＜年次報告書＞
第 1 年度（2025 年度）
日 ASEAN 科学技術・イノベーション協働連携事業（NEXUS）
日本－シンガポール「AI」研究領域

研究課題名（和文）	マルチリンガル･マルチモーダルな大規模言語モデルを用いた表現豊かで共感的な人間と AI のインタラクション
研究課題名（英文）	Expressive and Empathetic Human-AI Interaction by Enhancing Multilingual, Multimodal Large Language Models
研究期間	2025 年 4 月 1 日～2028 年 3 月 31 日

氏名	所属 部署　役職
河原 達也	京都大学 大学院情報学研究科　教授
SAKTI Sakriani	奈良先端科学技術大学院大学 先端科学技術研究科　教授

1. 研究進捗状況と成果

以下の４テーマについて研究を実施した。

＜研究テーマ１＞　多言語及び方言に対応した音声認識

音声認識を多数の言語に対応できるようにするために、国際音声記号(IPA)に基づくモデルを設計・学習した。データ量が限られたアイヌ語や沖縄語（首里方言）においても、多言語で事前学習されたモデルを５時間程度のデータでファインチューニングすることで、90%程度の認識精度を得られることを示した。

＜研究テーマ２＞　音声対話における個性・感情のモデル化

音声対話において、感情に加えて、個性のアノテーションを行ったデータセットを構築し公開した。個性と感情の関係をモデル化した上で、このデータセットを用いてモデル学習を行い、両者を統合的に認識するシステムを実現した。

＜研究テーマ３＞　マルチモーダルな共感的応答生成

感情認識に基づいて、ユーザに共感的な応答を生成する方法を実装した。バリデーションというカウンセリングで使用されている手法に着目し、これを行うタイミングの判定及び応答生成の研究を行った。言語情報と音声情報を各々活用し、両者を統合する方法を検討した。また、多言語のモデル化も行った。

＜研究テーマ 4＞　ロボット・エージェントにおける実装

上記に基づく音声対話システムを人間型ロボットや仮想エージェントに実装し、評価実験を行った。相槌や頷きを含む共感的な応答生成が高い評価を得ることを確認した。

＜「国際研究交流」活動＞

- 2025 年 5 月、7 月、2026 年 1 月にオンラインでミーティングを行い、研究発表や議論を行った。
- 2025 年 10 月に、日本側研究者がシンガポールに出張し、対面でミーティングを開催した。
- 日本側から２名の博士課程学生をシンガポール側に派遣し、各々３カ月間実習生として本研究に従事した。各々の成果は論文にとりまとめて投稿した。
- 国際会議 ICMI2025 及び AAAI2026 において、本研究テーマに関連するワークショップを共同で開催した。

2. 代表的な成果事例

1. R.Magoshi, S.Sakai, J.Lee, and T.Kawahara. Multi-lingual and zero-shot speech recognition by incorporating classification of language-independent articulatory features. In Proc. INTERSPEECH, pp.91--95, 2025.
   言語に独立な調音特徴の分類を併用することで未知の少数言語の音声認識も実現した。
2. C. Tran and S. Sakti.Toward High-Quality Cross-lingual Text-to-Speech Synthesis in Low-Resource Scenarios. In Proc. AAAI Workshop on Audio-Centric AI: Towards Real-World Multimodal Reasoning and Application Use Cases (Audio-AAAI), 2026.
   多言語のデータを効果的に利用することで低資源言語の音声合成を実現した。