

ABCI 3.0開発加速利用（2024年度）成果概要（公開用）

課題名：音声・音響・振動データの大規模基盤モデル構築および利活用

実施時期：2025年1月～2025年3月

所属機関名：産総研、東大、早大、慶應大、東京農工大

代表者氏名：深山 覚（産総研）

成果概要：

大規模言語モデル (Large Language Model; LLM) の汎用的な言語処理能力を活用した音声言語処理システムを開発した。LLM の自然言語処理タスクにおける推論過程を音声処理モデルに組み込み、音声と言語の情報を相補的に統合するととて、効率的かつ高精度な音声言語処理を実現した。またAIで生成された音声の品質を評価するための深層学習モデルを活用した音声合成技術を開発した。

成果のポイント：

① 大規模言語モデルの汎用的な言語処理能力を活用した音声言語処理システム

音声認識・音声翻訳システムにおける自然言語処理タスクを LLM に解かせることでその言語知識を活用した音声言語処理を実現した。具体的には、音声認識では単語誤り訂正を、音声翻訳ではテキスト翻訳の実行を LLM に実行させるプロンプトを設計した（図1）。主要なベンチマークを用いた評価実験の結果、提案システムは音声認識・音声翻訳の両タスクで従来法を大きく上回る性能を示した。今回の開発によって、LLM の汎用的な言語処理能力は音声言語処理に転用できることがわかった。

② AI 生成音声の自動品質評価モデルを活用した音声合成技術の開発

音声を生成する基盤モデルの研究開発を効率化するためには、その評価を自動化する機械学習モデルを用いた音声合成技術の開発を行なった。高品質な AI 生成音声の自動品質評価に特化した深層学習モデル「UTMOSv2」を活用し、音声合成モデルにより生成された音声の品質を改善するように再生成しなおすアルゴリズムを開発した。

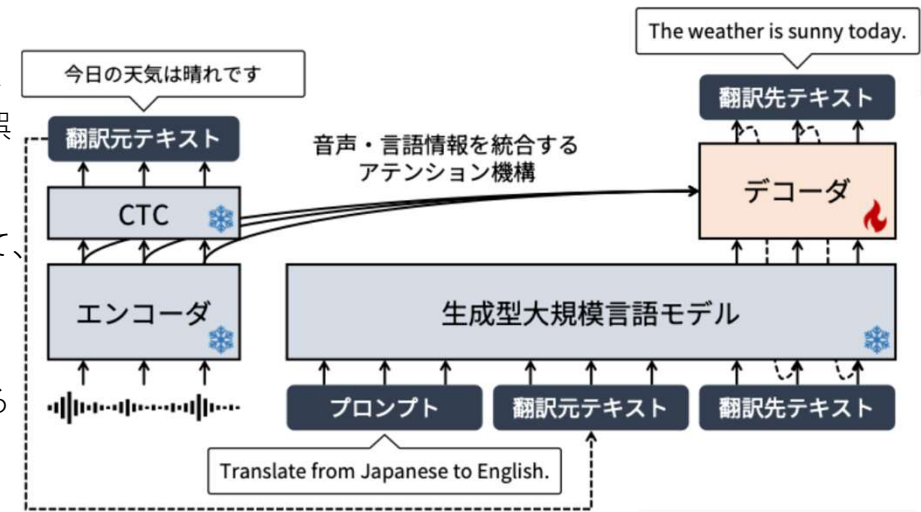


図1. 大規模言語モデルを活用した音声翻訳システム

成果についてより詳細な情報を提供しているWebページ、発表論文などの情報：

- ・ 音声翻訳に関する研究成果を、音声言語処理分野の主要国際会議である 26th Annual Conference of International Speech Communication Association (INTERSPEECH) に投稿。
- ・ 高品質な AI 生成音声の自動品質評価に特化した深層学習モデル「UTMOSv2」：GitHub <https://github.com/sarulab-speech/UTMOSv2>