【基盤研究（S）】

総合系（情報学）

研究課題名　次世代音声翻訳の研究

奈良先端科学技術大学院大学・情報科学研究科・教授　中村（なかむら）　哲（さとし）

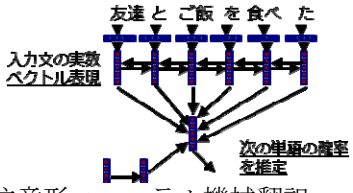
研究課題番号：　17H06101　研究者番号：30263429
研　究　分　野：　情報学、人間情報学、知覚情報処理、音声情報処理
キ　ー　ワ　ー　ド：　音声翻訳

【研究の背景・目的】

短い旅行会話を対象に一発話終了毎に翻訳する音声翻訳は実用化が進んでいるが、人間の通訳者が行うような同時通訳は格段に困難である。特に文構造が異なる日本語から英語の通訳では、文末に来る動詞や否定を待つか予測しなければ訳出ができない。本研究では、講演、講義、会議を対象に、人間の同時通訳者のように文末を待たずに即座に通訳出力ができ、文構造の違いで破綻することなく、発話者の意図を伝えることができる次世代音声翻訳の研究開発を行う。

【研究の方法】

① 雑音下常時音声認識、自動音声同時通訳、音声翻訳の高度化：A) 既知雑音ビッグデータに基づく深層学習（DNN）による雑音抑圧法、「独立低ランク行列分析」の導入・融合、会議状況で複数の分散型・位置不定マイクセンサ群から対象話者の音声を抽出する方法の発展、B) 衆議院の自動音声認識速記システムをベースに常時音声認識が講演、講義同時通訳用に動作するように改良、C) 次発話の部分木構造を現時点までの構文解析結果から予測し、Tree-to-string 翻訳モデルで高精度に翻訳を行う手法による同時通訳の高度化、D) LSTM による注意形ニューラル機械翻訳（NMT）の入出力層を圧縮する機械翻訳の高度化について検討、E) 対話制御の導入



② パラ言語音声翻訳：A) 発話の強調のパラ言語翻訳および感情翻訳への展開、B) パラ言語情報を対象言語に付与した音声合成技術の構築

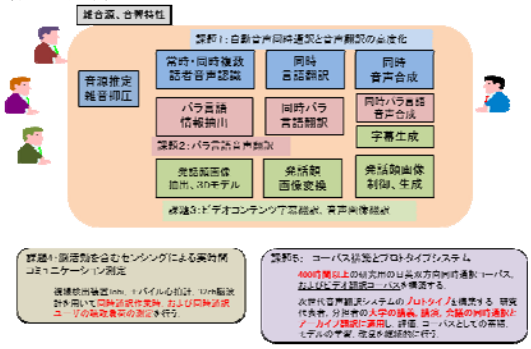
③ ビデオコンテンツ字幕翻訳、音声画像翻訳：A) 自動要約を適用した音声からテキストの音声翻訳、B) 音声リップシンク画像による音声画像翻訳の実現

Model Fitting

Mouth Model

④ 脳活動を含むセンシングによるリアルタイムコミュニケーション測定：同時通訳作業時、および同時通訳ユーザの聴取負荷の観察および測定

⑤ コーパス構築とプロトタイプシステム：A) 400 時間以上の研究用の日英双方向同時通訳およびビデオ翻訳コーパスの構築、B) 課題①～④の技術の統合



【期待される成果と意義】

①雑音下での複数発話者の音声を常時音声認識し、言語間での文構造の違いを考慮して五月雨式に通訳する自動音声同時通訳と音声翻訳方式の確立

②発話者の感情、強調、話者性等を抽出、保持、生成するパラ言語音声翻訳の実現

③講演、映像などのビデオコンテンツの字幕翻訳、音声画像翻訳方式の導入

④脳活動を含むセンシングによる通訳時認知負荷測定データの分析

⑤400 時間以上の研究用の日英双方向同時通訳コーパス、およびビデオ翻訳コーパスとプロトタイプシステム構築

【当該研究課題と関連の深い論文・著書】

・Yusuke Oda, Graham Neubig, Sakriani Sakti, Tomoki Toda, Satoshi Nakamura, "Syntax-based Simultaneous Translation through Prediction of Unseen Syntactic Constituents", The 53rd Annual Meeting of the Association for Computational Linguistics (ACL) Long Paper Track. pp. 198-207. July 2015.
・中村 哲, “話し言葉の音声翻訳技術,” 電子情報通信学会誌, vol.96, no.11, pp865-873, 2013.11

【研究期間と研究経費】

平成 29 年度－33 年度　157,100 千円

【ホームページ等】

特になし