

徳田 恵一

名古屋工業大学大学院工学研究科
教授

コンテンツ生成の循環系を軸とした次世代音声技術基盤の確立

§ 1. 研究実施体制

(1) 名古屋工業大学・徳田グループ

- ① 研究代表者: 徳田 恵一 (名古屋工業大学大学院工学研究科、教授)
- ② 研究項目
 - ・ 研究グループ間の連携調整
 - ・ 基盤技術および関連ソフトウェアの高度化
 - ・ ユーザによるコンテンツ生成循環の構築

(2) 名古屋工業大学・山本グループ

- ① 主たる共同研究者: 山本 大介 (名古屋工業大学大学院工学研究科、准教授)
- ② 研究項目
 - ・ 音声対話サービスのための情報基盤ネットワークの設計と活用
 - ・ 統一データベースに基づく音声対話コンテンツの動的作成
 - ・ Web インタフェースを用いた音声対話コンテンツの編集と共有

(3) エジンバラ大学・レナルズグループ

- ① 主たる共同研究者: Steve Renals (University of Edinburgh, School of Informatics, Professor)
- ② 研究項目
 - ・ 多言語化
 - ・ 対話コンテンツの自動インデキシング／リンキングおよびそれぞれに基づくユーザ分析

(4) 国立情報学研究所・山岸グループ

- ① 主たる共同研究者: 山岸 順一 (国立情報学研究所コンテンツ科学研究系、准教授)
- ② 研究項目
 - ・ 音声合成とユーザのフィードバックに関する基礎研究

§ 2. 研究実施の概要

本プロジェクトでは、魅力的な音声対話システムを作成するための音声基盤技術・基盤ソフトウェアの高度化、音声対話コンテンツを生成するための仕組み作り、実証実験に基づくコンテンツの分析と評価に取り組んでいる。本年度は、基盤技術・基盤ソフトウェアの高度化に取り組むと同時に、音声対話コンテンツ共有サービスの開発、音声対話コンテンツ生成環境の拡充に取り組んだ。さらに、公共空間における実証実験を実施した。

1) 基盤技術・基盤ソフトウェアの高度化

より魅力的な音声対話システムを構築するためには音声基盤技術や基盤ソフトウェアの高度化が必要である。本年度は、ディープニューラルネットワークに基づく音声合成による合成音声の品質改善や、リカレントニューラルネットワーク型言語モデルの高精度化による音声認識性能の改善等の音声言語情報処理の基礎研究を実施した。また、音声インタラクションシステム構築ツールキット MMDAgent をはじめとした関連ソフトウェア (HMM 音声合成ツールキット HTS、HMM 音声合成エンジン hts_engine API、日本語音声合成システム Open JTalk、歌声合成システム Sinsy、音声信号処理ツールキット SPTK) を改良し、新バージョンを公開した。また、MMDAgent を iOS へと対応させ、マルチプラットフォーム化を実現した。

2) 音声対話コンテンツを生成するための仕組みづくり

本年度は、音声対話コンテンツをユーザ間で共有可能にするための仕組みとして、音声対話コンテンツ共有サービスを立ち上げた(右図)。音声対話コンテンツ用ファイル形式の定義、インストーラの開発、オープン ID との連携、投稿時のエラー検出、利用規約の整備、著作権への配慮、マニュアルの整備等、音声対話コンテンツを容易に共有できるようにするために必要となる対応を行った。この他、利用頻度の高い対話や実用性を考慮した対話などの基本的な音声対話コンテンツをまとめた基本音声対話ライブラリの構築、MMDAgent で動作可能なライセンスフリー3D モデルの作成、モジュール型音声対話システムのアーキテクチャの構築、対話履歴を考慮した音声対話システムの構



築、クラウド方式の音声対話コンテンツの編集環境の構築、ネットワークエージェント技術を用いた音声対話システム間の連携基盤の開発に取り組んだ。

3) 実証実験に基づくコンテンツの分析と評価

前年度に引き続き、本年度も名古屋工業大学正門前、名古屋工業大学学生スペース、半田市観光協会、半田市市役所新庁舎において音声対話システムを継続的に運用し、音声対話コンテンツの収集・分析に取り組んだ。本年度はさらに、NHK名古屋放送局と共同で、TV番組やイベント等において実証実験を実施した。その際、複数のキャラクターを連携させて動作できるようにするなどの拡張を行った。さらに、国立情報学研究所キャラクター「情報犬ビットくん」による音声対話システムを構築した。このシステムは平成28年度NIIオープンハウス等にて展示する予定である。また、音声対話コンテンツ成立のためのユーザ誘引要素の分析、一問一答型音声対話システムの社会実験を行った。

4) エンサイクロペディア MMDAgent

本研究の成果物を統合・拡張することによりソフトウェア、ツール、設計指針に加えマニュアルや講習用スライド等を総合的・統一的にまとめたオールインワン・パッケージ「エンサイクロペディア MMDAgent」の作成を進めた。来年度正式に公開する予定である。

1. Kei Hashimoto, Keiichiro Oura, Yoshihiko Nankaku and Keiichi Tokuda, "Trajectory training considering global variance for speech synthesis based on neural networks," Proceedings of 2016 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2016), pp.5600-5604, Shanghai, China, March 20-25, 2016.
2. Siva Reddy Gangireddy, Steve Renals, Yoshihiko Nankaku and Akinobu Lee, "Prosodically enhanced recurrent neural network language models," Proceedings of INTERSPEECH 2015, Dresden, September 2015.
3. Keitaro Wakabayashi, Daisuke Yamamoto, Naohisa Takahashi, "A Voice Dialog Editor Based on Finite State Transducer Using Composite State for Tablet Devices," Computer and Information Science 2015, Studies in Computational Intelligence, Vol. 614, pp.125-139, 2016.