

# デモンストレーション：音声・音楽情報処理の研究紹介 I

中村 栄太<sup>1</sup> 虫鹿 弘二<sup>2</sup> 田中 翔平<sup>3</sup> Junichi Yamagishi<sup>4</sup> 佃 洗撰<sup>5</sup> 加藤 淳<sup>5</sup> 尾形 正泰<sup>5</sup>  
和田 雄介<sup>1</sup> 橋田 光代<sup>6</sup> 濱崎 雅弘<sup>5</sup> 吉久 怜子<sup>7</sup> 錦見 亮<sup>1</sup>

**概要：**音声・音楽情報処理の研究分野のさらなる発展に向けて、同分野に関する最先端の研究を紹介するセッションを、デモンストレーション形式で実施する。本稿では、このセッションで発表予定のデモシステムについて、各発表者による概要の紹介を行う。

## デモセッションの開催によせて

中村 栄太 (京都大学)

本セッションでは、音声・音楽情報処理研究分野における多様かつ最新の研究例を、デモンストレーション形式で紹介する。発表者が自らのシステムを実演、参加者がそれらを体験し、両者が直接ディスカッションを行えることは、一般発表にはない特色である。音声言語情報処理研究会と音楽情報科学研究会の合同研究会である本研究会では、音声情報と音楽情報の両方の分野からデモンストレーション発表が集まっている。

音楽情報科学研究会では、夏の研究発表会に合わせて例年デモセッション開催されているが、

- 新しい形態の楽器
  - 新しい音楽体験のためのインタフェース
- など、実際に触れてみないとわかりにくいものだけでなく、
- 既存の楽器を用い、演奏を伴うもの
  - 開発中のもの (未完成のものも含む)
  - 以前に発表したけど、それ以降デモンストレーションの機会に恵まれていないもの
  - 本セッションと同回に一般発表がなされるもの

などの発表も積極的に受け入れている。また、過去の開催回では一般発表では発表しにくいような、

- 思いつきで作ってしまったもの
- 作ってみたものの、発表されることなくお蔵入りとなってしまったもの

なども少なからず披露されている。そのようなものも発表の機会を得ることで、参加者とのディスカッションを通じて新たな気づきが生じ、これを契機としてのちに日の目をみる研究へと発展する可能性がある。本セッションを継続的に実施することは、本研究分野における多様な研究の発展に資するものと考えている。今後もより多くの研究に発表の機会を提供できるよう努めていく所存である。

音楽情報科学研究会では、ニコニコ生放送を通じて一般発表を動画中継・アーカイブする取り組みを進めてきた。しかし本セッションの各発表者が実際にデモを行う場面は動画中継されないため、発表内容を体験できるのは、発表会場に居合わせた参加者のみである。研究の成果を同じ時間と場所で共有しながらディスカッションを行える体験型の発表形態は、時系列メディアである音楽を扱う研究発表において、一般発表とは異なる側面からの意見を得られる有益な機会となる。また、参加者にとっても、本研究分野で扱われる多様かつ最新の話題に、最も近いところで触れられる機会となりうる。本セッションを通じ、多くの方に、何らかの新たな気づきを得ていただければ幸いである。

## 発表内容の概要

今回のデモセッションでは、全 11 件の発表が予定されている。この内、3 件は音声情報処理に関するもの、8 件は音楽情報処理に関するものであり、各発表者 (敬称略) の発表内容は、概ね以下のように分類される：

**音声合成：** 虫鹿, Yamagishi

**対話：** 田中

**歌詞探索：** 佃

**音楽関連制作支援：** 加藤

**音楽連動制御：** 尾形

**歌声分析・カラオケ：** 和田, 錦見

**演奏表現：** 橋田

<sup>1</sup> 京都大学  
<sup>2</sup> HOYA  
<sup>3</sup> 名古屋工業大学  
<sup>4</sup> National Institute of Informatics  
<sup>5</sup> 産業技術総合研究所  
<sup>6</sup> 相愛大学  
<sup>7</sup> レコチョク

音楽視聴支援： 濱崎

音楽可視化： 吉久

このように、各発表の研究内容は多岐にわたることから、発表会場では多様なディスカッションが期待される。

以降、各発表者による概要の紹介を掲載する。

## VoiceText の紹介

虫鹿 弘二 (HOYA), 小沼 海 (HOYA)

### 概要

VoiceText は様々な業界・企業・用途で1,300社以上に導入実績があります。電話自動応答、ロボット、アプリケーション、放送分野、Web など、多様なシーン利用されています。製品、導入事例、開発中の技術などについてご紹介いたします。

### 関連 Web ページ URL

<http://voicetext.jp/>

## しゃべってらくらく 洲浜商店

田中 翔平 (名古屋工業大学), 寺本 裕成 (愛知工業大学),  
李 晃伸 (名古屋工業大学)

### 概要

洲浜商店は、日々の買い物に不便を感じているお年寄りのために作られた、音声対話アプリケーションです。開発には MMDAgent を利用しています。現在日本では多くのお年寄りが日々の買い物に不便を感じています。このような人々のための買い物支援は徐々に充実してきていますが、赤字に悩む買い物支援事業者も多いです。洲浜商店はそういったお年寄り、事業者両方の悩みを音声対話で解決します。洲浜商店では・お年寄りからの支援依頼・支援業者からの買い物支援情報発信という2つのメイン機能を、キャラクターとの音声対話で利用できます。洲浜商店の機能の一つ、支援依頼を使えば、お年寄りは簡単に買い物を行えます。支援依頼を受け取った買い物代行業者は、内容をもとに依頼品をお届けします。洲浜商店のもう一つの機能として、買い物支援情報の発信機能があります。買い物支援情報とは「駅前に移動販売車が来ています」といった情報です。支援業者が登録した情報がアプリ上で表示されるので、支援業者は自分たちのことを知ってもらう機会が増えます。洲浜商店はお年寄りと事業者をつなぎ、どちらのくらしもスマートにする次世代のコミュニケーションツールです。

### 関連 Web ページ URL

<http://site460713-8694-3651.strikingly.com/>

<https://vimeo.com/239294559>

## Sneak Preview of the 2nd Voice Conversion Challenge 2018

Junichi Yamagishi, Jaime Lorenzo-Trueba  
(National Institute of Informatics),  
Tomoki Toda (Nagoya University),  
Daisuke Saito (Tokyo University),  
Fernando Villavicencio (ObEN),  
Tomi Kinnunen (University of Eastern Finland),  
Zhenhua Ling (University of Science and  
Technology of China)

### 概要

現在、声質変換の性能を共通のコーパスで競う「2nd Voice Conversion Challenge」が進行中であり、国際会議 Speaker Odyssey 2018 にて特別セッションが開かれる予定である。本デモでは、この 2nd Voice Conversion Challenge の要約と提出された 25 システムの結果を説明する。声質変換された音声の品質の評価では最大 MOS 値 4.1 が出るなど、これまで声質変換技術を大きく上回る結果となっている。

### 関連 Web ページ URL

<http://www.vc-challenge.org>

## Lyric Jumper : アーティストごとの歌詞の傾向を考慮したトピックモデルに基づく歌詞探索サービス

佃 洗根, 石田 啓介, 後藤 真孝 (産業技術総合研究所)

### 概要

各アーティストには、「恋愛」や「友情」といったトピックの歌詞が書かれやすい、といった傾向が存在する。本発表ではそうしたアーティストごとの歌詞トピックの傾向に基づいて様々な歌詞に出会える歌詞探索サービス Lyric Jumper (<https://lyric-jumper.petitlyrics.com/>) を紹介する。Lyric Jumper では、アーティストごとのトピックの傾向の可視化や、トピックの類似度に基づくアーティストの推薦などの機能を提供する。そうした機能を利用することで、Lyric Jumper によって様々なアーティストや楽曲の歌詞へのユーザの理解が深まることを目指している。Lyric Jumper の実現にあたり、我々はアーティスト

トゴとの歌詞トピックの傾向を推定するためのモデルを提案する。提案モデルでは、各アーティストがトピックの分布を持ち、その分布に応じて各歌詞にひとつのトピックが割り当てられる。歌詞データセットを用いた実験により、従来研究において歌詞トピックを推定する主要な手法である latent Dirichlet allocation (LDA) よりも提案モデルが優れていることを定量的に示した。

#### 関連 Web ページ URL

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

### TextAlive: 楽曲歌詞の Kinetic Typography 動画のための統合制作環境

加藤 淳, 中野 倫靖, 後藤 真孝 (産業技術総合研究所)

#### 概要

Kinetic Typography は、文字を動かして見せることで意味をより深く伝えられる手法として、楽曲のプロモーション動画などで頻繁に利用されている動画演出手法である。本発表では、楽曲に応じて歌詞の Kinetic Typography 動画を自動生成して自由に編集できる動画制作環境 TextAlive (<http://textalive.jp>) を紹介する。従来は、まず動画制作ツールの使い方に慣れる必要があった。さらに、歌詞を文字ごとに時刻同期させて動きを手付けするため、多くの手間がかかっていた。そこで、TextAlive は Web 上に公開されている楽曲と歌詞の時間的対応付けを自動で推定し、Kinetic Typography を用いた動画を自動生成する。また、生成された動画をインタラクティブに編集でき、動画制作の手間を従来より大幅に削減できる。さらに、文字の動きを決めるアルゴリズムをライブプログラミングでき、プリセットのテンプレートに縛られない自由な表現が可能である。デモ参加者は TextAlive で制作された動画を視聴したり、好みの楽曲から動画を自動生成したり、既存の動画を編集したりできる。

#### 関連 Web ページ URL

<http://textalive.jp>

### Songle Sync: 音楽に連動させて多様なデバイスを 大規模に制御できるプラットフォーム

尾形 正泰, 井上 隆広, 加藤 淳, 後藤 真孝  
(産業技術総合研究所)

#### 概要

音楽に連動してインターネット経由で 100 台以上のス

マートフォン・パソコン・IoT デバイスを制御する Songle Sync (<http://api.songle.jp/sync>) を開発した。音楽理解技術で得たビートやサビなどの情報に応じて表示・動きが変化する多様な演出を手軽に体験できる。CG を用いた映像・ロボット・照明などを音楽に連動させて新たな演出を生み出せるデモンストレーションを開発したほか、音楽連動アプリケーションを手軽に作成するための開発キットも公開した。音楽イベントでの実証実験を経て、音楽連動制御技術で同期した映像を来場者のスマートフォンに映して音楽体験を拡張できること、また混雑した通信環境でも利用可能なことを確認した。デモ会場では、スピーカーから流れる音楽に合わせてノート PC のほかに iPad, iPod touch といったモバイル端末の画面が一斉に同期するデモンストレーションを展示する。また、来場者がブースに表示された QR コードを読み取って、各自のスマートフォンでも Songle Sync のアプリケーションを体験できるデモンストレーションを提供する。

#### 関連 Web ページ URL

<http://api.songle.jp/sync>

### 楽曲中の歌声とユーザ歌唱のリアルタイムア ラインメントに基づく伴奏追従型カラオケシス テム

和田 雄介, 坂東 宜昭, 中村 栄太, 糸山 克寿, 吉井 和佳  
(京都大学)

#### 概要

本稿では、入力された音楽音響信号から伴奏音を抽出し、ユーザ歌唱のテンポ変化に自動で追従して再生するカラオケシステムを提案する。このシステムによって、ユーザは任意の楽曲を、テンポを自由にアレンジしながら歌うことが可能になる。このシステムの主な利点は、ユーザが楽譜 (MIDI ファイル) を用意する必要がないことと、システムを起動した後すぐにカラオケを楽しめることである。これらを実現するために、このシステムでは音源分離手法および audio-to-audio アラインメント手法をオンラインで並列に実行する。まず、入力された音楽音響信号が、ロバスト非負値行列因子分解 (RNMF) のオンライン版を用いて歌声と伴奏音に分解される。その後、分離された歌声信号とユーザ歌唱が、動的時間伸縮 (DTW) によって時間方向に同期される。最後に、DTW によって推定されたワーピングパスを用いて伴奏音が伸縮され、再生される。被験者実験により、このシステムの有効性が確認され、このシステムは新しい歌唱の楽しみ方を提示しうることが示された。

## 関連 Web ページ URL

<http://sap.ist.i.kyoto-u.ac.jp/members/wada/smc2017/index.html>

## フレーズ表現を多角的に俯瞰するための演奏表現支援システム

橋田 光代 (相愛大学)

### 概要

フレーズングはいわゆる発声や呼吸に関連した音楽的表現であり、言語に依らない意図の伝達手段として直感的に取り組みやすいものである。その表現に関わる方法論は、伝統的に往年の演奏家らの言を通じて経験的に語り継がれてきた知見が演奏解釈理論としていくつか取りまとめられているが、その使用者が優れたリスナーであることを前提としており、解決すべき多義性や曖昧性が残されている。本発表では、フレーズングに焦点を当てたユーザ主導型の演奏デザインにおいて、煩雑な作業の一部を自動化技術で代替するという方式の演奏デザインシステム Mixtract を介して、音楽・音声表現生成に共通する課題について多角的に議論する。

## 関連 Web ページ URL

<http://mixtract.m-use.net/>

## Songrium: Web-native music の俯瞰的な鑑賞を支援する音楽視聴支援サービス

濱崎 雅弘, 石田 啓介, 佃 洸撰, 深山 寛, 中野 倫靖,  
後藤 真孝 (産業技術総合研究所)

### 概要

Web で公開され、Web で視聴されてフィードバックを受け、Web で新しい派生作品が公開される、そうしたオープンコラボレーションが創発する音楽を、我々は Web-native music と呼んでいる。本発表では、そうした Web-native music の俯瞰的な鑑賞を支援する音楽視聴支援サービス Songrium (<http://songrium.jp>) を紹介する。Songrium では自動的に Web 上の楽曲を収集・分類し、様々な観点に基づく可視化インタフェースを提供する。音楽星図および Songrium3D では空間的な広がりを用いた可視化によりコンテンツの膨大さを感じられるインタフェースを提供する。バブルプレーヤおよび超歴史プレーヤではコンテンツの時間的な変化を感じられるインタフェースを提供する。さらに生成ランキングでは、公開された派生作品数に基づく新しいランキングを提供する。Songrium

の全ての機能は Web 公開中であり、デモンストレーションでは実際に触って各機能を参加者に体験していただく。

## 関連 Web ページ URL

<http://songrium.jp>

## リスナー間の楽曲嗜好傾向の可視化の一手法

吉久 怜子, 大矢隼士 (レコチョク),  
伊藤 貴之 (お茶の水女子大学), 山内 和樹 (レコチョク)

### 概要

リスナーが保有している楽曲のリスト群を用い、リスナー間での楽曲の視聴傾向の可視化する一手法を提案する。具体的には、リスナーと楽曲をノード、リスナー-楽曲間の視聴関係をエッジとし、リスナーはリスナー同士、楽曲は楽曲同士でクラスタリングする。このエッジがより束化するようにクラスタを配置することを目指す。

## 調とリズムを考慮した階層隠れセミマルコフモデルに基づく歌声 F0 軌跡に対する音符推定

錦見 亮, 中村 栄太, 後藤 真孝, 糸山 克寿, 吉井 和佳  
(京都大学)

### 概要

本稿では歌声 F0 軌跡から音楽的に自然な音符系列を推定する統計的手法を示す。歌声の発音時刻や F0 は楽譜に示されたビート時刻や音符の音高からの大きな逸脱を含むため、歌声 F0 軌跡の時間・周波数方向への離散化による音符推定の精度を向上するためには、楽譜の音楽的な自然さを表現する楽譜モデルが重要である。我々は調とリズムに依存する音符の音高を表現する楽譜モデルと楽譜 (音符系列) から時間・周波数方向に逸脱する歌声 F0 軌跡を表現する F0 モデルとを統合した準ビート同期階層隠れセミマルコフモデル (HHSMM: hierarchical hidden semi-Markov model) を提案する。楽譜モデルでは、確率的に生成された調に従って音符の音高が生成される。さらに、音符の開始位置はビートの 1 次元格子に定義されたマルコフ過程に従って生成される。F0 モデルでは、歌声の発音時刻の時間方向の逸脱、音符間における F0 の滑らかな遷移、F0 の周波数方向の逸脱が確率的に生成され、楽譜に付与される。提案法では、楽譜モデルと F0 モデルが音符推定に与える影響を考慮しながら、入力 of 歌声 F0 軌跡から尤もらしい音符系列を推定する。実験結果から調やリズムを考慮しない場合と比較して、提案法による音符系列の推定精度が向上することを示した。