

ポスト5G基盤強化の研究開発・事業成果概要

実施期間 2024年度～2025年度

実施者 株式会社Kotoba Technologies Japan

事業成果

事業概要

音声をリアルタイムにEnd-to-Endでやりとりできる音声基盤モデルの開発を行った。音声基盤モデルの日本語音声能力・推論速度を飛躍的に高め、同時通訳（Speech-to-Text, Speech-to-Speech）を中心した音声アプリケーションへの応用も進めた。

事業概要に沿い、リアルタイム音声基盤モデルの開発を行った。計画に基づき、初期は合成データを含めた学習データの構築に重点を置き、フィルタリングを施しながら高品質データを準備した。モデル開発においては、リアルタイムのtwo-stream音声基盤モデルの開発に予定通り成功し、同時音声翻訳や音声生成において、大きな成果を出した。とりわけ音声翻訳での成果が著しく、音声からテキスト、または一部音声から音声の翻訳が日英中韓4言語で達成された。日韓においては、言語としての語順が近いため、より正確かつ高速な翻訳が可能となった。このようなリアルタイムな音声基盤モデルの開発例は、国内外で非常に珍しく、国際的競争力のある知見蓄積に繋がった。

技術開発だけでなく、社会実装に関しても大きな進捗を生むことができた。本事業のモデルを活用した同時通訳アプリは幅広く利用されており、すでに数万ダウンロードを超えており、継続的に利用するユーザーも数千人に昇る。またテレ東Bizなどでも取り上げられ、特に日英翻訳で使用されている。エンターテインメント（韓国アイドルのライブイベントなど）領域でもユーザーから使用され始めている。当社の今後の事業開発においても、重要なデータポイントとなっている。

リアルタイム音声基盤モデルの開発を行い、国内外でも極めて挑戦的で競争力の高いモデルの開発に取り組んだ。より具体的には、一般的な基盤モデルとは大きく異なり、左図のようなtwo-streamモデル（「聴きながら話す」ことができるモデル）のインプットアウトプットを活用できるモデルを開発し、知見を深めていった。同時音声翻訳や音声生成のような社会的経済的にインパクトの大きい下流タスクでも検証を行った。

社会実装イメージ

開発に取り組んだ音声基盤モデルは、非常に大きな社会実装の可能性を秘めている。例えば、コールセンターにおいては、高速書き起こしだけでなく、外国人対応のニーズなどが挙げられる。また、エンターテインメント領域においては、多言語ライブ配信、また将来的に大きく成長することが期待される音声エージェントへの導入も考えられる。

