



2020年度 グループ2 活動報告

研究開発項目 2：自在音声対話の研究開発

研究開発課題 1：自在遠隔音声対話の研究開発

実環境下における音声検出・認識

本年度は音声強調・検出と認識を統合した End-to-End 処理系の深層学習による最適化を検討するとともに、音声データベースの設計とベースラインモデルの構築を行った。音声強調については複数の手法を組み合わせる方法を検討し、音声認識との統合学習も実装した。

人間レベルの自律音声対話

相槌の生成や場をなごませるための共有笑いの生成について、入力から生成までの End-to-End 処理系の深層学習による最適化を検討するとともに、タスク設計とデータ収録環境の整備を行った。同時に複数人の傾聴を行うシステムにおいて、相槌生成を実装するとともに、共有笑いの生成も試みた。

自律対話と遠隔操作対話の切替え・融合

タスクとシステムの設計とデータ収録環境の整備を行った。また、同時に複数人の傾聴を行うシステムを設計し、プロトタイプを実装した。基本的な動作を確認するとともに、音声合成や対話制御などにおいて、技術的な課題を明らかにすることができた。

研究開発課題 2：音響情報処理・音声変換の研究開発

多様な利用者に対応できる自律 CA 用音声合成

自律 CA 用の対話音声合成の基礎検討を行った。そのために、これまで課題推進者が培ってきた音声合成技術を進化させ、十分な音声品質・個性再現度・応答速度を持つ音声合成技術の詳細設計に取り組んだ。具体的には、言語予測機能を持つ音声合成システムを実装し、その初期検討を行った。また、多様な音声個性を持つ、数百時間規模の音声コーパスの設計及び構築を進めた。

自律 CA 発話と遠隔操作発話を同化させる音声変換

本年度は、遠隔音声に対する音声変換技術の詳細設計に取り組み、同化のための基盤技術を構想した。来年度以降にその実装・評価を行う。

実環境下における音声分離・強調

本年度は基礎的な検討を行うとともに、CA と利用者の音響環境を模擬したデータベースの設計とベースラインモデルの構築を行った。具体的には、拡散性雑音中の音声抽出技術に関する初期検討を行い、また、大規模環境音データベースを購入・整備してその利用可能性を模索した。

研究開発課題 3 : 対話知識処理の研究開発

対話知識獲得

基礎的な検討を進めるための人間同士の対話データ収集を行った。傾聴、相談、接客という対話タスクを設定し、オペレータ役の二人の人間が一定時間ごとに入れ替わりながら、ユーザ役の二人の人間に対して対話サービスを行う対話を収集した。入れ替わる際は引継ぎのためのメモを交換することにした。対話データや引継ぎメモを分析することにより、CA が行うべき対話や引継ぎ方に対するの知見を得ることができる。本データ収集では 10 分程度の対話を 30 個収集することができた。対話データベースの構築指針として、収集された対話データをもとに対話シナリオを構築するという指針を立てた。

対話状況理解および可視化

対話状況理解および可視化におけるタスクとして、対話要約に着目することにした。タスク指向型対話および非タスク指向型対話を念頭に置き、操作者が途中から対話に参加できるようにするための対話要約の実現を目指す。この基礎検討のため、対話データ収集ではオペレータ役の引継ぎメモを収集し、可視化の方式を検討する準備を整えた。加えて、対話実験環境の整備として、CA を表示するためのディスプレイとユーザのマルチモーダル情報を取得するためのモーションキャプチャデバイスを準備した。

対話制御および制御インターフェース

対話制御および制御インターフェースのタスクとして、データに基づく対話制御の研究に加え、操作者から CA に対話を引き継ぐ際に必要となる入力方式の研究に取り組むこととした。対話データ収集で得られた引継ぎメモを参考にし、操作者の話した内容を CA に伝達できるようにするための方式を検討していく。

研究開発課題 4 : CG-CA 特有対話の研究開発

CG エージェントの認知の研究

音声対話用の CG キャラクターとして特に求められるデザイン要件を抽出し整理した。過去の対話性認知に関する研究や、映画等のコンテンツ業界におけるセオリー等を調査し、対話用 CG キャラクターに共通に求められる設計要件を「見た目・姿勢」「動作、表情」「反応性・アニメーション」「会話様式」に分類して各項目の洗い出しを行った。

また、一般人を対象としたインタラクションのデータ収集のための実験シナリオとして、受付案内、キャンパス案内、カウンセリング等について検討した。学内交渉の結果、名工大のキャンパス案内システム「正門メイちゃん」を利用したキャンパス案内実験を将来行うめどを立てた。また、ストリーミング送受信の処理基盤である Apache Kafka を用いた大規模データ収集の仕組みのプロトタイプを実装した。

また、キャラクターの対話への影響を調べるために、実験に用いる CG キャラクターの類型に関する検討を行った。CG エージェントの表現上の禁忌等についてユーザの特性と合わせた調査が必要である。今年はその調査のために、CG エージェントの表現の可能性を広くカバーする典型的なデザインセットについて考案した。来年度はこれに基づいて CG キャラクターのデザイン設計・発注を行う計画である。

CG エージェントの対話生成

来年度に予定しているプロの CG 操演俳優（いわゆる VTuber）による CG キャラクター操演対話のデータ収集の実施に向けて、対話タスクの検討およびデータベースの設計、収録内容に関する検討、機材の検証、システムの予備実装を行った。VTuber を幅広く視聴している学生を中心にチームを研究室内で立ち上げ、VTuber に関する基礎調査を行った。実際の収

録機材として、自動表情解析ツール OpenFace および全身モーショントラッキングデバイス Perception Neuron Pro の動作検証を行い、精度と遅延を検証するとともに、対話システム MMDAgent との接続を試行し、可用性の評価を行った。また、来年度からのデータ分析に向けて GPU マシンの調達・整備を行った。

CG-CA 対話システムの研究開発

遠隔操作対話のための操作者動作のリアルタイム伝送システムの設計と試作を行った。OpenFace と MMDAgent をソケット通信で繋ぐことで、オペレータの頭の動作、表情、口形を遠隔地の MMDAgent にリアルタイム伝送できることを確認した。また、遠隔操作時の音声とアバターの同期について、OpenFace の検出する Action Unit を口形に変換・伝送する方法、および音声認識（簡易な音素認識）を駆動して伝送音声から直接リアルタイムにリップシンクを行う方法を実装レベルで考案した。これらの開発のために、開発用 PC と Git サーバ用 PC を調達・整備し、システム開発を素早く加速して行うための分散開発環境の整備を行った。来年度以降、早期にプロトタイプシステムの実装を行い、研究開発グループ内で共有していく予定である。