

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 (名古屋工業大学大学院工学研究科、助教)

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. 研究実施内容

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うるシステムの要件と、

2)コンテンツクリエイタひ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、音声区間検出、赤外線センサー・レーザーレンジファインダ・CCD カメラ等の情報と音声情報によるマルチモーダル情報処理、合成音声の更なる高品質化、多様な感情を表す音声合成、合成音声と表示キャラクターとの唇及び動作の同期、ネットワークを用いた対話情報の送受信等の課題に取り組んだ。また、研究基盤ソフトウェアに関して、以下の通り開発を進め、新バージョンを公開した。

1. 音声信号処理ツールキット SPTK-3.5 (<http://sp-tk.sourceforge.net/>)
2. 音声認識エンジン Julius-4.2.1 (<http://julius.sourceforge.jp/>)
3. ランタイム音声合成エンジン hts\_engine API-1.06 (<http://hts-engine.sourceforge.net/>)  
本ソフトウェアの一部は NTTドコモの携帯電話 20 機種以上に採用された。
4. 日本語テキスト音声合成システム Open JTalk-1.05  
(<http://open-jtalk.sourceforge.net/>)
5. 英語テキスト音声合成システム Flite+hts\_engine-1.03  
(<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】

ユーザ生成型音声対話コンテンツ作成環境に関する研究では、音声対話コンテンツを作成するための基本的な Web インタフェースを開発し、さらに、キャンパス内に設置した双方向音声案内デジタルサイネージ「正門メイちゃん」(図2)と連携するための FST スクリプトの自動生成機能を実

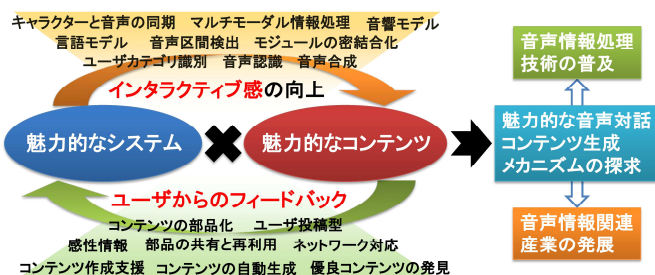


図1 コンテンツ生成の循環系

現した(図3)。これらの機能を、名古屋工業大学情報基盤センターが提供しているイベント共有システムである「イベントカレンダー」と連携させることによって、学生や教職員が実際に利用可能な環境を構築した。実際に、学



図2 双方向音声案内  
デジタルサイネージ

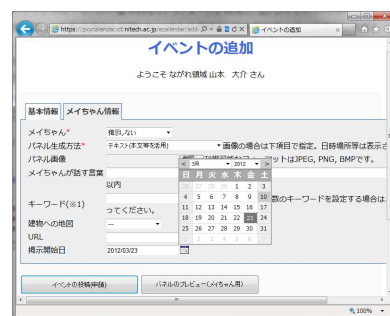


図3 音声対話コンテンツ  
作成環境

生や教職員による多数のユーザによる音声対話コンテンツの生成が実現され、従来の音声対話を用いない旧イベントカレンダーに比べて約2倍の数のイベントを学内で共有可能になった(図4)。

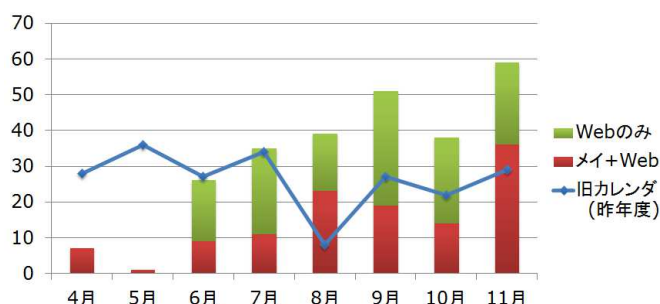


図4 ユーザによる音声対話コンテンツ登録数

また、モバイル環境における音声対話システムとユーザ生成型音声対話コンテンツ作成環境に関する研究を実施した。モバイル環境における音声対話システムに関連

する研究では、iPhoneなどのスマートフォンを用いてCGキャラクターと音声による対話を可能にするプロトタイプシステムである「モバイルメイちゃん」を開発した(図5)。本システムは、Skypeのビデオ電話機能とMMDAgentを連携させることによって実現している。さらに、第74回情報処理学会全国大会において公開実験とアンケート調査を実施し、121名分のアンケート結果を回収した(図6)。



図5 モバイル環境音声対話システム



図6 公開実験の様子

このように、音声対話システムに関するプロトタイプシステムの開発や公開実験等を実施することによって、魅力的な音声対話コンテンツを実現するための課題やノウハウ等の知見を得ることができた。

### 【課題3: 多言語化】

対話コンテンツの多言語化は、言語に依存しない普遍的な知見を得るための必要不可欠な実証実験であり、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。

2012年3半期に「双方向音声案内デジタルサイネージ」をエジンバラ大学 Informatics Forum(図7)に設置予定であり、これに向けて、英語版の音声合成、音声認識、対話システム、対話エージェント等のモジュール群を、MMDAgentと統合する作業を開始した。

エジンバラ大学音声技術研究所において開発されている英語音声認識および合成システムを、MMDAgentに組み込むためには、まず、音響モデルに使用する音声データのライセンス上の問題を解決する必要がある。そこで今年度は、イギリス英語大規模音声データ(60名、合計60時間分)を収録し、それらを音声認識/合成用コーパスとして構築した。また、このコーパスから音声認識システム/音声合成システムを構築した。このコーパスは、MMDAgentと同一のModified BSDライセンスで無償配布する予定であり、イギリス英語研究用音声データベースとして今後様々な分野において種々多様な用途に利用されることが期待される。

なお、【課題4:自動インデキシングとリンクング】および【課題5:魅力の評価と信頼性の確保】に関しては、次年度以降の開始予定である。



図7 Informatics Forum