

時を超えて蘇る 50 年前の歌声 ——スモールデータを用いたタスク混合深層学習による歌唱再現——

【発表概要】

東京大学大学院情報理工学系研究科 猿渡・小山研究室の高道慎之介助教らの研究チームは、東映株式会社ツクン研究所からの受託研究において、歌手の松任谷由実氏が 50 年前にデビューした当時の歌声を人工再現する技術を開発しました。

今回研究グループは、1970 年代から 80 年代にマルチトラック収録された松任谷由実氏のヴォーカルデータから声情報処理技術と機械学習技術を用いて、当時の声色と歌唱表現を忠実に再現することに成功しました。

本研究によって再現された歌声はデビュー当時のアーティスト名「荒井由実」として、松任谷由実氏の現在の歌声とデュエットを行います。このデュエットは 2022 年 10 月 1 日に楽曲「Call me back」のミュージックビデオ (MV) として YouTube で一般公開されました。

【発表のポイント】

- 複数の音声変換タスクを多段階に学習する機械学習と、これまで蓄積してきたデータベースおよび編集法により、1970 年代～80 年代に収録された僅かな歌声のみからその声色や歌いまわしを現代に対応する音質で新しい楽曲上で再現しました。
- この再現技術を使い、松任谷由実氏が「荒井由実」として活動していたデビュー当時の歌声を人工的に再現しました。
- 今回の研究成果の活用で、今後過去の歌声・音声データと現在の文化との相互作用による歌謡の新たな創作方法や鑑賞方法が生まれることが期待されます。

【発表内容】

直近 50 年の日本の歌謡文化を振り返ると、1970 年代から 1980 年代のニューミュージックブームを経て、1990 年代からは J-POP が人気を博しています。この間日本の歌謡曲は欧米の歌謡文化にリアルタイムに影響されることで発展し、ニューミュージックや、日本独自の J-POP 文化が形成されてきました。このように、文化間の作用は新たな文化を生み出すことに貢献します。一方、現存する過去の音資料は限定的であることから、数少ない貴重な音資料を、現代の音質で再現することができれば、現代を生きる我々にとっても、過去の歌謡文化を知り、新たな文化を生み出す源ともなることが期待されます。

高道慎之介助教、丹治尚子学術専門職員、研究室所属の学生が起業した Parakeet 株式会社から構成される研究グループは、音声情報処理と機械学習を専門としており、音声や歌声を高精度に再現する合成変換技術と、合成変換技術に資する音声・歌声データ設計技術を中心に研究してきました。

今回研究グループは、松任谷由実氏の 50 年前の歌声を人工的に再現することを目的に研究を開始しました。松任谷由実氏は、ニューミュージック時代から J-POP 全盛期の現代においても第一線で活躍し続ける音楽家ですが、現存するデビュー当時の音資料は限定的であり、また当然ながら歌声再現を目的とした資料ではないため、従来の情報工学技術では再現が困難でした。

この問題解決のため、研究グループは①多段階合成変換タスク混合学習アルゴリズムと、②実データに利用可能なデータ編集・枝刈り法（機械学習に不要なデータを半自動的に削除する方法）を使った手法を提案しました。

①では、テキストから話声を合成するテキスト音声合成、歌詞から歌声を合成する歌声合成、歌声を別の歌声に変換する歌声変換を一括で機械学習できるアルゴリズムを用います。今回研究グループは当時の音資料の少なさを補うため、グループが設計し、2020年に発表した、JUST・JVSなどの多数の話声・歌声データベースを用いました※1。この大規模な話声データベースを用いたテキスト音声合成により「読み上げとしての声らしさ」を機械学習し、また、歌声データベースを用いた歌声合成には「歌い上げとしての声らしさ」を機械学習しています。最終的な当時の歌声の再現には、歌声変換を利用しています。この歌声変換により、ある歌唱者の歌声をもとに、松任谷由実氏の当時の声色や韻律をもつ歌声を生成しました。

②では、①の機械学習に資するよう、当時の音資料を半自動的に編集・枝刈りする手法を用いています。歌声の抑揚のゆらぎは歌手特有の表現になりえる一方、①の機械学習で扱うことが困難です。これに対し、我々がこれまで提案した「ゆらぎを機械学習する」方法もありますが、今回は逆に、データに対してゆらぎを抑える方法を提案し、機械学習の得意不得意に合わせて音資料を編集・枝刈りしています。

本研究による歌声再現技術でデビュー当時の「荒井由実」として再現された歌声は、「松任谷由実 with 荒井由実」のアーティスト名義で、松任谷由実氏のMV、また50周年記念アルバム内でも採用されました。MVでは東映株式会社トークン研究所が制作したデジタルヒューマンが本研究で再現された50年前の声で歌います。※2

本研究は、長きにわたり日本の音楽シーンの第一人者である松任谷由実氏の協力を得て実施され、過去と現代の歌唱の文化間作用を模索する第一歩になることが期待されます。また、これまでの映像制作の現場では難しかった「過去の声質の再現」を可能にしたことで、今後の映像表現の可能性が広がるだけでなく、過去の歌声が現代に蘇ることによる新たな文化の創造が期待されます。研究グループは今後も音声情報処理技術と深層学習技術を活用し、現存する過去の音楽資産を現代の音環境と歌謡文化に採用できる手法の開発に取り組み、多様な歌声と保存環境にも対応できる技術を活用することで、将来的に歌謡文化に限らずより広い音声言語文化に研究成果を展開することを目指しています。

【発表者】

高道 慎之介（東京大学 大学院情報理工学系研究科システム情報学専攻 助教）

【参考情報】

※1 “JSUT and JVS: free Japanese voice corpora for accelerating speech synthesis research,”

Shinnosuke Takamichi*, Ryosuke Sonobe, Kentaro Mitsui, Yuki Saito, Tomoki Koriyama, Naoko Tanji, Hiroshi Saruwatari, Acoustical Science and Technology, Vol.41, No.5, pp.761--768, Sep. 2020.

DOI:<https://doi.org/10.1250/ast.41.761>

※2 「Call me back」の動画 URL

<https://www.youtube.com/watch?v=oWo-TabDt8w>

“50年の時を超えて”をテーマに、今回の研究で再現された「荒井由実」の歌声と松任谷由実氏のデュエットによる楽曲「Call me back」は、1980年のオリジナルトラックを元に新たに歌詞を書き下ろしたもの。“松任谷由実 with 荒井由実”のアーティスト名で発表された。

【添付資料】

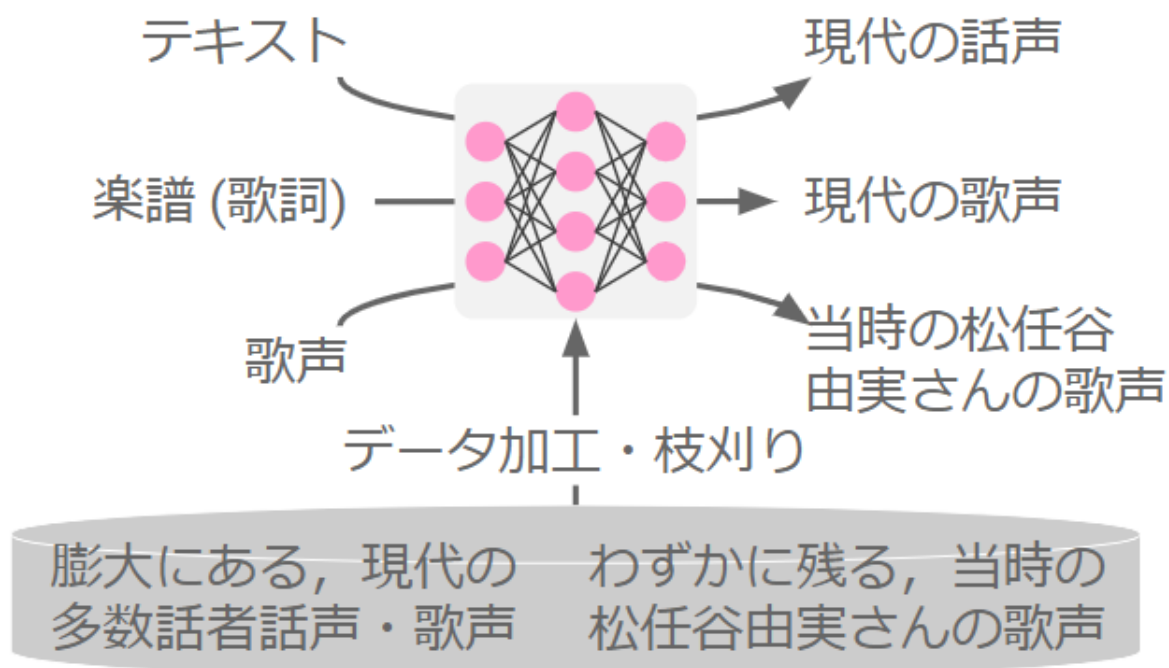


図1：タスク混合学習アルゴリズム（中央）では、テキスト音声合成（テキストから現代の音声を合成）や歌声変換（ある歌声を、当時の荒井由実氏の歌声に変換）など複数のタスクを扱う。またデータ編集・枝刈り法（下）ではアルゴリズムに適するように当時の音声データを編集・選択する。