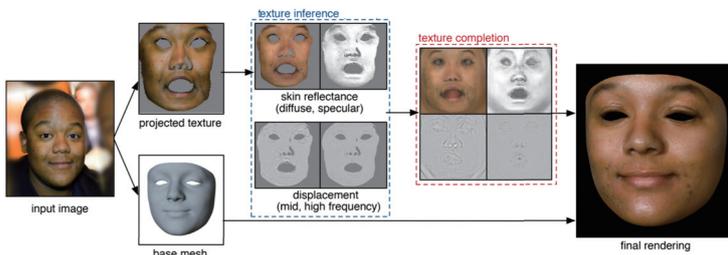


提案手法

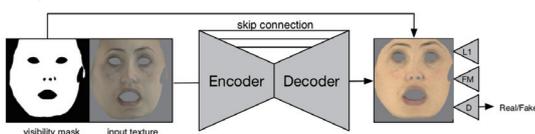
近年、特に映画のプロダクションでは不気味の谷を超えるデジタルアバターが活用され、視聴者がCGの存在にすら気づかないレベルの映像が表現されている。しかし、このようなデジタルアバターを作成するためにはLight Stageのような三次元撮影装置を利用する必要があり、個人が自分のデジタルアバターを作成するのは費用等の点から困難である。

そこで、一枚の顔写真から高画質の皮膚表面反射率と形状を推定するディープラーニングベースの手法を提案する。画像にはオクルージョンや任意の照明条件が含まれていても対応可能である。再構成された高解像度テクスチャは、Diffuse (拡散反射) アルベドと Specular (鏡面反射) アルベドの高解像度の皮膚表面反射率マップ、中・高周波 Displacement マップを含む。これにより、新しい照明条件の下でフォトリアルなデジタルアバターをレンダリングすることが可能である。

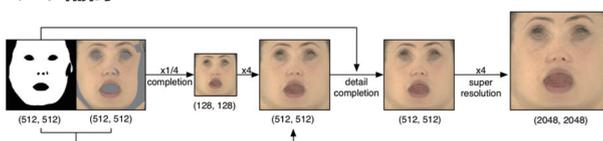
ワークフロー



テクスチャ推定



テクスチャ補間



テクスチャ推定結果



照明・方向に対する頑健性



先行研究との比較



method	PSNR	RSME
[Thies et al. 2016a]	17.6354	0.1369
[Saito et al. 2017]	15.6308	0.1767
[Mohammed et al. 2009]	18.34	0.1271
ours	19.333	0.1102

歌唱動画創作支援：音楽からの個性ある歌唱アニメーション生成技術

岩瀬 翔平 加藤 卓哉*1 森島 繁生

早稲田大学 *1 ExaWizards Inc.

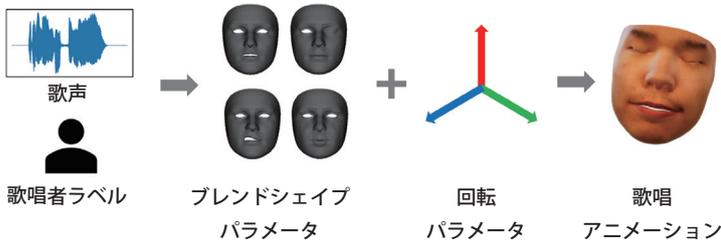
はじめに

歌唱時表情アニメーションの制作

3DCG コンテンツの一つとして歌唱アニメーションがある。音楽に合わせて踊るダンスアニメーションの制作は盛んに行われる一方で、歌唱時の表情アニメーションの制作は難易度が高く多大な時間を要するため、表情は固定されてしまうことが多い。また、写実的な3DCGモデルでは、人間的に自然な動作がより高い水準で求められると同時にキャラクタとなる歌唱者の個性を保持することが要求されるため、制作における技術的障壁がさらに高いと言える。そこで我々は歌声データから直接、自然かつ個性を反映した表情アニメーションを生成可能なシステムを実現した。

システム概要

- 本手法では深層学習モデルを用いて歌声から、表情を制御するブレンドシェイプパラメータと顔回転のパラメータを推定する。
- 学習データセット中の各歌唱者に対応する歌唱者ラベルによって出力アニメーションの個性の制御が可能である。
- これらパラメータを用いて高精細な3次元顔モデルのアニメーションが制作可能である。

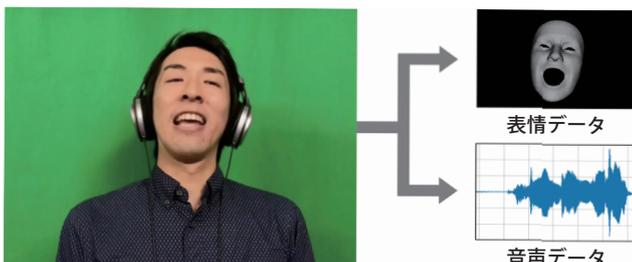


歌唱データセット

データセット構成

男性歌唱者5名 x J-POP 楽曲5曲

- ブレンドシェイプによって表される表情データ
- 表情データと同期された音声



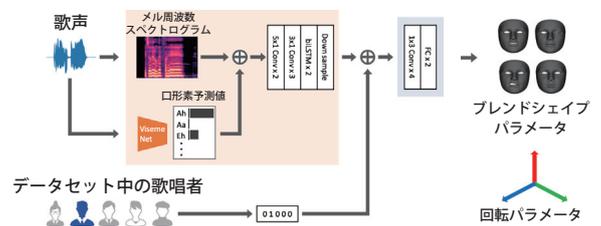
提案手法

生成ネットワーク

表情を音声的要因と個性的要因に分解する

例) 口形→音声的要因、 顔の開閉→個性的要因

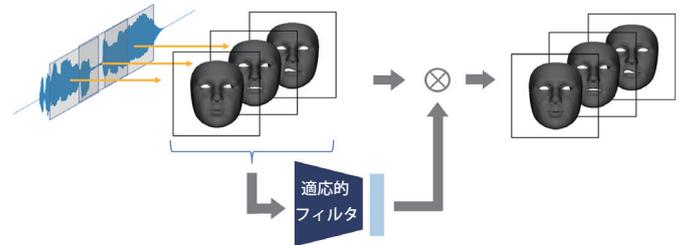
- 歌声から音声特徴量を抽出し、歌唱ラベルより歌唱者情報を与える
- 歌唱者ラベルを入れ替えることで出力アニメーションの個性が変更可能



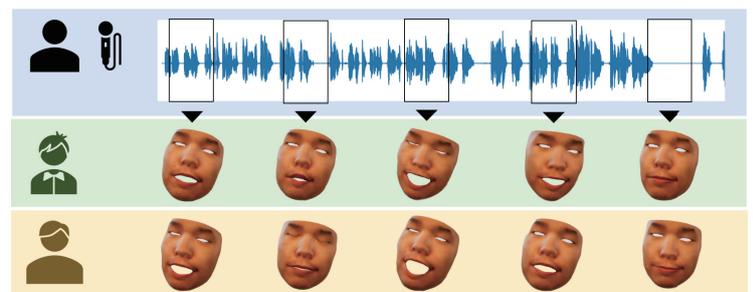
時間安定性

フレーム単位で推定されたパラメータは時間的に不安定

- 適応的にフィルタを生成・適応することで時間方向に平滑化を行う
- 重みを求める適応的フィルタネットワークは生成ネットワークと同時に学習を通して獲得する



生成結果



- 歌声と一致した自然な動作の生成が可能
- 歌唱者ラベルの入れ替えにより顔の開閉のタイミングや顔の傾きが変化
- 極端に大きな首振り動作は再現できていないため今後の課題となる

音楽理解技術に基づくリズムを利用した体験型コンテンツの創出

横山 幸大 伊藤 理紗 當麻 僚太郎 中村 聡史 後藤 真孝*1 井上 隆広*1
 明治大学 産業技術総合研究所*1

問題意識と目的

音楽は様々な可能性を秘めているが、その体験としてはまだ十分に力を発揮できていない。

音楽理解技術を活用し、その解析結果に基づき音楽体験（ミュージックエクスペリエンス：MX）を拡張！

ミュージックエクスペリエンスを高めるサービスやアプリケーションの実現！

具体的な方法

- 好きな音楽を利用して高齢者の運動を支援
- 好きな音楽をゲームとして楽しみ社会貢献
- その場で流れる音楽を認識して新たな体験へ
- 音楽との偶然の出会いを利用したコミュニケーション

高齢者向け音楽体験システム

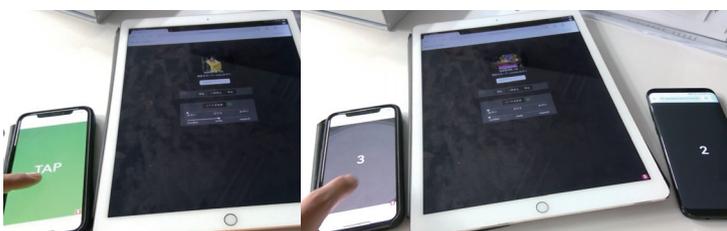
踏んでリズムを楽しむ音楽ゲーム

- 高齢者が懐かしむことができる音楽を利用し、Songle による音楽の解析結果をもとに、リズムに合わせて踏んで楽しむゲームを生成
- 踏むときに利用するタイルの数や、難易度に応じて、タイルに割り当てられる踏むタイミングが動的に変化
- 運動を促すことが目的のため、無理なく楽しめる様々なモードを搭載



スマートデバイスを利用した複数人で楽しむリズムゲーム

- 高齢者が懐かしむことができる音楽を利用し、その音楽の Songle による認識結果をもとに、リズムに合わせてみんなで触って楽しむゲームを生成
- Songle Sync の仕組みにより、たくさんのデバイスが完全に同期し、それぞれのタイミングでタップすることで音楽をすすめることが可能
- 高価なデバイスを使わずとも、施設のひとがもつスマートデバイスだけで実施することが可能



JusTune

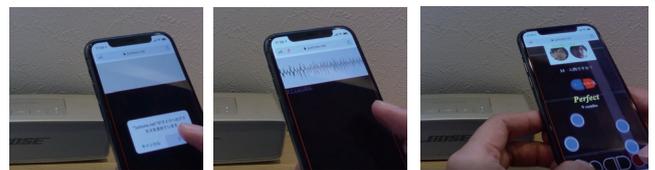
好きな音楽で新たな体験を (<https://justune.net>)

- 好きな音楽について、Songle の音楽解析結果を利用し、そのリズムやサビなどの認識結果に基づき音楽ゲームを自動生成
- 音楽ゲームとして楽しむだけでなく、マイクロタスクをゲーム内に取り込むことによって、ゲームを楽しみながら社会貢献可能



その場に流れる音楽で新たな体験を (<https://justune.net/live>)

- ラジオやテレビ、街なかのスピーカーなどから流れてくる音楽が何なのか、そして音楽の何ミリ秒目が再生されているのかを Songle API により認識し、Songle の解析結果を利用してリアルタイムに音楽ゲームを楽しめるように！
- マイクロタスクについても同じく貢献可能！



その場に流れる音楽で体験共有 (<https://justune.net/sync>)

- ライブや花火大会の会場など、多くのひとが参加する環境において、ある音楽があった時、その音楽を Songle API により認識して、Songle の音楽解析結果を利用し音楽ゲームを自動生成。また Songle Sync の仕組みを利用することで、周りのみんなと一緒に音楽をゲームとして楽しむことが可能に！



離れたひとと同じ音楽で体験共有 (<https://justune.net/comm>)

- ラジオやテレビ、ライブ配信など、同じ音楽をまったく同じタイミングで聴いているひとを Songle API の認識から判断し、今流れている音楽について、他者を感じることを可能に！

その他応用

自身にのみ聞こえる音楽によるプレゼンテーション支援

- 音楽を聴きながらプレゼンすることで、プレゼンの質を向上
- 緊張しない・時間を守りやすい・音量制御が可能・テンポ制御が可能など

自身にのみ聞こえる音楽によるコミュニケーション支援

- 緊張してしまうコミュニケーションを、音楽により改善
- 間を取りやすいなどのポジティブな効果

音楽を聴く行為に着目した 音楽キュレーション技術

野中 滉介 松山 直人 高橋 拓 齊藤 絢基
新納 真次郎 田村 柁優紀 白鳥 裕士 中村 聡史
明治大学

問題意識と目的

日々膨大な音楽が制作されていくため、有名な音楽への再生が集まり、せっかくの素晴らしい音楽が埋もれてしまう。

色々な観点で音楽をキュレーションすることにより、音楽を再発見できるようにするとともに、音楽の世界をより豊かにする！

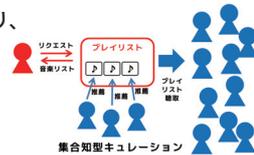
具体的な方法

- ひとの推薦してほしいというニーズを満たすサービスの実現
- 他者により深く興味を持ってもらうための推薦手法とサービスの実現
- グループで音楽を楽しむことを支援する手法の実現
- まだメジャーではないアーティストを応援するサービスの実現

mureq(ミュージック)

質問応答型音楽キュレーションサービス (https://mureq.com)

- マニアックなコンテキストなどを含む質問に様々なユーザが回答して集合知として音楽が集まり、音楽プレイリストを完成させるサービス



クエリの例

- 朝に外を走りたくなるような曲をください
- カラオケに入ってるマイナーなボカロ曲で、歌いやすい盛り上がる曲覚えたい



- 質問に対して音楽が集まることで、結果的に音楽に対するアノテーションとなり、検索のインデックスとして利用可能に

otoniwa(おとにわ)

音楽を庭に植え、育て、聴いてもらうサービス

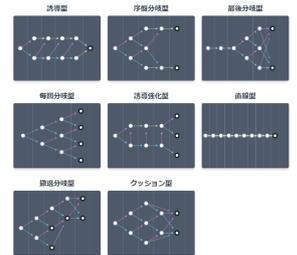
- 自身の好きなアーティストの音楽を聴いてもらうため庭に植える
- 庭で音楽が再生されると、ポイントがもらえる(動画の視聴数も向上)
- 再生数の伸び率によりポイントがもらえるため、音楽を発掘してすすめることを促進



reco.mu(レコミュ)

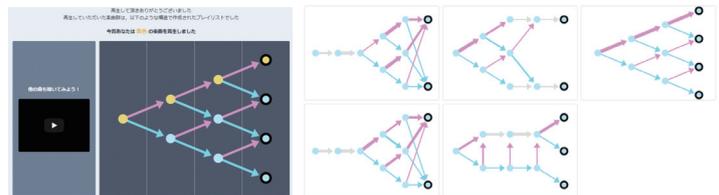
分岐型音楽キュレーションサービス (https://reco.mu)

- 他者に音楽を薦めるとき、その他者が音楽をどう興味を持つかということに着目した音楽推薦サービス
- この曲が好きならこの曲、この曲がいまいちならこの曲といった感じで推薦を行う
- ユーザの聴き方に応じて音楽が分岐
- 分岐型は後半の満足度が向上



ユーザの聴き方から再キュレーション

- 他者がどう音楽を楽しんだかを記録し、分岐の改善が可能
- 他者がどう音楽をすすめているかを参考に、プレイリストを Fork して改良



Uniotto(ユニオット)

グループ型音楽鑑賞アプリ (https://uniotto.org)

- 音楽をひとりで楽しむのではなく、お互いに音楽をすすめあうことを可能にするグループ型音楽鑑賞支援アプリケーション(iOS用)
- 音楽をグループで楽しむための様々なミニアプリケーションの提案

リレー音楽プレイヤー

- 参加者がリレー形式で音楽を選びその音楽の連なりなどをうまく利用しグループの会話を盛り上げるプレイヤー



音札

- 参加者のもつ音楽をカルタの手札としてその音楽が再生されてからいかに早くそのカルタを取ることができるかを競いまたその音楽をすすめあうゲーム

ミュージック人狼

- 参加者がそれぞれ音楽を選択し再生されている音楽が誰によるものかを再生されているひとは自分じゃないと演技しそうでないひとは推理する人狼ゲーム



文字の数式化と融合による 未来の歌詞表現の実現

齊藤 絢基 新納 真次郎 佐藤 剣太 野中 滉介 佐々木 美香子 鈴木 正明 中村 聡史
明治大学

問題意識と目的

ミュージックビデオで提示される歌詞や、カラオケで提示される歌詞など、文字による表現は様々な場面に存在しているが画一的なことが多い。

こうした文字表現を、音楽に合わせてより豊かな表現をできるようにし、音楽の印象を増幅可能とする。

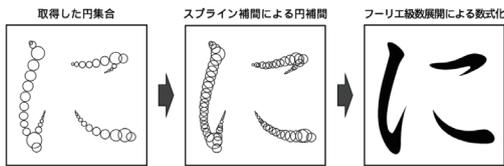
具体的な方法

- 文字を数式として表現し、演算可能に
- 演算により、状況に応じた新たな文字表現を生成

文字の数式化と文字表現の生成

フーリエ級数展開による文字の数式化

- 文字を芯線の座標と、その太さで表現し、フーリエ級数展開により数式化



オリジナルの円集合
 $(x, y, r) = \{(x_1, y_1, r_1), (x_2, y_2, r_2), \dots, (x_n, y_n, r_n)\}$

ストロークの数式
 $X_1 = f_1(t), Y_1 = g_1(t), R_1 = h_1(t)$
 $X_2 = f_2(t), Y_2 = g_2(t), R_2 = h_2(t)$
 $X_3 = f_3(t), Y_3 = g_3(t), R_3 = h_3(t)$

スプラインによって補間された円集合
 $(x', y', r') = \{(x_{1,1}, y_{1,1}, r_{1,1}), \dots, (x_{1,m}, y_{1,m}, r_{1,m}), (x_2, y_2, r_2), \dots, (x_n, y_n, r_n)\}$

数式化された文字の加重平均化

- 数式化された複数の文字表現を任意の割合で融合し、新しい文字表現を生成



文字表現の印象を形容詞対の値で表現し、任意の点で融合

- 18種のフォントについて、35対の形容詞対を用意し、17人に印象評価依頼
- フォントをベクトル表現可能とするとともに、因子分析により4軸を抽出

- 形容詞対の選択により表現する平面を選定し、任意の点の指定により新しい文字表現を生成可能に



印象に基づく歌詞生成

音楽印象に基づくフォント生成

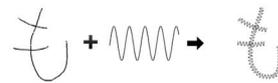
- 音楽から推定された印象に基づき、歌詞に文字表現を適用し、その文字表現に基づき印象の拡張を行う

C3(sadness), C6(cuteness) 私たちがあいに愛し合って 私たちがあいに愛し合って 私たちがあいに愛し合って	C2(vigorousness), C6(cuteness) あふれぬほどにまた会えよう あふれぬほどにまた会えよう あふれぬほどにまた会えよう
C1(grandness), C4(violence) 野放しに突進し 野放しに突進し 野放しに突進し	C2(vigorousness), C5(funniness) この世界わりの日には この世界わりの日には この世界わりの日には
C2(vigorousness), C6(cuteness) あふれぬほどにまた会えよう あふれぬほどにまた会えよう あふれぬほどにまた会えよう	C2(vigorousness), C4(violence) あふれぬほどにまた会えよう あふれぬほどにまた会えよう あふれぬほどにまた会えよう

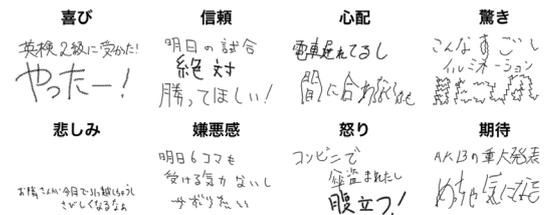
- 音楽の印象を利用して、歌詞のフォントを生成し適用した実験を実施し、切ない・滑稽・可愛いについては印象強調可能だが、元気が出る・激しいについては強調が難しく、アニメーションなどを利用する必要あり

数式化された文字への演算によるアニメーション表現

- 芯線に対して正弦波、矩形波などを付与することで、文字アニメーションを生成



- 印象に適切なアニメーションタイプを生成して多様な表現を可能に



応用

手書き・フォントの融合による温かみのある手紙作成

- 手書きとフォントの融合により、送り手にとって恥ずかしくなく、また受け手にとってうれしい手紙が作成可能

コミックのセリフや字幕などへの適用

- コミックのセリフや、映画の字幕など、様々な文字表現が行われる場面において、フォント融合や手書き融合の技術を利用し、好感度の高い文字表現や、印象評価値を上げる文字表現などが可能に

融合インターフェース

- フォントを絵の具のように融合可能なユーザインターフェースの実現



ダンス理解のための基礎的研究と 応用のためのデータベース構築

又吉康綱 古市冨佳 斎藤光 阿部和樹 中村聡史

小山裕己 *1 深山覚 *1 後藤真孝 *1 森島繁生 *2

明治大学 産業技術総合研究所 *1 早稲田大学 *2

問題意識と目的

- ・中学校の体育の教育におけるダンスの必修化や、踊ってみた動画の隆盛、テレビ番組などにおける参加を促すダンス映像、ダンスのプロリーグに関するD.LEAGUEの発足など、ダンスに対する注目が高まっている。
- ・独自にダンスを楽しむ場合、そのひとの個性にあったダンスを利用する必要があるうえ、ニュアンスを理解するための仕組みも必要になる。
- ・参考にできるダンス動画は膨大にあるが、キーワード以外での探索はできないため、その検索・探索を可能にするための基盤技術が必要になる。

具体的な方法

- ・ダンスの個性について理解する
- ・ダンスにおいてニュアンスを習得する適切な方法を模索
- ・大規模ダンスデータベースを構築し、修正可能とする

ダンスの個性

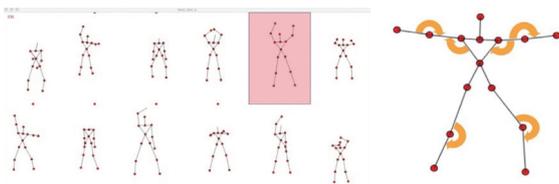
ダンスにおけるそのひとらしさ、個性とは何なのか？

- ・ダンスにおいて自身を判断する材料が個性の一部といえる？
- ダンスの骨格情報だけを抽出したものを提示し、自身がどのものかを判断してもらったところ、熟達したダンサーは自身のダンスを高精度に判断できていた

Participant	U's dance	M's dance	N's dance	O's dance	P's dance	Q's dance	R's dance	S's dance	T's dance	U's dance	average	dance
L	0.95	0.70	0.60	0.60	0.05	0.10	1.10	0.45	1.05	1.65	0.00	0.35
M	0.40	0.85	0.40	0.10	0.30	1.10	1.00	0.15	0.20	0.45	0.35	0.30
N	1.10	0.85	0.75	0.75	0.45	0.20	1.20	0.00	0.45	0.30	0.45	1.15
O	0.40	1.20	0.75	0.75	0.45	1.15	2.05	0.00	0.25	0.00	0.30	0.30
P	0.05	0.55	0.15	0.10	0.70	0.10	0.70	1.30	0.40	0.15	0.30	0.60
Q	0.20	0.60	0.00	0.75	1.90	0.75	0.15	0.20	0.10	0.00	1.55	0.15
R	0.65	1.05	0.40	0.35	1.35	0.40	1.60	0.45	0.65	0.30	0.90	0.00
S	0.45	1.00	1.60	0.50	0.25	0.15	1.55	1.70	1.10	0.15	0.55	0.00
T	0.35	0.65	0.40	0.15	0.30	0.15	0.50	0.10	0.50	0.50	0.55	0.35
U	0.40	1.95	0.25	0.15	0.15	0.15	0.60	0.05	0.20	0.65	0.40	0.05
V	0.95	1.05	0.80	0.70	0.50	0.85	1.35	0.30	0.20	0.55	1.20	0.00

ダンスにおけるそのひとらしさを機械が判断可能か？

- ・ダンスにおいてひとはどこで自分を判断しており、機械はそれを判定可能か？
- 結果：経験豊富なダンサー→経験が浅いダンサー
- 経験豊富なダンサーは、動きではなく、体の形を判別に利用していた



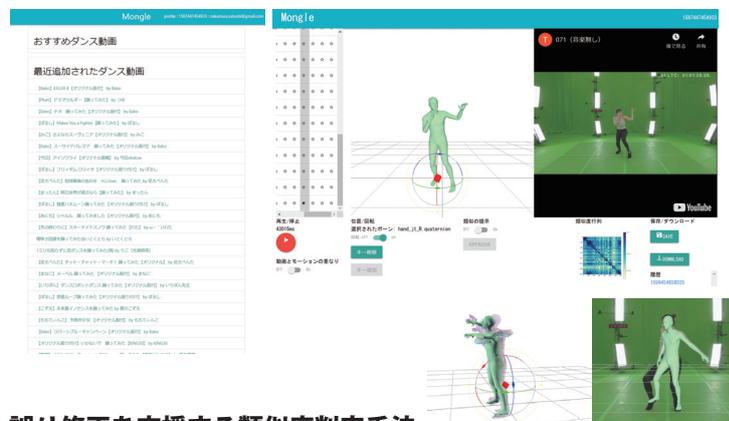
・角度特徴を使うことで機械も高精度に判定可能

ダンス経験が浅いグループ 名家経験力乏のダンス														ダンス経験が豊富なグループ 名家経験力乏のダンス													
A	B	C	D	E	F	G	H	I	J	K	L	M	N	A	B	C	D	E	F	G	H	I	J	K	L	M	N
0.25	0.07	0.10	0.05	0.09	0.08	0.07	0.05	0.09	0.05	0.06	0.05	0.04	0.06	0.26	0.04	0.14	0.09	0.05	0.04	0.08	0.09	0.08	0.05	0.03	0.03	0.07	
0.05	0.40	0.06	0.04	0.07	0.07	0.06	0.02	0.06	0.04	0.06	0.08	0.07	0.03	0.03	0.37	0.07	0.04	0.04	0.07	0.06	0.05	0.05	0.06	0.09	0.06		
0.08	0.07	0.03	0.04	0.08	0.07	0.06	0.09	0.04	0.07	0.06	0.08	0.08	0.08	0.04	0.07	0.32	0.04	0.04	0.05	0.09	0.03	0.06	0.10	0.08	0.08		
0.06	0.06	0.04	0.11	0.08	0.08	0.08	0.07	0.06	0.06	0.03	0.06	0.06	0.03	0.14	0.04	0.05	0.24	0.10	0.05	0.05	0.08	0.10	0.08	0.05	0.05		
0.05	0.08	0.04	0.07	0.06	0.08	0.09	0.06	0.05	0.08	0.09	0.05	0.08	0.05	0.08	0.04	0.04	0.09	0.37	0.08	0.03	0.07	0.08	0.07	0.04	0.02		
0.07	0.06	0.06	0.04	0.06	0.24	0.09	0.10	0.04	0.09	0.08	0.06	0.06	0.06	0.05	0.07	0.05	0.05	0.08	0.34	0.03	0.05	0.08	0.09	0.09	0.03		
0.06	0.06	0.06	0.08	0.09	0.30	0.07	0.05	0.07	0.07	0.07	0.07	0.07	0.07	0.04	0.07	0.09	0.04	0.03	0.05	0.24	0.06	0.06	0.06	0.09	0.15		
0.05	0.07	0.06	0.05	0.10	0.07	0.06	0.06	0.08	0.08	0.09	0.09	0.04	0.03	0.09	0.07	0.03	0.09	0.06	0.05	0.08	0.10	0.08	0.06	0.03			
0.05	0.03	0.07	0.07	0.05	0.07	0.07	0.28	0.11	0.09	0.04	0.06	0.06	0.05	0.08	0.05	0.06	0.09	0.07	0.05	0.11	0.14	0.10	0.08	0.05			
0.06	0.05	0.08	0.04	0.06	0.09	0.07	0.24	0.08	0.06	0.06	0.06	0.06	0.06	0.08	0.05	0.08	0.07	0.08	0.07	0.07	0.09	0.20	0.08	0.06			
0.03	0.03	0.05	0.04	0.07	0.08	0.07	0.05	0.38	0.06	0.06	0.06	0.06	0.06	0.05	0.09	0.08	0.04	0.04	0.08	0.07	0.06	0.08	0.09	0.25	0.08		
0.04	0.06	0.06	0.02	0.05	0.05	0.07	0.10	0.03	0.07	0.08	0.06	0.06	0.06	0.03	0.08	0.10	0.03	0.03	0.04	0.16	0.05	0.06	0.08	0.10	0.25		

Mongle(モングル)

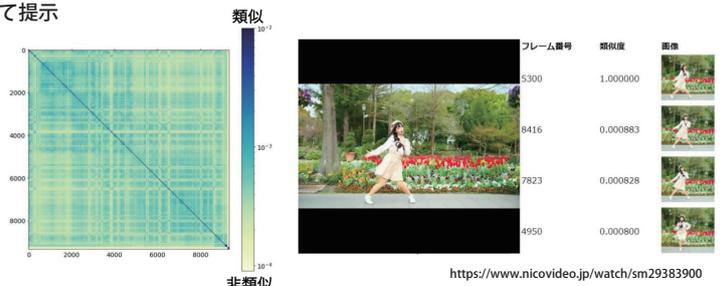
ダンスモーション誤り修正インターフェイス (https://mongle.jp)

- ・森島研究室のダンス動画からの骨格認識手法をもとに、ダンス動画を骨格情報とともに情報を蓄積
- ・ウェブインターフェイス上で誤り修正可能とする



誤り修正を支援する類似度判定手法

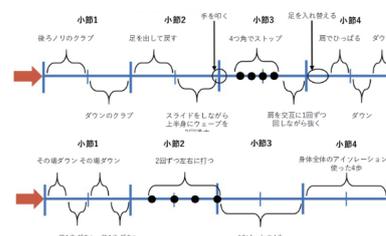
- ・ダンス動画内における類似シーンを画像ベースで発見し、そのシーンにおける骨格推定結果は類似していると判断し、その骨格情報を誤り修正のための候補として提示



ダンスの習得容易性向上

ダンスにおいて難しいニュアンスを動画からどう習得するか？

- ・ダンスにおいて習得するのが難しいニュアンスをいかにして動画だけを視聴して習得を目指す
- 着目した音楽推薦サービス
- ・漫画的オノマトペ提示手法が最も効果的！
- ・ダンスの構造表現にも取り組む



Kiite: 楽曲と出会える音楽発掘サービス

石田 啓介 佐藤 浩輔*2 濱崎 雅弘 後藤 真孝 佃 洸撰 佐々木 渉*1
 高橋 卓見 渡邊 研斗 深山 覚 中村 聡史*2

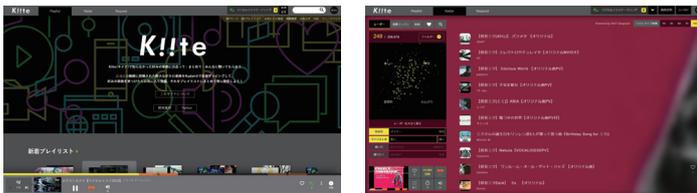
産業技術総合研究所 *1 クリプトン・フューチャー・メディア株式会社 *2 明治大学

Kiite(キイテ)とは

音楽発掘サービス Kiite (https://kiite.jp)

- 膨大な歌声合成楽曲の中を探索し、好みの楽曲に出会えるサービス
 - 音楽印象分析・音楽推薦技術を用いて、多様な探索・絞り込み・推薦を実現
 - ユーザが複数の推薦エンジンを作成でき、他のユーザのエンジンも利用可能
- 自分のプレイリストを作成して、自分のお気に入りの音楽を広めたり、他のユーザが作ったプレイリストを視聴することが可能

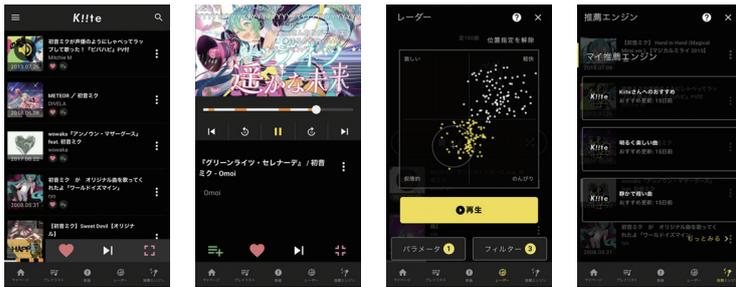
自分の「好き」を見つけて
 自分の「好き」を広める音楽発掘サービス **Kiite**



モバイル版 Kiite

スマートフォン・タブレット用 Kiite

- Kiiteの音楽レーダー機能、音楽推薦エンジン機能、新曲チェック機能などを、スマートフォン等のモバイル端末から手軽に利用可能



Kiite Cafe

「好き」が伝わる音楽発掘カフェ

- そこに居るだけであなたの「好き」がみんなに伝わるバーチャルカフェ
- Kiite Cafeにログインしているユーザのお気に入り曲やプレイリストの曲から、次に流れる曲が自動的に選曲され、みんなで同時に視聴できる
- いま流れている曲が「誰の好き」なのか、今その曲を聴いて「誰が好き」になったのかが伝わる



選曲理由の表示
 現在流れている曲が、誰のプレイリスト登録曲やお気に入り曲であるかが、すぐわかる

ユーザの反応を表示
 現在流れている曲をお気に入り登録すると、画面にハートマークが現れ登録したことが周りに伝わる。さらに吹き出しや回ったりもできる

Kiiteの様々な機能

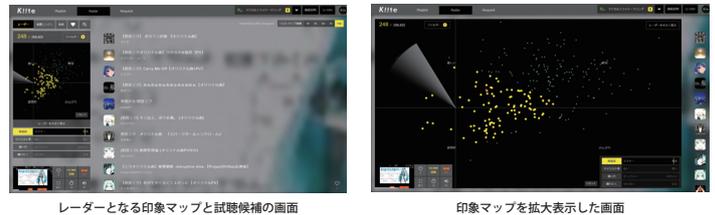
音楽発掘用プレーヤー機能：多数の楽曲を次々と試聴しながらプレイリストを作成・共有可能

- 効率的な試聴を可能にする音楽発掘用プレーヤー：
 - 楽曲のサビ区間を事前に自動検出し、音楽動画をサビから試聴できる
 - 再生中にボタンを一押しで手軽にプレイリストへ追加できる



音楽レーダー機能：音楽印象分析に基づく探索・絞り込みが可能

- 歌声合成の歌手名、投稿日、ジャンル、ソーシャルタグ、再生数などの属性をフィルター条件として組み合わせる絞り込める
- 音楽印象分析技術によって楽曲の音響信号からその印象を自動分析し、その結果も組み合わせながら絞り込める



複数の音楽推薦エンジンを使い分けて新たな楽曲に出会う

- 音楽推薦エンジンがユーザ毎に異なる「お勧め楽曲のプレイリスト」を自動生成
 - 視聴履歴やお気に入り、楽曲の音響信号を自動解析して得られる曲調を利用
- ユーザに自動でパーソナライズ(個人適応)した推薦エンジンに加え、ユーザが自分でカスタマイズ(改造)した「カスタム推薦エンジン」を追加可能
 - 自分の推薦エンジンだけでなく、他のユーザの推薦エンジンも利用可能



ボーカロイド楽曲レーベルKARENTとのコラボレーション企画を2020年8月に実施

Kiite Cafeがクリエイターとの新しい交流の場に

- 特定のクリエイターが作成したプレイリストの楽曲だけが選曲される特別な時間帯を用意し、特別企画として告知
- 当日はクリエイター本人も参加し、100名を超えるユーザが集合



OngaACCEL 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)	2016/08 - 2021/03
濱崎 雅弘	(産業技術総合研究所)	2016/08 - 2021/03
中野 倫靖	(産業技術総合研究所)	2016/08 - 2021/03
深山 覚	(産業技術総合研究所)	2016/08 - 2021/03
加藤 淳	(産業技術総合研究所)	2016/08 - 2021/03
佃 洸撰	(産業技術総合研究所)	2016/08 - 2021/03
小山 裕己	(産業技術総合研究所)	2017/04 - 2021/03
渡邊 研斗	(産業技術総合研究所)	2018/04 - 2021/03
Tian Cheng	(産業技術総合研究所)	2016/12 - 2021/03
土田 修平	(産業技術総合研究所)	2017/04 - 2021/03
Kin Wah Edward Lin	(産業技術総合研究所)	2018/05 - 2021/03
Fabrizio Pedersoli	(産業技術総合研究所)	2019/04 - 2021/03
川崎 裕太	(産業技術総合研究所)	2016/11 - 2021/03

石田 啓介	(産業技術総合研究所)	2016/11 - 2021/03
井上 隆広	(産業技術総合研究所)	2016/11 - 2021/03
田中 一大	(産業技術総合研究所)	2016/11 - 2021/03
笠井 志麻	(産業技術総合研究所)	2019/04 - 2021/03
高橋 卓見	(産業技術総合研究所)	2019/04 - 2021/03
中村 美慧	(産業技術総合研究所)	2019/04 - 2021/03
矢倉 大夢	(産業技術総合研究所)	2019/12 - 2021/03
田中 利枝	(産業技術総合研究所)	2017/04 - 2017/10
Jordan Smith	(産業技術総合研究所)	2016/11 - 2017/12
尾形 正泰	(産業技術総合研究所)	2016/08 - 2018/08
Paul Haimes	(産業技術総合研究所)	2017/04 - 2019/02
中村 裕美	(産業技術総合研究所)	2017/02 - 2019/09

吉井グループ

吉井 和佳	(京都大学)	2016/08 - 2021/03
中村 栄太	(京都大学)	2016/08 - 2021/03
Andrew McLeod	(EPFL)	2018/04 - 2021/03
呉 益明	(京都大学)	2018/10 - 2021/03
錦見 亮	(京都大学)	2016/08 - 2021/03
柴田 剛	(京都大学)	2018/04 - 2021/03
石塚 峻斗	(京都大学)	2019/04 - 2021/03
平松 祐紀	(京都大学)	2019/04 - 2021/03
大山 偉永	(京都大学)	2020/05 - 2021/03
寺尾 萌夢	(京都大学)	2020/05 - 2021/03
Florian Thalmann	(京都大学)	2020/06 - 2021/03

大喜多 美里	(京都大学)	2016/08 - 2017/03
福田 翼	(京都大学)	2016/08 - 2017/03
尾島 優太	(京都大学)	2016/08 - 2018/03
村田 叡	(京都大学)	2018/02 - 2018/03
津島 啓晃	(京都大学)	2016/08 - 2019/03
和田 雄介	(京都大学)	2016/08 - 2019/03
糸山 克寿	(東京工業大学)	2016/08 - 2019/03
柴田 健太郎	(京都大学)	2017/04 - 2020/03
上田 舜	(京都大学)	2017/04 - 2020/03
持橋 大地	(統計数理研究所)	2016/08 - 2020/03

森島グループ

森島 繁生	(早稲田大学)	2016/08 - 2021/03
Savkin Pavel	(早稲田大学)	2016/08 - 2021/03
加藤 久美子	(早稲田大学)	2018/04 - 2021/03
岩本 尚也	(早稲田大学)	2016/09 - 2021/03
加藤 卓哉	(エクサウィザーズ)	2016/09 - 2021/03
斎藤 隼介	(南カリフォルニア大学)	2016/11 - 2021/03
山口 周悟	(早稲田大学)	2016/11 - 2021/03
中塚 貴之	(早稲田大学)	2016/09 - 2021/03
野沢 綸佐	(早稲田大学)	2016/09 - 2021/03
福原 吉博	(早稲田大学)	2016/10 - 2021/03
金田 綾乃	(早稲田大学)	2016/09 - 2021/03
山本 晋太郎	(早稲田大学)	2016/08 - 2021/03
馮 起	(早稲田大学)	2017/04 - 2021/03
粥川 青汰	(早稲田大学)	2017/04 - 2021/03
島村 僚	(早稲田大学)	2018/04 - 2021/03
久家 隆宏	(早稲田大学)	2018/04 - 2021/03
石田 大地	(早稲田大学)	2018/04 - 2021/03
土屋 志高	(早稲田大学)	2018/04 - 2021/03
趙 慕絲	(早稲田大学)	2018/04 - 2021/03
松田 聡子	(早稲田大学)	2019/04 - 2021/03
谷田川 達也	(東京大学)	2019/04 - 2021/03
板摺 貴大	(早稲田大学)	2019/04 - 2021/03
岩瀬 翔平	(早稲田大学)	2019/04 - 2021/03
小池 達也	(早稲田大学)	2019/04 - 2021/03
田中 啓太郎	(早稲田大学)	2019/04 - 2021/03
辻 雄太	(早稲田大学)	2019/04 - 2021/03
中村 和也	(早稲田大学)	2019/04 - 2021/03
大矢 隆	(早稲田大学)	2019/04 - 2021/03
小林 篤史	(早稲田大学)	2019/04 - 2021/03
綱島 秀樹	(早稲田大学)	2020/04 - 2021/03
栗林 雅希	(早稲田大学)	2020/04 - 2021/03
久保谷 善記	(早稲田大学)	2020/04 - 2021/03
沖川 翔太	(早稲田大学)	2020/04 - 2021/03

佐藤 和仁	(早稲田大学)	2020/04 - 2021/03
平田 明日香	(早稲田大学)	2020/04 - 2021/03
辻 雄太	(早稲田大学)	2020/04 - 2021/03
朝比奈 わかな	(早稲田大学)	2016/09 - 2017/03
佐藤 晴紀	(早稲田大学)	2016/09 - 2017/03
赤堀 渉	(早稲田大学)	2016/11 - 2017/03
鍵山 裕貴	(早稲田大学)	2016/11 - 2017/03
成田 史弥	(早稲田大学)	2016/11 - 2017/03
福里 司	(早稲田大学)	2016/11 - 2017/08
杉原 大志	(早稲田大学)	2017/04 - 2017/09
小澤 禎裕	(早稲田大学)	2016/09 - 2018/03
柿塚 亮	(早稲田大学)	2016/08 - 2018/03
佐藤 優伍	(早稲田大学)	2016/11 - 2018/03
古川 翔一	(早稲田大学)	2016/11 - 2018/03
中村 優文	(早稲田大学)	2016/11 - 2018/03
持田 恵佑	(早稲田大学)	2016/08 - 2018/03
佐藤 樹	(早稲田大学)	2016/08 - 2018/03
松田 聡子	(早稲田大学)	2016/08 - 2018/09
山本 祥彰	(早稲田大学)	2018/04 - 2018/09
谷田川 達也	(早稲田大学)	2016/08 - 2019/03
小坂 昂大	(早稲田大学)	2016/11 - 2019/03
板摺 貴大	(早稲田大学)	2016/08 - 2019/03
高森 啓史	(早稲田大学)	2016/08 - 2019/03
宮川 翔貴	(早稲田大学)	2016/09 - 2019/03
又吉 祐	(早稲田大学)	2018/04 - 2019/08
野澤 直樹	(早稲田大学)	2016/11 - 2020/03
井上 和樹	(早稲田大学)	2017/04 - 2020/03
夏目 亮太	(早稲田大学)	2017/04 - 2020/03
山口 智也	(早稲田大学)	2017/04 - 2020/03
山岸 奏実	(早稲田大学)	2017/04 - 2020/03
矢代 達希	(早稲田大学)	2018/04 - 2020/03
岩住 有希子	(早稲田大学)	2019/04 - 2020/03

中村グループ

中村 聡史	(明治大学)	2016/08 - 2021/03
鈴木 正明	(明治大学)	2016/08 - 2021/03
佐藤 浩輔	(明治大学)	2017/01 - 2021/03
高橋 拓	(明治大学)	2017/04 - 2021/03
佐々木 美香子	(明治大学)	2017/04 - 2021/03
又吉 康綱	(明治大学)	2017/04 - 2021/03
菅野 一平	(明治大学)	2018/04 - 2021/03
桑原 樹蘭	(明治大学)	2018/04 - 2021/03
南里 英幸	(明治大学)	2018/04 - 2021/03
野中 滉介	(明治大学)	2018/04 - 2021/03
古市 冴佳	(明治大学)	2018/04 - 2021/03
細谷 美月	(明治大学)	2018/04 - 2021/03
松山 直人	(明治大学)	2018/04 - 2021/03
伊藤 理紗	(明治大学)	2019/04 - 2021/03
船崎 友稀奈	(明治大学)	2019/04 - 2021/03
杉本 知佳	(明治大学)	2019/04 - 2021/03
二宮 洸太	(明治大学)	2019/04 - 2021/03
濱野 花莉	(明治大学)	2019/04 - 2021/03
梶田 美帆	(明治大学)	2019/04 - 2021/03
藤原 優花	(明治大学)	2019/04 - 2021/03
横山 幸大	(明治大学)	2019/04 - 2021/03
新納 真次郎	(明治大学)	2019/11 - 2021/03
青木 由樹乃	(明治大学)	2020/04 - 2021/03
植木 里帆	(明治大学)	2020/04 - 2021/03
山崎 郁未	(明治大学)	2020/04 - 2021/03
飯島 雅也	(明治大学)	2020/04 - 2021/03
清水 亜美	(明治大学)	2020/04 - 2021/03

中村 瞭汰	(明治大学)	2020/04 - 2021/03
山口 真	(明治大学)	2020/07 - 2021/03
福地 翼	(明治大学)	2016/08 - 2018/03
久保田 夏美	(明治大学)	2016/08 - 2018/03
佐藤 剣太	(明治大学)	2016/08 - 2019/03
新納 真次郎	(明治大学)	2016/08 - 2019/03
牧 良樹	(明治大学)	2016/08 - 2019/03
松田 滉平	(明治大学)	2016/08 - 2019/03
松井 啓司	(明治大学)	2016/08 - 2019/03
田村 柁優紀	(明治大学)	2016/08 - 2019/03
大野 直紀	(明治大学)	2016/08 - 2019/03
土屋 駿貴	(明治大学)	2016/08 - 2019/03
前島 紘希	(明治大学)	2016/08 - 2019/03
田島 一樹	(明治大学)	2016/08 - 2019/03
白鳥 裕士	(明治大学)	2016/08 - 2019/03
上西 隆平	(明治大学)	2017/04 - 2019/03
神山 拓史	(明治大学)	2016/08 - 2020/03
斉藤 絢基	(明治大学)	2016/08 - 2020/03
樋川 一幸	(明治大学)	2016/08 - 2020/03
山浦 祐明	(明治大学)	2016/08 - 2020/03
阿部 和樹	(明治大学)	2016/08 - 2020/03
徳久 弘樹	(明治大学)	2018/04 - 2020/03
齋藤 光	(明治大学)	2017/04 - 2020/03
水野 颯	(明治大学)	2017/04 - 2020/03
大島 遼	(明治大学)	2016/08 - 2020/03
木頃 伸哉	(明治大学)	2018/04 - 2020/03

佐々木グループ

佐々木 渉	(クリプトン・フューチャー・メディア株式会社)	2016/08 - 2021/03
林 禎康	(クリプトン・フューチャー・メディア株式会社)	2016/08 - 2021/03
長澤 公平	(クリプトン・フューチャー・メディア株式会社)	2019/11 - 2021/03
田名部 茂	(クリプトン・フューチャー・メディア株式会社)	2016/08 - 2019/11



OngaACCELシンポジウム2020 (<https://ongaaccel.jp/symposium2020/>) 「音楽情報処理研究が切り拓く未来の音楽体験」予稿集

科学技術振興機構 (JST) 戦略的創造研究推進事業 ACCEL
2016 年度採択研究課題
「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」

発行日 2020 年9 月12 日
代表者 後藤 真孝
編集者 中村 美慧 笠井 志麻 加藤 淳
発行 OngaACCEL プロジェクト <info@mail.ongaaccel.jp>
国立研究開発法人 産業技術総合研究所
人間情報インタラクション研究部門 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 029-861-1344