

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(研究機関別)

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(研究機関別)

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(研究機関別)

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. 研究実施内容

(文中に番号がある場合は(3-1)に対応する)

ユーザによる音声対話コンテンツ生成という新しい概念を導入し、それが実際に機能するための仕組みや条件を実証的に探究する。音声対話システム全体をシステム自体とコンテンツに分離して考え、1)「魅力的」な音声対話を成立させるためのシステムの要件と、

2)コンテンツクリエイター引いてはユーザによって「魅力的」な音声対話コンテンツが次々と生成されるための条件を解明し、同時にそのための仕組みを確立する。コンテンツ生成の循環系(図1)が成立するための様々な知見を実証的に得るためには、システムを公共空間に常設し運用するとともに、インターネットを介したネットワーク連携の仕組みを活用しながら、ユーザがそのコンテンツを生成・登録・改善できる環境を構築しなければならない。

以下、設定した課題に関してそれぞれの達成状況及び進捗状況を説明する。

【課題1：基盤技術と関連ソフトウェアの高度化】

「魅力的」なコンテンツを生成するためには、システムがユーザにとって「魅力的」であり得る技術基盤をもたなければならない。このため、克服しなければならない個別の課題が数多くある。本年度は、音声対話システムをモバイル環境で利用可能にするために、以下に述べる二つの方式を開発した。一つは、図2左で示すように、MMDAgentをAndroid OSに移植し、スマートフォン単体で動作可能なモバイル音声対話システムを開発した。もう一つは、昨年度に引き続き、図2右に示すように、MMDAgentとビデオ通話機能(Skype)を連携させることにより、ビデオ通話によるモバイル音声対話を可能にした。前者は応答遅延の少ないより自然な音声対話が可能になり、後者は、追加のソフトウェアなどをインストールする必要なく手軽にモバイル音声対話が可能になった。

また、合成音声の更なる高品質化^{9), 10)}、多様な感情を表す音声合成^{15), 17)}等の音声対話システムに必要な基盤技術の改善に取り組むと共に、研究基盤ソフトウェアについても開発を進め、新バージョンを公開した。

1. 音声対話システム MMDAgent-1.3.1 (<http://www.mmdagent.jp/>)⁸⁾
2. 音声信号処理ツールキット SPTK-3.6 (<http://sp-tk.sourceforge.net/>)



図1 コンテンツ生成の循環系



図2 モバイル環境における音声対話システム

3. 音声認識エンジン Julius-4.2.2 (<http://julius.sourceforge.jp/>)
4. HMM 音声合成システム HTS-2.3 alpha (<http://hts.sp.nitech.ac.jp/>)¹⁾
5. ランタイム音声合成エンジン hts_engine API-1.07 (<http://hts-engine.sourceforge.net/>)
6. 日本語テキスト音声合成システム Open JTalk-1.06 (<http://open-jtalk.sourceforge.net/>)
7. 英語テキスト音声合成システム Flite+hts_engine-1.04 (<http://hts-engine.sourceforge.net/>)

【課題 2 : ユーザによるコンテンツ生成環境の構築】



図 3 サイエンスアゴラにおいて利用した音声対話入力インタフェースの例

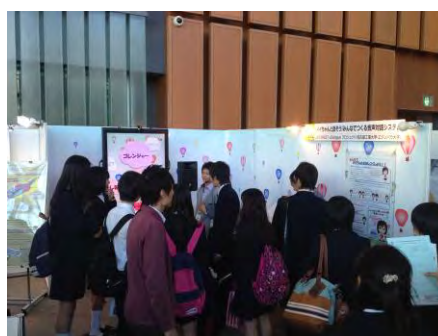


図 4 サイエンスアゴラにおける展示の様子

Web インタフェースを用いたより魅力的な音声対話を構築できる枠組みについて検討した。具体的には、従来からある認識キーワードとそれに対応する音声案内文だけでなく、声質やキャラクターの表情・モーションなども追加可能な Web インタフェースを試作した。サイエンスアゴラ 2012 では図 3 に示すデモシステムを実際に使用し、モーションや声質の組み合わせの変化が音声対話の魅力に大きく影響することを実感した(図 4)。

【課題 3 : 多言語化】

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。この際には言語的な違いだけでなく、文化的な違いに関しても十分考慮する必要がある。

今年度は、エジンバラ大学の Informatics Forum 内のオープンスペース (図 5) に英語音声対話システムを設置することを念頭に置きながら、エジンバラ大で開発された英語音声認識システムおよび音声合成システムの統合を行った。来年度も継続して、対話エージェントの開発も含めた 英語音声対話システムの開発を行う。対話エージェントに関してはデザインのみでなく、英国の文化に適したジェスチャーなどについても考慮していく。



図 5 Informatics Forum

【課題4：自動インデキシングとリンクング】

ウェブ上のユーザにより作成される大量の対話コンテンツ群は、音声対話システムにとっての集合知である。今年度は、情報検索および言語モデルを専門とするポスドク研究員を採用し、これらの対話コンテンツ群を活かした新たな研究を開始した。具体的には、大量にある対話コンテンツ群の中から、自動で適切な対話コンテンツを探し出せるよういくつかの要素技術の検討を開始した。

【課題5：魅力の評価と信頼性の確保】

本学の正門前に設置済みの「双方向音声案内デジタルサイネージシステム⁶⁾」に関する使い勝手や評判について、全学の学生・教員・職員を対象とした詳細なアンケートを実施し、改善点について議論した。

【実証実験フェイズ1】

前年度に引き続き、本学の正門前に設置済みの「双方向音声案内デジタルサイネージ⁶⁾」を利用して、実証実験フェイズ1を実施した。ユーザ生成型コンテンツの投稿数は、前年度(209件)よりも今年度(249件)は増加し、特に、学生による投稿数が14件から33件と倍増した。また、本学のオープンキャンパスにおいてモバイル音声対話に関する公開実験を実施した。

§3. 成果発表等

(3-1) 原著論文発表

● 論文詳細情報

1. Keiichi Tokuda, Yoshihiko Nankaku, Tomoki Toda, Heiga Zen, Junichi Yamagishi, Keiichiro Oura, "Speech synthesis based on hidden Markov models", Proceedings of IEEE, 2013. (in press)
2. Syo Itazuro, Takahiro Uchiya, Tetsuo Kinoshita, Ichi Takumi, "Development support of learning agent on repository-based agent framework", International Journal of Software Science and Computational Intelligence, 2013. (accepted)
3. Wenpeng Wei, Hideyuki Takahashi, Takahiro Uchiya, Tetsuo Kinoshita, "Cooperation protocol design method for repository-based multi-agent applications", International Journal of Cognitive Informatics & Natural Intelligence, 2013. (accepted)
4. Pablo Martinez Lerin, Daisuke Yamamoto, Naohisa Takahashi, "Encoding network-constrained travel trajectories using routing algorithms", International Journal of Knowledge and Web Intelligence, Vol.4, No.1, pp. 34-49, 2013. (DOI: 10.1504/IJKWI.2013.052724)
5. Hiroya Mizutani, Daisuke Yamamoto, Naohisa Takahashi, "A preventing method for overlapping focuses in a Focus+Glue+Context map", International Journal of

- Knowledge and Web Intelligence, Vol.4, No.1, pp. 50-69, 2013. (DOI: 10.1504/IJKWI.2013.052725)
6. 大浦圭一郎, 山本大介, 内匠逸, 李晃伸, 徳田恵一, “キャンパスの公共空間におけるユーザ参加型双方向音声案内デジタルサイネージシステム”, 人工知能学会誌, Vol.28, No.1, pp.60—67, 2013 年 1 月.
 7. Pablo Martinez Lerin, Daisuke Yamamoto, Naohisa Takahashi, "Mouse-sensitive following path suggestion for drawing travel routes in web map systems", Journal of Geographic Information System, Vol.4, No.5, pp. 393-402, October 2012. (DOI: 10.4236/jgis.2012.45045)
 8. Akinobu Lee, Keiichiro Oura, Keiichi Tokuda, “MMDAgent — A FULLY OPEN-SOURCE TOOLKIT FOR VOICE INTERACTION SYSTEMS—,” 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013), Vancouver, Canada, May 26–31. (accepted)
 9. Shinji Takaki, Yoshihiko Nankaku, Keiichi Tokuda, “Contextual partial additive structure for HMM-based speech synthesis,” 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013), Vancouver, Canada, May 26–31. (accepted)
 10. Kazuhiro Nakamura, Kei Hashimoto, Yoshihiko Nankaku, Keiichi Tokuda, “Integration of acoustic modeling and mel-cepstral analysis for HMM-based speech synthesis,” 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013), Vancouver, Canada, May 26–31. (accepted)
 11. Takahiro Uchiya, Ichi Takumi, Tetsuo Kinoshita, "Circulation mechanism of multiagent system based on agent repository", Proc. of IEEE BWCCA-2012, pp.668-672, December 2012.
 12. Naoya Tatematsu, Takahiro Uchiya, Ichi Takumi, Tetsuo Kinoshita, "Enhancement of repository based agent framework for ubiquitous environment", Proc. of IEEE BWCCA-2012, pp.673-678, December 2012.
 13. Syo Itazuro, Takahiro Uchiya, Ichi Takumi, Tetsuo Kinoshita, "Design Environment of reinforcement learning agents for intelligent multiagent system", Proc. of IEEE BWCCA-2012, pp.679-683, December 2012.
 14. Takahiro Uchiya, Kinoshita, "Enhancement of agent design environment for agent interoperability", Proc. of IW-STEIC2012, pp.89-94, October 2012.
 15. Shinji Takaki, Kei Sawada, Kei Hashimoto, Keiichiro Oura, Keiichi Tokuda, “Overview of NIT HMM-based speech synthesis system for Blizzard Challenge 2012,” Blizzard Challenge 2012 Workshop, Portland, USA, September 14, 2012 (web proceedings).

16. Takafumi Hattori, Kei Hashimoto, Yoshihiko Nankaku, Keiichi Tokuda, "A Bayesian approach to speaker recognition based on GMMs using multiple model structures," Interspeech 2012, Portland, USA, September 9–13, 2012.
17. Viviane de Franca Oliveira, Sayaka Shiota, Yoshihiko Nankaku, Keiichi Tokuda, "Crosslingual speaker adaptation for HMM-based speech synthesis based on perceptual characteristics and speaker interpolation," Interspeech 2012, Portland, USA, September 9–13, 2012.
18. Syo Itazuro, Takahiro Uchiya, Ichi Takumi, Tetsuo Kinoshita, "Design support mechanism of learning agent on repository-based agent framework", Proc. of ICCI*CC2012, pp.300-304, August 2012.
19. Wenpeng Wei, Hideyuki Takahashi, Takahiro Uchiya and Tetsuo Kinoshita, "Repository-based methodology of cooperation protocol design for multi-agent system", Proc. of ICCI*CC2012, pp.283-288, August 2012.
20. Pablo Martinez Lerin, Daisuke Yamamoto and Naohisa Takahashi, "Encoding travel traces by using road networks and routing algorithms", Proc. of KES IIMSS 2012, Gifu, pp.233-243, May 2012.
21. Hiroya Mizutani, Daisuke Yamamoto and Naohisa Takahashi, "A fusion of multiple focuses on a Focus + Glue + Context map fusion", Proc. of KES IIMSS 2012, Gifu, pp.11-21, May 2012.
22. Pablo Martinez Lerin, Daisuke Yamamoto and Naohisa Takahashi, "Making a pictorial and verbal travel trace from a GPS trace", Proc. of W2GIS 2012, Naples, Italy, pp.98-115, April 2012.