

## 音声情報案内システム「たけまるくん」の運用

鹿野清宏<sup>†1</sup>, Tobias Cincarek<sup>†1</sup>, 竹内翔大<sup>†1</sup>, 川波弘道<sup>†1</sup>, 李晃伸<sup>†2</sup>, 西村竜一<sup>†3</sup>, 猿渡洋<sup>†1</sup>

奈良先端科学技術大学院大学情報科学研究科<sup>†1</sup>,  
名古屋工業大学情報工学科<sup>†2</sup>, 和歌山大学システム工学部<sup>†3</sup>

### 1. はじめに

音声認識技術は、過去 30 年余りの研究開発努力によっても、まだ、広く利用されている状況になっているとは言えない。一方、IT の普及により、IT を利用できる人とできない人の格差が広がり、いわゆる「デジタルデバイド」の問題が社会問題となってきている。著者らは、音声認識技術を IT 利用のインフラとしてとらえ、10 数年余り、フリーソフトウェアの観点から、音声認識ソフトウェアの研究開発を行ってきた [1, 2, 3, 4].

2003.4 から文部省リーディングプロジェクト e-Society 基盤ソフトウェアの総合開発の一環として、「ユーザ負担のない話者・環境適応性を実現する自然な音声対話処理技術」で、IT 利用のインフラとしての音声認識・合成のフリーソフトウェアの開発を行っている [4]. ソフトウェアのマイコン、DSP への実装も目指して、企業とも連携をとって研究開発を行っている。誰でも容易にネットワーク情報にアクセスできるための音声認識・合成のフリーソフトウェアおよび廉価なデバイスの開発を目指している。このプロジェクトの研究開発は、2003 年度からの 5 年である。研究開発は、(a) 教師なし話者・環境適応、(b) 大語彙連続音声認識、(c) ハンズフリー音声認識、(d) 実システム構築・利用ガイドライン、(e) 多様な音声合成、の 5 項目からなる。システム開発の概念図を図 1 に示す。

このプロジェクトでは、実環境で使える音声対話システムの実現を中心に備えてきた。本稿では、大語彙連続音声認識プログラム Julius を利用して、5 年以上に渡って運用している音声情報案内システム「たけまるくん」について、運用状況に重点をおいて紹介する。

### 2. 音声情報案内システム「たけまるくん」

「たけまるくん」システム [5] は、この 5 年余り、生駒市の北コミュニティセンターに常設され、種々の情報案内サービスを行っている。当初は、300 発話/日であったが、種々の改良により 600 発話/日まで、利用頻度があがっている。

Speech Guidance System “Takemaru-kun”  
K.Shikano, T.Cincarek, S.Takeuchi, H.Kawanami  
H.Saruwatari: Nara Institute of Science and Technology.  
A.Lee: Nagoya Institute of Technology.  
R.Nisimura: Wakayama University

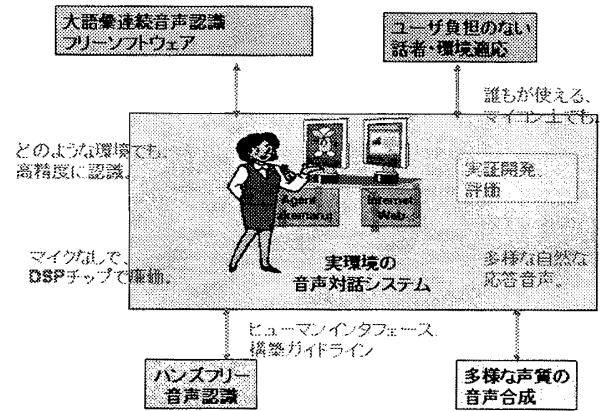


図 1 : 文部科学省 e-Society 「人にやさしい音声認識・合成」プロジェクトの概要

### 2. 1 「たけまるくん」のシステム構成

システムの構成を図 2 に示す。大語彙連続音声認識プログラム Julius [3] の改良版を用いている。語彙数は、4 万語で、大人と子供のそれぞれの音韻モデルと言語モデルを利用した並列デコーディングである。さらに、笑い声、背景雑音などを発声長や雑音 GMM で検出している。音声応答は、一問一答型で、対話制御は行っていない。大量(約 1 万)の質問文例をもっており、各質問文例は、応答文と対応する Web アドレスをもっている。応答文の種類は、約 300 で、大人と子供別に用意している。Julius の認識結果から単語の確信度を計算して、語順を無視して、最も近い質問文例を見つけ出し、対応する応答音声を TTS で合成し、かつ、Web の表示を行う。

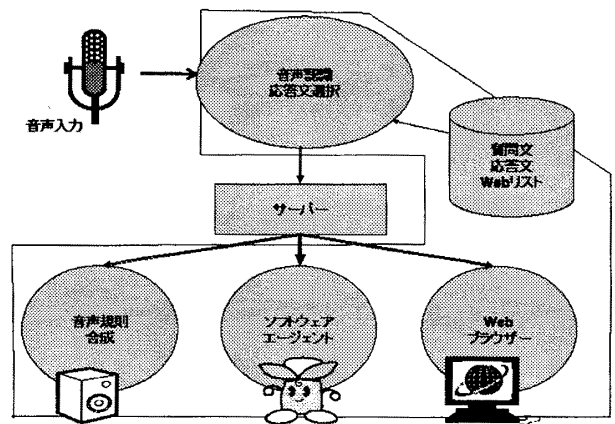


図 2 : 公共機関に設置されている音声情報案内システム「たけまるくん」の構成

## 2.2 「たけまるくん」の運用状況

「たけまるくん」は、2002年11月に稼働を開始した。稼働状況を図3に示す。2004年3月末に、GMMによる笑い声・雑音の識別、大人と子供の並列デコーディングによる改良を行った。2004年4月以降の利用頻度が約2倍に向上していることが分かる。現在の応答正解率は70%程度と推定される。誤った応答もニヤミスが多く、完全に間違えた場合でも、一問一答であるので、システムへの影響は少ない。システムは、非常に安定に動作している。システムの監視として、プロセス数などのチェックをネットワーク経由で10分おきに行っており、異常がある場合には、研究室のスタッフにメールが入る。スタッフは、ネットワーク経由でリブートなどの対処がどこからでもできる。また、日々の発声数（子供、大人）、雑音の種類数（笑い声、背景雑音など）も自動集計しており、日々の利用状況を提供している。現在までに、約2年分のすべての入力データに、発話内容の書き起こしと、雑音の種類などのラベルと応答文が付与されている。

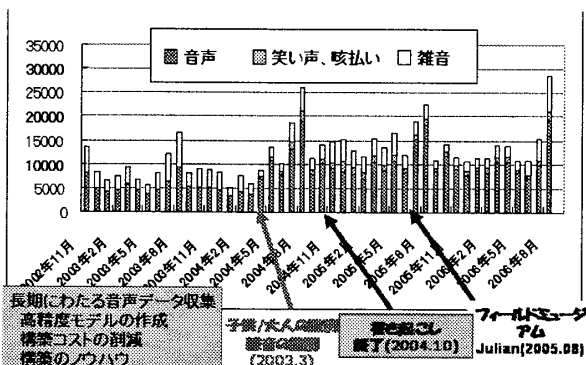


図3：音声情報案内システム「たけまるくん」の月あたりの入力数

## 2.3 波及効果

「たけまるくん」は、生駒市のキャラクターであり、子供に好かれており、大量の子供と幼児の音声データが収録できた。当初の子供の単語認識率は60%程度であったが、現在では84%程度までに向上している。また、幼児の認識は、できなかったが、現在では60%程度にまで向上している(表1)。この2年間分の音声データベースを活用して、ビジネスモデルの観点からの音韻モデルの教師なし学習や既存の音声データベースの活用の方策も検討してきた[6, 7, 8]。

たけまるくんへの発話の内容についても調査を行い、表2に示すように、たけまるくん自身への問いかけや挨拶も多く、おしゃべりが音声対話システムでも重要であることが伺える。

表1：収録音声データによる単語認識率の向上

大人		
2年間分の書き起こし収録データを利用		
データベース	JNAS (新聞記事読み上げ音声)	たけまるくん (対話音声)
音声データ量	14,809	14,800
単語認識率	84.9%	88.5%

子ども			
データベース	CSRC (単語発声)		たけまるくん (対話音声)
音声データ量	40,000		78,626
単語認識率	子ども	幼児	子ども
	78.8%	41.3%	84.2%
			60.8%

表2：発話内容の分析

2006. 8. 1 から 8. 20 までの入力				
ドメイン	大人		子供	
汎用的(観光など)	279	24%	1,604	15%
たけまるくん会話	415	36%	5,132	47%
ウェブ検索、HP	35	3%	122	1%
天気・ニュース	65	6%	505	5%
日付・時間	46	4%	265	2%
生駒市関係	106	9%	966	7%
北コミ関係	102	9%	709	9%
Field Museum	115	10%	1,633	15%
合計	1,163	100%	10,936	100%

## 3. むすび

Julius を中心に構成されている「たけまるくん」システムは、他の情報案内システムへの移植も可能で、企業も含めて数機関の情報案内システムに移植されている。

たけまるくんの音声データベースも公開する予定である。

## 参考文献

- [1] 鹿野, 伊藤, 河原, 武田, 山本 編著, “音声認識システム”, オーム社, 2001-05
- [2] A.Lee, T.Kawahara, K.Shikano, “Julius--- An Open Source Real-Time Large Vocabulary Recognition Engine”, Proceedings of 7th European Conference on Speech Communication and Technology, pp.1691-1694, September 2001
- [3] オープンソースの高性能汎用大語彙連続音声認識 Julius : <http://julius.sourceforge.jp/>
- [4] 鹿野, 河原, 河原, 武田, 猿渡, “人にやさしい音声認識・合成基盤ソフトウェアの研究開発”, 情報処理学会研究報告, 2003-SLP-49-55, 2003-12
- [5] 西村, 西原, 鶴身, 李, 猿渡, 鹿野, “実環境研究プラットフォームとしての音声情報案内システムの運用”, 電子情報通信学会論文誌, Vol.J87-D-II, No.3, pp.789-798, 2004-03
- [6] T.Cincarek, T.Toda, H.Saruwatari, K.Shikano, “Utterance-based Selective Training for the Automatic Creation of Task-Dependent Acoustic Models,” IEICE Trans. Information and Systems, Vol.E89-D, No.3, pp.962-969, 2006
- [7] T.Cincarek, T.Toda, H.Saruwatari, K.Shikano, “Cost Reduction of Acoustic Modeling for Real-Environment Applications Using Unsupervised and Selective Training”, IEICE Trans. Information and Systems, March 2008
- [8] T.Cincarek, H.Kawanami, R.Nishimura, A.Lee, H.Saruwatari, K.Shikano, “Development, Long-Term Operation and Portability of a Real-Environment Speech-oriented Guidance System”, IEICE Trans. Information and Systems, March 2008