

TSUBAME 共同利用 令和7年度 産業利用 成果報告書

利用課題名 深層学習を用いたリアルタイム AI 音声合成プラットフォームの技術開発
英文: Real-time AI Voice Synthesis Platform using Deep Learning

利用課題責任者 竹内 雅樹
Masaki Takeuchi

所属 株式会社 CoeFont
CoeFont Co., Ltd.
URL <https://www.coeFont.com/ja/>

邦文抄録(300 字程度)

本研究では、深層学習技術を活用したリアルタイム AI 音声合成プラットフォームの高度化を目的とし、大規模音声モデルの開発および AI 通訳システムの機能拡張を行った。具体的には、感情表現と自然な発話リズムを両立した最新音声モデル「CoeFont v3 Fuji」を開発・公開した。また、リアルタイム AI 通訳サービス「CoeFont 通訳」において、iOS 版の提供開始、多言語対応の拡充、企業向けのコンテキスト理解機能の追加、Microsoft Teams 連携などの機能拡張を実施した。TSUBAME の計算資源を活用した大規模深層学習により、音声合成および翻訳処理の高速化と高品質化を実現し、リアルタイム多言語コミュニケーション基盤の社会実装を推進した。

英文抄録(100 words 程度)

In this study, we advanced a real-time AI voice synthesis platform using deep learning techniques. We developed and released the latest speech synthesis model, **CoeFont v3 Fuji**, which improves emotional expressiveness and natural speech rhythm. In addition, we expanded the real-time AI interpretation service **CoeFont Interpreter**, including the release of an iOS application, multilingual support expansion, context-aware interpretation using uploaded documents, and integration with Microsoft Teams. By utilizing the TSUBAME supercomputing environment, we accelerated large-scale model training and inference. These achievements contribute to the practical deployment of real-time multilingual communication platforms integrating speech synthesis and interpretation technologies.

Keywords: Deep learning, AI voice synthesis, real-time speech synthesis, multilingual interpretation, voice AI platform

背景と目的

近年、深層学習技術の発展により、音声合成や機械翻訳の精度は大きく向上している。特に、オンライン会議や国際ビジネス、教育、観光などの分野では、リアルタイムで自然な音声生成と多言語翻訳を実現する技術の需要が急速に高まっている。

しかし、従来の音声合成技術では、発話の自然さや感情表現、長時間生成の安定性、リアルタイム推論の処理速度などに課題が存在していた。また、多言語コミュニケーションの分野では、通訳人材の不足やコストの高さ、即時対応の難しさが課題となっている。

本プロジェクトでは、これらの課題を解決するため、深層学習を用いたリアルタイム AI 音声合成プラットフォ

ームの技術開発を行う。具体的には、大規模音声モデルの開発とリアルタイム翻訳機能を統合した AI 音声プラットフォームを構築し、TSUBAME の大規模計算資源を用いたモデル学習および推論高速化により、高品質かつリアルタイムな音声合成・通訳システムの実現を目的とした。

概要

本研究では、深層学習を用いた AI 音声プラットフォームの開発に向けて、以下の取り組みを行った。

1. 最新 AI 音声モデル「CoeFont v3 Fuji」の開発

CoeFont v3 Fuji は、人間の発声特性を解析し、抑揚・間・リズムなどを自然に再現する音声合成モデルで

あり、感情表現の再現性を大きく向上させた。

本モデルは、喜怒哀楽などの感情表現を繊細に再現できる点が特徴であり、動画制作、ナレーション、アプリケーション UI など多様な用途に対応可能である[1]。TSUBAME の GPU 環境を利用した大規模学習により、音声品質の向上と推論速度の最適化を実現した。

2. リアルタイム AI 通訳「CoeFont 通訳」の開発と拡張

AI ライブ翻訳サービス「CoeFont 通訳」を開発し、リアルタイム音声認識・翻訳・音声合成を統合した通訳システムを実装した。

主な機能拡張は以下の通りである。

- iOS アプリ版の提供開始
- 通訳開始時間の高速化(最短短約 1 秒)
- Word / PDF / PowerPoint 等の資料を学習するコンテキスト理解機能
- 日本語、英語、中国語、韓国語、フランス語、スペイン語、ドイツ語、ベトナム語、タイ語、ポルトガル語といった多言語対応の拡張
- Microsoft Teams 向けプラグインの提供

これにより、オンライン会議や国際コミュニケーションにおいて、AI 通訳を容易に導入できる環境を整備した[2][3][4][5]。

結果および考察

本研究により、以下の成果を得た。

高品質音声合成モデルの実現

CoeFont v3 Fuji の開発により、従来モデルと比較して以下の改善が確認された。

- 人間らしい発話リズムの再現
- 感情表現(喜怒哀楽)の精度向上
- 長時間でも聞き疲れしない音質

これにより、コンテンツ制作やナレーションなどの用途において、より自然な音声生成が可能となった。

リアルタイム AI 通訳システムの実用化

AI 通訳システム「CoeFont 通訳」では、

- 通訳開始まで最短短約 1 秒
- 多言語対応(最大 10 言語)
- モバイル環境やオンライン会議環境での利用

などの改善を行い、ビジネス現場での利用に適したリアルタイム通訳環境を構築した。

さらに、資料アップロード機能により、企業固有の用語や文脈を理解した通訳が可能となり、従来の機械翻訳よりも実用性の高い通訳システムを実現した。

まとめ、今後の課題

本研究では、深層学習を用いたリアルタイム AI 音声合成プラットフォームの技術開発を行い、最新音声モデル「CoeFont v3 Fuji」の開発および AI 通訳サービス「CoeFont 通訳」の機能拡張を実現した。

TSUBAME の計算資源を活用した大規模モデル学習により、高品質な音声合成とリアルタイム翻訳を統合した AI 音声プラットフォームの基盤を構築できた。

今後の課題としては、以下が挙げられる。

- より多様な言語への対応拡大
- 感情表現や話者特徴のさらなる高度化
- リアルタイム推論のさらなる低遅延化
- 医療・教育・観光など多様な分野への応用展開

今後も TSUBAME との連携を継続し、大規模深層学習モデルの研究開発を進めることで、AI 音声技術の社会実装を推進していく予定である。

参考文献

[1] AI 音声はついに”感情”の領域へ！ CoeFont、最新音声モデル「CoeFont v3 Fuji」を正式リリース
<https://prtimes.jp/main/html/rd/p/000000103.000078329.html>

[2] 国内発の AI ライブ翻訳サービス『CoeFont 通訳』、iOS 版を正式リリース
<https://prtimes.jp/main/html/rd/p/000000104.000078329.html>

[3] 日本発の AI ライブ翻訳サービス「CoeFont 通訳」、ファイルアップロード対応で“会社事情を理解する専属 AI 通訳”へ進化
<https://prtimes.jp/main/html/rd/p/000000107.000078329.html>

[4] 日本発の AI ライブ翻訳サービス「CoeFont 通訳」、PC・iOS 版で新たに韓国語・ベトナム語に対応

[https://prtimes.jp/main/html/rd/p/000000109.00078329.html](https://prt看imes.jp/main/html/rd/p/000000109.00078329.html)

[5] CoeFont、リアルタイム翻訳サービス「CoeFont 通訳」に Microsoft Teams 向けプラグイン提供を開始

<https://prtimes.jp/main/html/rd/p/000000117.00078329.html>