

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速化
 - ・ユーザによるコンテンツ生成環境の構築

(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, 教授)
- 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に基づくユーザ分析
 - ・ユーザが作成した対話コンテンツデータベースを利用した能動的対話システム

§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基盤ソフトウェアとして、音声対話システム構築ツールキット MMDAgent を開発[11]してきた。MMDAgent は、音声認識、音声合成、3D グラフィックス、対話制御などの機能を備え、ユーザによる対話の記述や拡張が可能な、40 万行のソースコードから成る、オープンソースソフトウェアである。本年度は、主に、モバイル対応化と音声基盤技術の改良、コンテンツ生成環境とネットワーク連携、多言語化とトピックモデルに基づく音声対話コンテンツの分析、および、公共空間における実証実験に取り組んだ。

1) モバイル対応化と音声基盤技術の改良

本年度は、MMDAgent の拡張と、各種音声モデルを改良することにより、新たに Android スマートフォンでの動作も可能にした MMDAgent 1.4 を公開した。また、各種オープンソースソフトウェア(音声認識エンジン Julius、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、音声信号処理ツールキット SPTK)の新バージョンも公開した。さらに、HMM 音声合成[12]のための特徴抽出とモデリングの統合手法[1]や、加算構造によるスペクトラムモデルの改良[3]等の音声処理の基礎研究も実施した。

2) コンテンツ生成環境とネットワーク連携

ユーザによる音声対話コンテンツ生成と利用を促進させるためには、ネットワーク連携可能な音声対話システムの実現が必須である。そこで、本年度は、モバイル版の MMDAgent を基盤として、音声対話システムとネットワーク(特にモバイル環境)の連携に関する研究を実施した。具体的には、スマートフォンのための音声対話開発基盤を整備すると同時に、マルチエージェント技術[8]に基づくネットワーク連携基盤、タブレットを用いた音声対話コンテンツ編集インタフェース、クラウド環境のための音声対話コンテンツ生成、MMDAgent と VoIP の連携、音声対話を用いた学内見学支援、関連ワード推薦などの研究に取り組んだ。

3) 多言語化とトピックモデルに基づく音声対話コンテンツの分析

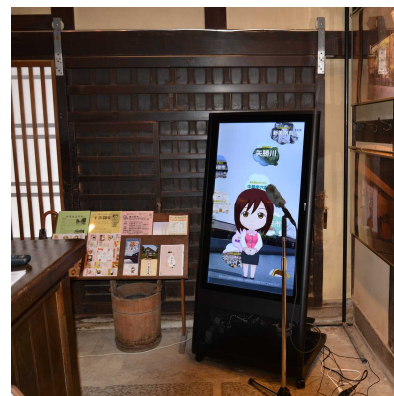
言語に依存しない普遍的な知見を得るためには、日本語のみならず他の言語による実証実験を並行して進める必要がある。そこで、エジンバラ大学のオープンスペースに音声対話システムを設置することを目的に、英語音声対話システムを構築し、さらに、対話調英語音声合成技術、ニューラルネットワークを用いた言語モデルの高精度化等、基盤技術の向上に取り組んでいる。また、音声対話コンテンツ群をインデキシング・リンキングすることを目標に、言語モデルの一種であるトピックモデルを応用し、FSTM というトピック事後確率ベクトルを計算する方法を導入した。また、本手法を用いて、音声対話コンテンツ内のトピックを確率的に推定し、音声対話コンテンツを分類、類似コンテンツを探索する実験にも挑戦した。

4) 公共空間における実証実験

本年度は、昨年度から引き続き、名古屋工業大学の正門前に双方向音声案内デジタルサイネ

ージ「メイちゃん」の運用を継続し、学生・教職員から多数の音声対話コンテンツの投稿があった。

また、新たに、ユーザ生成型音声対話コンテンツに関する実証実験システムを、愛知県半田市「蔵のまち観光案内所（国登録有形文化財小栗家住宅内）」に設置した。本実証実験では、ユーザ（ここでは、観光案内所の職員）が簡単かつ即座に音声対話内容等を追加・更新できる仕組みを実現することで、ユーザの創意工夫で、現場に即した、より面白い音声観光案内が実現可能になった。なお、本実証実験システムは、テレビや新聞などで紹介される、半田市の公式広報誌の表紙に掲載されるなど、観光 PR に繋がるものとして地域からも期待されている。



§ 3. 成果発表等

(3-1) 原著論文発表

論文詳細情報(国際)

1. Kazuhiro Nakamura, Kei Hashimoto, Yoshihiko Nankaku, and Keiiti Tokuda, “Integration of spectral feature extraction and modeling for HMM-based speech synthesis,” *IEICE Transactions on Information and Systems*, June 2014. (in press)
2. Rasmus Dall, Junichi Yamagishi, Simon King, “Rating naturalness in speech synthesis: The effect of style and expectation,” *Proceedings of Speech Prosody 7*, May 2014. (in press)
3. Shinji Takaki, Yoshihiko Nankaku and Keiichi Tokuda, “Spectral modeling with contextual additive structure for HMM-based speech synthesis,” *IEEE Journal of Selected Topics in Signal Processing*, vol.8, issue.2, pp.229-238, April 2014. (DOI:10.1109/JSTSP.2014.2305919)
4. Shinji Takaki, Kei Sawada, Kei Hashimoto, Keiichiro Oura, and Keiichi Tokuda, “Overview of NIT HMM-based speech synthesis system for Blizzard Challenge 2013,” *Blizzard Challenge 2013 Workshop*, Barcelona, Spain, September 3, 2013 (web proceedings).
5. Takenori Yoshimura, Kei Hashimoto, Keiichiro Oura, Yoshihiko Nankaku, and Keiichi Tokuda, “Cross-lingual speaker adaptation based on factor analysis using bilingual speech data for HMM-based speech synthesis,” *8th ISCA Speech Synthesis Workshop*, pp.317-322, Lyon, France, August 31-September 2, 2013.
6. N. d'Alessandro, J. Tilmanne, M. Astrinaki, T. Hueber, R. Dall, T. Ravet, A. Moinet, H. Cakmak, O. Babacan, A. Barbulescu, V. Parfait, V. Huguenin, E. S. Kalayci, and

- Q. Hu, "Reactive statistical mapping: Towards the sketching of performative control with data," IFIP Advances in Information and Communication Technology, Special Issue: Proceedings of the 9th eINTERFACE Workshop, August 2013.
7. Motohiro Shibakawa, Takahiro Uchiya, Ichi Takumi, Tetsuo Kinoshita, "Design and implementation of multiagent-based distributed backup system," Proceedings of the 12th IEEE/ACIS International Conference on Computer and Information Science, pp.234-240, June 2013. (DOI: 10.1109/ICIS.2013.6607847)
 8. Takahiro Uchiya, Syo Itazuro, Ichi Takumi, Tetsuo Kinoshita, "Design environment of intelligent multiagent systems," Proceedings of the 12th IEEE/ACIS International Conference on Computer and Information Science, pp.241-246, June 2013. (DOI: 10.1109/ICIS.2013.6607848)
 9. Shinji Takaki, Yoshihiko Nankaku, and Keiichi Tokuda, "Contextual partial additive structure for HMM-based speech synthesis," Proceedings of ICASSP 2013, pp.7878-7882, Vancouver, Canada, May 26-31, 2013. (DOI:10.1109/ICASSP.2013.6639198)
 10. Kazuhiro Nakamura, Kei Hashimoto, Yoshihiko Nankaku, and Keiichi Tokuda, "Integration of acoustic modeling and mel-cepstral analysis for HMM-based speech synthesis," Proceedings of ICASSP 2013, pp.7883-7887, Vancouver, Canada, May 26-31, 2013. (DOI:10.1109/ICASSP.2013.6639199)
 11. Akinobu Lee, Keiichiro Oura, and Keiichi Tokuda, "MMDAgent -a fully open-source toolkit for voice interaction systems-," Proceedings of ICASSP 2013, pp.8382-8385, Vancouver, Canada, May 26-31, 2013. (DOI:10.1109/ICASSP.2013.6639300)
 12. Keiichi Tokuda, Yoshihiko Nankaku, Tomoki Toda, Heiga Zen, Junichi Yamagishi, and Keiichiro Oura, "Speech synthesis based on hidden Markov models," Proceedings of the IEEE, vol.101, no.5, pp.1234-1252, May 2013. (招待論文) (DOI:10.1109/JPROC.2013.2251852)