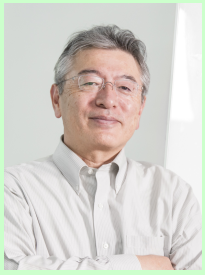
科学研究費助成事業（基盤研究（Ｓ））公表用資料
〔令和２(2020)年度 研究進捗評価用〕

平成２９年度採択分
令和２年３月３１日現在

次世代音声翻訳の研究

Next Generation Speech Translation Research



課題番号：17H06101

中村　哲（NAKAMURA, SATOSHI）

奈良先端科学技術大学院大学・データ駆動型サイエンス創造センター・教授

研究の概要（４行以内）
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，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を始めとする音声翻訳の高度化の研究を行う．

研　究　分　野：情報学，人間情報学，知覚情報処理

キ　ー　ワ　ー　ド：音声情報処理

１．研究開始当初の背景

我が国は 2020 年には 4000 万人の訪日外国人を受け入れる計画である．このような，急激な訪日外国人の増加，日本の社会，企業の国際化に伴い，外国人とのコミュニケーションが必要な場面が急増している．旅行会話のような単純な文に対する音声翻訳技術については TOEIC650 点に匹敵する性能を達成するなど実用化が進んでいる．しかし，講演，講義，会議の自動音声同時通訳は格段に困難で研究が進んでいない．

２．研究の目的

本研究では，講演，講義，会議を対象に，①雑音下での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の高度化の研究を中心に，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，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，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，の研究を行い，⑤同時通訳，ビデオ翻訳コーパス構築とプロトタイプシステムを構築する．

３．研究の方法

人間の同時通訳者の認知モデルを考慮しつつ，新たな確率モデルや再帰的深層ニューラルネットワークに基づく高精度な音声翻訳の処理モデルを構築する．同時にこのモデルを学習するための 400 時間の大規模同時通訳、ビデオ翻訳コーパスを収集する．

４．これまでの成果

自動音声翻訳の研究を著書にまとめた[1].

4.1.1 雑音下音声認識の研究: 独立低ランク行列分析（ILRMA）の理論拡張（複素 Student’ｓt 分布生成モデルへの拡張）および DNN 音源モデルとの融合として独立深層学習行列分析（IDLMA）を提案した。IDLMA の成果が IEEE/ACM Trans. ASLP に採択され、2019 年 6 月の公開以来 2200 件以上のダウンロードを達成した[2].

4.1.2 音声認識：単語単位の End-to-End 音声認識を提案し、従来の DNN-HMM に基づく方式に比べて、はるかに単純な構成で、30 倍以上の高速化（リアルタイム比）を実現した。また，低遅延のインクリメンタル音声認識を提案した[3]．注視型 Encoder-decoder 型をベースに文全体を入力して注視するモデルを教師(teacher)とし，漸進的処理のために短いセグメント単位で注視を行うモデルを生徒(student)として，モデル学習を行う手法を提案し，文単位の入力を利用した場合からの精度低下を抑えられることを確認した[4]．さらに，Encoder-decoder 型ニューラルネットワークに基づくテキスト音声合成に漸進的な処理が可能な改良を加え主観評価実験により少量の遅延により漸進的音声合成が実現できることを示した．

4.2 同時通訳：語順の違う言語対に対しても適応的に遅延をコントロールして同時通訳を実現する方法，同時通訳学習データのデータ拡張法を提案した.「順送りの訳」という，英文の要素を前から小分けにして訳出し，関係詞節はその手前で一旦文を区切

る方策により，入力トークン列に対して k トークンの入力を待ってから翻訳出力する既存の wait-k 法を，英語と日本語の間のような語順の差が大きい場合にも適用できるように，デコーダの出力記号の一つにトークンを出力せず次の入力を待つ特殊記号を追加する方式を提案した．適応的に入力待機を行い漸進的な翻訳による精度低下を小さく抑えられることを確認した[5]．

4.3.1 発話の強調のパラ言語翻訳： LSTM ニューラルネットを用いたシステムをベースに，強調情報を抽出し言語翻訳モデルの中の注意情報として扱い目的言語の音声を強調する方法を提案し有効性を示した[6]．

4.3.2 元言語の発話の話者性，感情などパラ言語情報を対象言語に付与し，適切に意図を伝達できる同時通訳用の音声合成技術の構築を行った．音声変換技術に対して、国際的な技術評価会 Voice Conversion Challenge 2018 を開催し音声変換技術の改善に大きく貢献した[7]。

4.4.1 字幕翻訳：大学の講義アーカイブの音声（日本語）を書き起こし，人手翻訳を行い，機械翻訳の学習データを作成した．このデータを用いて，翻訳字幕を表示する日英講義アーカイブ翻訳システムを NAIST の協力を得て実現し学内で運用中である[8]．

4.4.2 音声画像翻訳：人物のインスタントモデリングとして一切のセンサー類を利用することなく、１枚の顔画像のみから、顔の３次元形状とアルベド、ディスプレースメント、スペキュラー情報を推定して、新しい照明環境下においてアバタをフォトリアリスティックに実現することを可能とした。

4.5 同時通訳中の認知負荷：ASSR（auditory steady-state response）を同時通訳中に呈示し，ASSR から誘発される脳波計信号の位相同期を計測することで，同時通訳中の認知負荷の定量化に ASSR が有効であることが示した．

4.6.1 同時通訳コーパス：国際シンポジウムの同時通訳（英日・日英合計 5.5 時間），Web 上で公開されている講演の同時通訳（英日 79 時間），日本語話し言葉コーパスの同時通訳（日英 45 時間）および日本語記者会見の同時通訳（日英 5.5 時間），講演同時通訳（英日 50 時間，日英 40 時間）および日本語記者会見の同時通訳（37 時間）の合計 262 時間分を収録した．

4.6.2 全体技術を統合し，同時通訳システム，ビデオコンテンツ字幕翻訳システムのプロトタイプを構築した．本試作システムにおける各モジュール間の接続は (1) テキストによる標準入出力（パイプ接続） (2) 処理統括サーバとの相互通信 のいずれかで行う設計となっている[9]．

５．今後の計画

講演，講義，会議に対し，言語間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の高度化を進める．同時通訳技術については，方式研究だけでなくコーパスの構築，共有，通訳者の作業支援，通訳の品質評価なども，通訳者の協力を得ながら進めていく予定である．

６．これまでの発表論文等（受賞等も含む）

(1) 中村 哲，Sakriani Sakti，Graham Neubig，戸田智基，高道慎之介，“音響サイエンスシリーズ 18 音声言語の自動翻訳，” 日本音響学会編，コロナ社 2018.7.10

(2) Naoki Makishima, Shinichi Mogami, Norihiro Takamune, Daichi Kitamura, Hayato Sumino, Shinnosuke Takamichi, Hiroshi Saruwatari, Nobutaka Ono, ″Independent deeply learned matrix analysis for determined audio source separation,″ IEEE/ACM TASLP, vol. 27, no. 10, pp. 1601-1615, 2019.

(3) M. Mimura, S. Sakai, T. Kawahara, “Forward-backward attention decoder,” INTERSPEECH, Sep, 2018

(4) Sashi Novitasari, Andros Tjandra, Sakriani Sakti, Satoshi Nakamura, “Sequence-to-sequence Learning via Attention Transfer for Incremental Speech Recognition,” Proc. of Interspeech 2019, 2019

(5) 帖佐 克己，須藤 克仁，中村 哲，“英日同時通訳におけるニューラル機械翻訳の検討，” 言語処理学会 第 25 回年次大会 (NLP2019), Mar, 2019

(6) Quoc Truong Do, Sakriani Sakti, Satoshi Nakamura, “Sequence-to-Sequence Models for Emphasis Speech Translation,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 26, 2018, pp. 1873 - 1883

(7) P. L. Tobing, Y.-C. Wu, T. Hayashi, K. Kobayashi, T. Toda, “Voice conversion with CycleRNN-based spectral mapping and finely-tuned WaveNet vocoder,” IEEE Access, Vol. 7, No. 1, pp. 171114-171125, 2019.

(8) 須藤 克仁，林 輝昭，西村 優汰，中村 哲，“授業アーカイブの翻訳字幕自動作成システムの試作，” 情報処理学会研究報告, Vol. 2019-NL-240, No. 15

(9) 中村哲，Sashi Novitasari，帖佐克己，柳田智也，二又航介，須藤克仁，Sakriani Sakti，“漸進的な音声認識・機械翻訳・テキスト音声合成に基づく音声から音声への同時翻訳，”令和２年度日本音響学会大会 3-4-4 2020 言語処理学会論文賞，優秀発表賞等．

7．ホームページ等

https://ahcweb01.naist.jp/