

## どんな研究

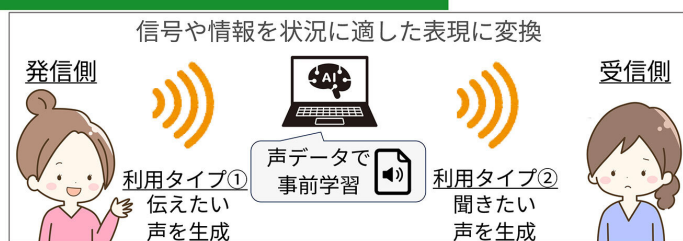
音声生成AIの研究開発用データは、声優など声による表現を職業の中核とする話者による演技的な発話に偏りがちです。各AIユーザにとって魅力的な音声の生成の実現をめざす第一歩として、**多様な魅力をもつライブアイドルを話者とする多数話者音声データセットを構築しました。**

## どこが凄い

本データセットは、研究利用のために契約処理が適切になされた**初の大規模アイドル音声データセット**です（話者数200人以上、総時間約30時間）。他の音声データセットのような読み上げ音声や日常会話のみならず、チェキ会での発話などアイドル特有の魅力的な発話も含んでいます。

## めざす未来

多様な人の音声を学習することで、**誰もが自分の声を柔軟に魅力的に整えられる音声生成AI**、いわば「声のメイク」の実現をめざしています。この試みにより、声へのコンプレックスや発話へのためらいを軽減し、積極的な自己表現や活発なコミュニケーションを促すことをめざします。

音声生成AIによる  
コミュニケーション機能拡張

実現したい機能拡張：**日常に溶け込む魅力的な声の生成（声のメイク）**

魅力的な声とは？

- ①聴取者によって異なる  
②声質や話し方により複合的に決まる } 魅力のあり方は多様

研究の第一歩として、**多様な魅力をもつ声を収集**

	JVS	CSJ	JIS (本成果)
話者	声優/ ナレーター	一般話者	ライブ アイドル
規模	約30時間 100人	約660時間 1417人	約30時間 204人
日常会話らしさ	△	○	○
魅力	○	△	○

※ JVS: Japanese Versatile Speech corpus、CSJ: Corpus of Spontaneous Japanese

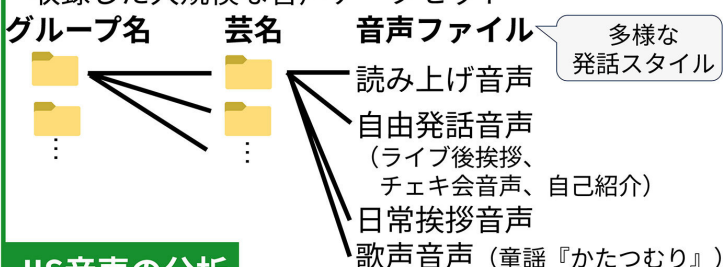
## 選定した話者ジャンル

ライブアイドル: ライブを主な活動とするアイドル  
音声生成AI研究を行う上での利点

- ・ **プライバシーの保護とファンへのアクセス可能性の両立**（理由：芸名）
- ・ **魅力の理解者によるきめ細やかな主観評価が可能**（理由：明確なキャラクター性とファンの存在）
- ・ **多数話者の収録が容易**

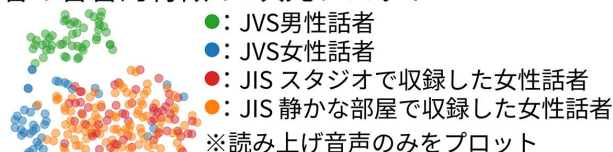
## Japanese Idol Speech Corpus (JIS)

- ・ 多数の女性ライブアイドル音声を収録した大規模な音声データセット



## JIS音声の分析

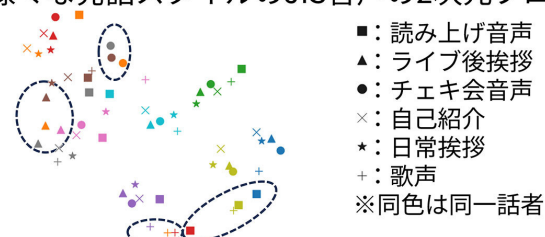
- ・ 話者の音響的特徴の2次元プロット



JVS女性とJISが分かれて分布

➡ JVSに十分に含まれていない話者ジャンルをカバー

- ・ 様々な発話スタイルのJIS音声の2次元プロット



話者ごとの違いに加え、発話スタイルごとの特徴も分布上に現れている (○の箇所)

- ➡ スタイルの違いが聞き手の印象と関連している可能性を示唆 (例: ライブ後挨拶⇒張りのある声、チェキ会音声⇒親密な声)
- ➡ **利用シーンに応じた魅力的な声の生成への活用可能性**

## 関連文献

[1] Y. Kondo, H. Kameoka, K. Tanaka, T. Kaneko, "JIS: A Speech Corpus of Japanese Idol Speakers with Various Speaking Styles," in *Proc. INTERSPEECH*, pp. 4783-4787, 2025.

## 連絡先

近藤 祐斗 (Yuto Kondo) メディア情報研究部 事象モデリング研究グループ

## 共同研究先・外部資金

本研究のうち、音声収録作業（話者募集、収録、データセット編集）に関する部分はJST CREST (JPMJCR19A3) の支援を受けて実施しました。