2006－SLP－63（7）
2006／10／20

音声情報案内システム「たけまるくん」および「キタちゃん」の開発

鹿野　清宏[†]　Cincarek Tobias[†]　川波　弘道[†]　西村　竜一[††]　李　晃伸[†††]

[†]奈良先端科学技術大学院大学　情報科学研究科　〒630-0192 奈良県生駒市高山町 8916-5
（shikano, cincar-t, kawanami）@is.naist.jp

[††] 和歌山大学　システム工学部, [†††]名古屋工業大学 大学院工学研究科

あらまし　筆者らは生駒市北コミュニティセンターに、音声情報案内システム「たけまるくん」を設置して、2002 年 11 月から４年間運用している。最初の１年半あまり、精力的にシステムの改良を行い、現在では、子供を中心とする多くの市民に利用されている。このシステムは、大語彙連続音声認識プログラム Julius を用いた４万語あまりの大語彙の連続発声認識を中心に構築された本格的な自由発話による音声情報案内システムである。また、入力された音声や雑音はすべて収録され、とくに最初の２年間は書き起こしが終了している。この２年間の書き起こしデータを用いることによるシステムの性能の向上の予備評価についても報告する。

このたけまるくんの成果を活かして、今年の３月末に、奈良先端大の近くの近鉄の駅「学研北生駒」に、独立した２つの音声情報案内システム「キタちゃん」と「キタロボ」を設置した。駅は 60dBA と、コミュニティセンターに比べて、騒音レベルが 10dB 程度高く、厳しい音声認識の利用条件である。「キタちゃん」は、たけまるくんと同様に CG エージェントが応答する型で、タッチパネルも併用できる大人向けのシステムである。「キタロボ」は、ロボット型インタフェースで、どちらかというと子供向きのシステムである。この両システムの運用も６ヶ月間になるが、良好に動作している。たけまるくんからこの両システムへのポータビリティについても述べる。

キーワード　音声情報案内システム，大語彙連続音声認識，書き起こし音声データベース，ポータビリティ，応答生成，実環境システム，環境雑音認識，一問一答知識ベース

Development and Evaluation of Takemaru-kun Spoken Guidance System and Portability to Kita-chan and Kita-robo Systems

Kiyohiro Shikano[†]　Tobias Cincarek[†]　Hiromichi KAWANAMI[†]

Ryuichi Nisimura[††]　Akinobu Lee[†††]

[†]Graduate School of Information Science, Nara Institute of Science and Technology,
[††]Wakayama University　[†††]Nagoya Institute of Technology

Abstract　We have been developing and operating “Takemaru-kun” spoken information guidance system in North Community Center in Ikoma city these four years. Takemaru-kun, which is composed of large vocabulary continuous speech recognition program Julius and Q-A database, is now widely used by Ikoma citizens, mainly children. All inputs have been recorded and the first two-year data are annotated. Takemaru-kun system improvement based on two year annotated data is also reported.

Takemaru-kun was successfully ported to two spoken information guidance systems in Gakken North Ikoma railway station in the end of this March. These two systems are CG agent type “Kita-chan” , and robot type “Kita-robo”. The portability of acoustic models from Takemaru-kun to Kita-robo is also discussed.

Keyword　Spoken information guidance system, Large vocabulary continuous speech recognition, Annotated speech data, System portability, Response generation, Question-response database

１．はじめに

音声認識の研究が３０年あまり精力的に行われているにもかかわらず、音声対話システムはほとんど利用されていない。筆者らは、コストによる制約の少ない大学のメリットを活かして、エージェント型の音声情報案内システム「たけまるくん」を、2002 年１１月に生駒市北コミュニティセンターに設置して、４年間運用している[2,3,4,5,6,7]。当初の予想を上回って、１日平均 600 入力と、市民とくに子供に利用されている。

さらに、今年の３月に近鉄学研北生駒駅に、エージェント型の「キタちゃん」とロボット型の「キタロボ」を設置して、運用を開始した[16]。

これらは、大語彙連続音声認識のメリットを生かした自由発話による情報案内システムである。システムの作成者だけでなく、ユーザ、導入者の三者が満足できるシステムを目指している。幼児から高齢者まで、気楽に利用されることを目指し、タスクはユーザが決めるとの立場をとっている。ユーザの発話に応じて、タスク領域が柔軟に対応できるように、一問一答型の戦略[1]を主体としている。これらの音声対話システムの開発では、システムを作成して運用するだけでなく、大学としてのビジネスモデルも意識している。

２．音声情報案内システム「たけまるくん」と「キタちゃん&キタロボ」

たけまるくんは、大語彙連続音声認識プログラム Julius [8,9,10,11,12]を主体とした CG エージェント型の音声情報案内システムである。このたけまるくんの外観を図１に示す。システム構成を図２に示す。たけまるくんは、二つの画面をもち、左には生駒市のキャラクターのたけまるくんを表示し、右の画面には、Web などの情報案内の画面を表示する。よって、応答は、ディスプレイと併用して、比較的短い規則合成の応答音声によって行われる。知識ベースは、約 10,000 文の質問文データベースとして蓄えられ、それぞれの質問文には、応答文、Web アドレスがつけられている。１０位までの認識結果により形態素単位での語順を無視したスコアリングにより質問文が選択され、リンクされた応答文を出力する。

図 1. たけまるくんの外観

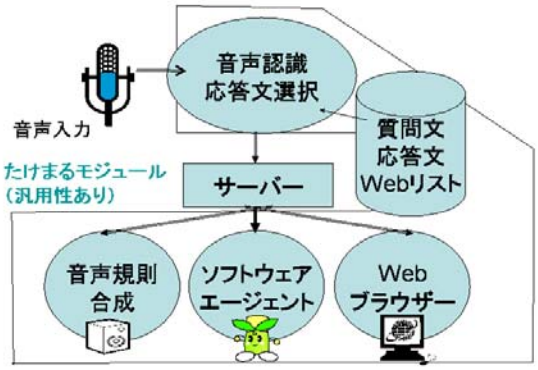

図 2. たけまるくんの構成

キタちゃんおよびキタロボは、たけまるくんのモジュールをもとに作成され、騒音レベルの高い近鉄北生駒駅に設置された。キタちゃんは、たけまるくんと同様に、CG エージェント型で、マウスに替えて、タッチパネルが使用できる。キタロボは、ロボットタイプで、タッチパネルは利用できない。両者とも、たけまるくんをもとに作成した一般的な質問に加え、駅周辺の案内、駅の探索などが、Julian ベースのネットワーク文法で追加されている。両者とも、大きな液晶ディスプレイを備えている。キタちゃんの外観を図３に、キタロボの外観を図４に示す。

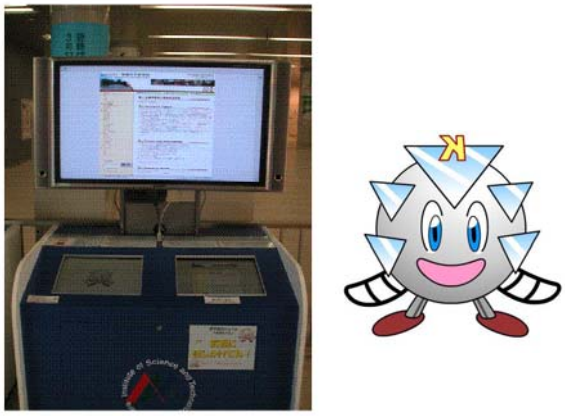

図 3. キタちゃんの外観

図 4. キタロボの外観

キタちゃんは、タッチパネルディスプレーの位置が高く、小学校高学年から大人向けのシステムである。キタロボはマイク位置が低くなっており、幼児、小学生向けになっている。

たけまるくんは、インターネット接続されており、１０分おきにプロセス数、使用メモリー量、入力数などが監視システムによりチェックされ、異常の可能性がある場合には、管理者にメールが届くようになっている。システムのリブートやチェックなどのシステムメンテナンスは、ネットワークを介して遠隔地からもできる。北生駒駅は、大学のネットワークが拡張されており、キタちゃんとキタロボも、同様の監視システムを備えている。Julius を含めてたけまるパッケージの信頼性は高く、トラブルは非常に少ない。

３．運用状況

ここでは、たけまるくんの４年間の改善に伴う運用状況を述べる。各システムへの入力の種類、サービスタスクなどについても述べる。

３．１．たけまるくん

たけまるくんは、当初一日当たり 250 発話程度の入力数であったのが、雑音や笑い声の識別、大人と子供別の応答などの改善にともない、ここ２年間は、一日当たり約 600 発話になっている。

この４年間の音声入力数、笑い声と咳払いの入力数、雑音の入力数を図５に示す。

また、最近(2006.8.1～8.20)の入力数を図６に示す。夏休みの期間でもあり、音声入力のおおよその打ち分けは、子どもが 90%、大人が 10%である。雑音のうち､約半数は､800ms 以下の短い入力であり、長さのみで瞬時にリジェクトされている。さらに、サービスタスクの内訳を表１に示す。サービス内容が多岐に渡っていることが分かる。

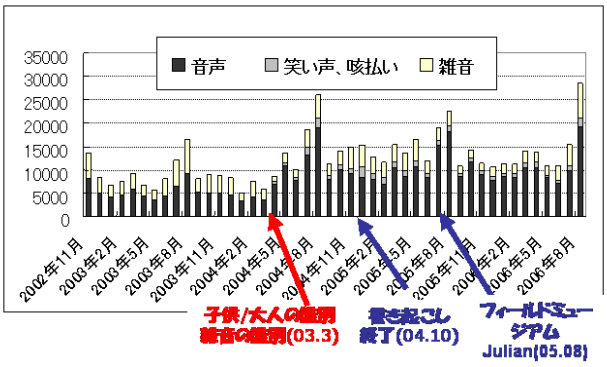

図 5. たけまるくんの運用状況（月別の入力数）

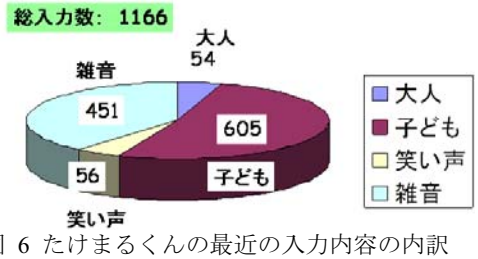

図 6 たけまるくんの最近の入力内容の内訳

表１　たけまるくんの最近のサービスタスクの内訳

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%

３．２．キタロボとキタちゃん

次にキタロボとキタちゃんの３月から８月までの運用状況を図７に示す。一日あたりの入力数はキタロボが約 200 発話、キタちゃんが 100 発話である。雑音環境は、たけまるくんの 50dBA に比較して、約 60dBA と高いにもかかわらず、音声入力に対する雑音の比（SNR）は変わっていない。雑音レベルが高いのでユーザがたけまるよりも比較的大きな声で発声する傾向にあり、かつ、発声レベルに合わせてＡＤのレベルも調整したことによると推測される。800ｍｓ以下の短いリジェクトの比率もたけまるくんとほぼ同様である。次にキタロボとキタちゃんの最近(2006.8.1～8.20)の入力の内訳を図８に示す。キタちゃんの大人と子どもの比率は、１：２、キタロボは１：４であり、どちらも駅という状況のため大人の比率がたけまるくんよりも高い。キタロボのサービスタスクの内訳を表２に、キタちゃんの内訳を表３に示しておく。新たに追加したタスクで、周辺案内はある程度利用されているが、路線検索/案内はあまり利用されていないことがわかる。たけまるくんと同様、サービス内容は多岐に渡っている。

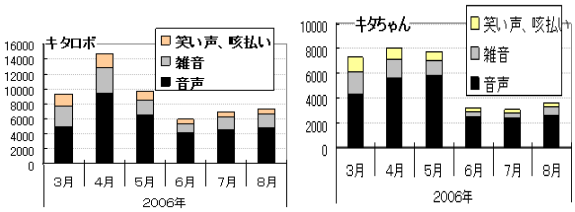

図 7　キタちゃん、キタロボの運用状況

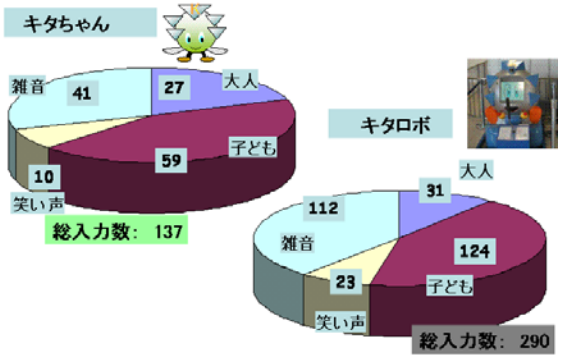

図 8. キタちゃんとキタロボの入力の内訳

表２　キタロボのサービスタスクの入力の内訳

ドメイン	大人		子供	
汎用的・時間など	266	19%	914	18%
キタロボ会話	622	44%	2,752	53%
ウェブ検索、HP	27	2%	43	1%
天気・ニュース	128	9%	327	6%
生駒市関係	96	7%	260	5%
北生駒駅関係	180	13%	504	10%
周辺案内	34	2%	210	4%
路線表示	48	3%	174	3%
合計	1,401	100%	5,184	100%

表３　キタちゃんのサービスタスク入力の内訳

ドメイン	大人		子供	
汎用的・時間など	227	20%	508	22%
キタちゃん会話	440	39%	1,244	52%
ウェブ検索、HP	36	3%	41	2%
天気・ニュース	93	8%	91	4%
生駒市関係	95	9%	118	5%
北生駒駅関係	73	7%	153	7%
周辺案内	138	12%	207	9%
路線探索	14	1%	10	0%
合計	1,116	100%	2,372	100%

４．２年間の書き起こしデータによるたけまるくんの改善の予備検討

２年間の音声データの書き起こしが終了している。このデータには、書き起こしの他に、話者の属性（男性/女性、幼児/低学年/高学年/大人/高齢者）が、声を聞くことにより付与されている。２回以上の発話には、応答正解文を付けている。この書き起こしデータを利用して、音響モデル、言語モデル、質問データベースを新たに作成した。２年間の音声データベースの概要を表４に示す。また、評価セットは、書き起こし文の頻度の多いものから１つずつ大人 2000 発話、子ども 3000 発話を選択した。評価セットを表５に示す。応答正解率は、応答文付きのデータのみによって算出した。

表４ ２年間の書き起こし音声データ(学習セット)

学習セット	発話数	時間
大人	23,417	11.5 h
子供	120,671	68.7 h
全部	144,088	80.2 h

表５　評価に用いた音声データ（評価セット）

評価セット	発話数	応答文あり
大人	2,000	1,350
子供	3,000	2,596
全部	5,000	3,946

現在のたけまるくんシステムが動作している音韻モデル、言語モデル、質問データベースの仕様を表６に示しておく。このベースラインシステムの性能を表７にまとめておく。

これらの 125 日間の音声書き起こしデータを用いたベースラインシステムを比較対照として、表４の２年間の書き起こし音声データを利用して音響モデル、言語モデル、質問データベースを表８に示す手順で学習して、システムの更新を行った。学習テストによる性能評価の結果を表９に示す。大幅な性能の向上が期待できる。とくに、言語モデル、質問データベースが向上に寄与している。

表６ ベースラインシステムの学習の手順

ベースライン学習データ	2002 年１１月から 2003 年３月（１２５日間）の雑音なしの音声データ 大人：　8106 発話 子ども：　17392 発話
音韻モデル	大人：　ＪＮＡＳ（約４万発話）のＰＴＭをもとにＭＡＰ適応 子ども：　ＪＡＮＳ女性（約２万発話）をもとにＭＡＰ適応
言語モデル	生駒市関連の Web(1080k 文)、想定質問文(6.4k 文）の言語モデルをもとに、 大人、子どもの書き起こしテキストから作成した言語モデルを１：１で融合。 さらに、文法による単語制約の強化。
質問データベース	１２５日間の書き起こしテキストをもとに作成。大人：　３，３１３質問文(302 応答文) 子ども：　６，５７３質問文(298 応答文)

表 7 ベースラインシステムの性能

	Perplex-ity	単語正解精度	単語正解率	応答正解率(書き起こし)	応答正解率（認識）
大人	２１	81.0%	84.0%	81.8%	75.8%
子ども	１６	63.9%	70.2%	71.2%	59.4%

表８　２年間の音声データによる学習手順

学習データ	2002 年１１月から 2004 年 10 月（２年間）の雑音なしの音声データ 大人：　23,417 発話 子ども：　120,671 発話
音韻モデル	大人：　２年間の大人音声データで F-B アルゴリズムで学習 子ども：２年間の子ども音声データで F-B アルゴリズムで学習
言語モデル	２年間の書き起こし文で学習
質問データベース	２年間の書き起こし文で２回以上の質問文をベースラインの質問文に追加。 大人：　３，９４８質問文(393 応答文) 子ども：　１０，２７３質問(404 応答文)

表９　２年間の音声データに更新による性能

	Perplex-ity	単語正解精度	単語正解率	応答正解率(書き起こし)	応答正解率（認識）
大人	１１	84.2%	89.2%	90.1%	83.1%
子ども	9	70.3%	77.3%	94.4%	76.1%

２年間の書き起こし音声データによるモデル、質問データベースなどの更新により、単語認識率、応答正解率とも大幅に向上している。次に、各モデルの更新の効果を調べるために、ベースラインシステムに、各更新モデル/データを入れ替えた場合の性能を表１０に示す。各モデル/データの更新の効果があることが分かるが、とくに全てのモデル/データの更新を行った場合の相乗効果が大きいことが分かる。

上記の評価では、大量の２年間の書き起こしを利用しているが、評価データも学習データに含まれている。文献[17,18]では、従来の音韻モデルと言語モデルを用いて、質問データのみを増やして、１ヶ月間のすべての音声入力に対する複数の正解を許した応答正解率を求めている。大人で７９.２％、子どもで６１.４％の応答正解率を確認している。書き起こし文の入力では、大人で８５.３％、子どもで８０.９％の応答正解率である。これらの結果からも、たけまるくんタスクでは、常に１０数％の想定外発話が入力されていることが推測される。この想定外発話の検出も重要な問題である。この検出を、質問文選択のスコアと入力の形態素数から行う可能性を調べるために、図９のようなグラフを描いた。図で○が正解の応答、＋が誤りの応答である。誤りの応答にはタスク外発話が多く含まれており、スコアと入力音声の形態素数の関係から、誤りの多くがタスク外発話として検出ができる可能性が推測できる[18]。

表１０　音韻モデル(AM)、言語モデル(LM)、質問データベース(DB)の更新の効果

年齢	評価値	ベースライン	AM だけ更新	LM だけ更新	AM+LM 更新	DB だけ更新	全部更新
大人認識	正解精度	81.0	82.4	84.0	84.2	-	**84.2**
	正解率	84.0	87.6	86.3	89.2	-	**89.2**
	質問データ	3,307	-	-	-	3,948	**3,948**
	応答正解率	75.8	78.4	76.4	78.1	80.1	**83.1**
子供認識	正解精度	63.9	66.1	67.6	70.3	-	**70.3**
	正解率	70.2	74.4	72.6	77.3	-	**77.3**
	質問データ	6,573	-	-	-	10,273	**10,273**
	応答正解率	59.4	62.5	59.5	61.9	67.7	**76.1**

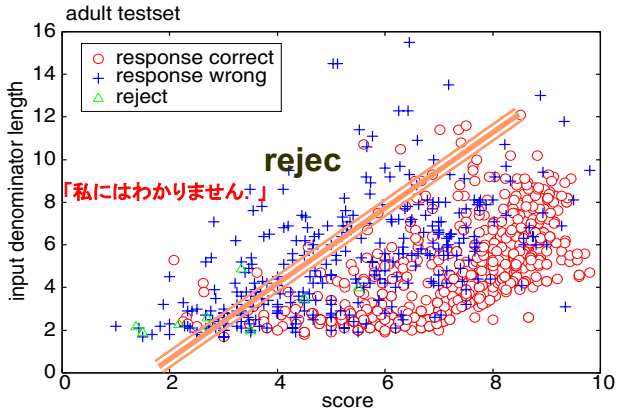

図９　質問文選択スコアと入力形態素数によるタスク外発話の検出　（大人の音声入力の場合）

４．キタちゃんへのポータビリティ

音声認識システムのポータビリティはビジネスモデルの観点からも重要である。音韻モデルのポータビリティについて、２０日間の音声データによる予備実験の結果について報告する[19]。

たけまるくんの設置されている生駒市北コミュニティセンターは比較的に静かで、約５０dBA である。キタロボの近鉄学研北生駒駅は周りの交通雑音などで約６０dBA である。キタロボ、キタちゃんの１ヶ月の書き起こしが完了したので、最初の２０日間の音声データ（大人：2864 発話、子ども：4733 発話）用いてたけまるくんのベースラインモデルからＭＬＬＲ適応を行った。評価データとしては、残りの１０日間のキタロボの音声データ（大人：1131 発話、子ども：1448 発話）を用いた。評価結果を図１０に示す。比較として、ＣＳＲＣで配布している子どもモデル、大人のＪＮＡＳモデルの性能も示した。ＭＬＬＲ適応による効果は若干あるが、騒音レベルの違いにもかかわらず、たけまるくんモデルも良好に動作していることが分かる。

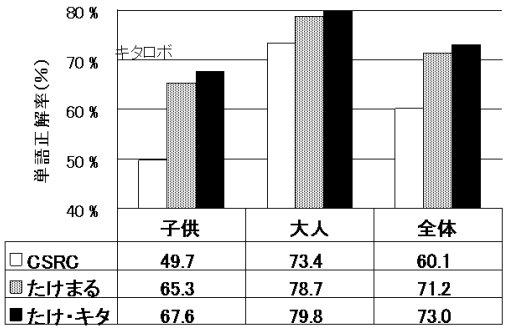

図１０　キタロボデータによる音響モデルの適応

騒音レベルが 10dB 高いにもかかわらず、たけまるくんモデルが良好に動作していることの検証のひとつとして、収録データの SNR のヒストグラムを作成した。結果を図１１に示す。指向性マイクと騒音下での発話レベルの上昇により、両者の SNR はほとんど同じであることが分かる。また、発話パワーに関しては、両者とも子どものパワーが大人よりも大きい。

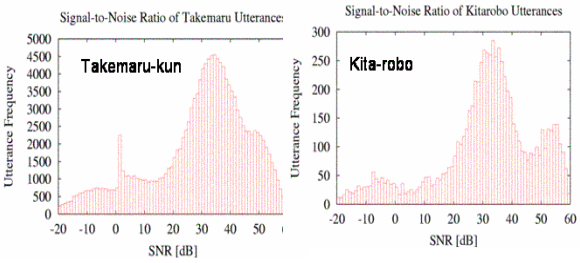

図１１　たけまるくんとキタロボとの SNR

５．今後の課題と計画

たけまるくんは、設置してから４年になるが、市民、とくに子どもに好かれて良好に動作してい

る。このような音声による情報案内システムは、応答内容に誤りがあっても、ユーザ自身がその誤りを瞬時に判別できることもあり、音声認識が有効に利用できる一つの実例である。この情報案内をさらに広範なＷｅｂ検索につなげてゆくのが今後の課題の一つである。２年間の音声データの書き起こしが終了しており、このデータを用いた効率のよい音韻モデル[13]、言語モデルの半教師なし学習の検討を行う。また、質問応答データベースの構築も進め、質問応答システムの構成法の検討も行う[17,18]。エージェント型やロボット型は、子どもや幼児に好かれ、子どものＩＴ教育にも有効である。子どもや幼児の音声認識も重要な研究テーマである[15,20]。

騒音レベルの高い駅にキタちゃんとキタロボを設置した。当初の予想より、良好に動作している。今後、たけまるくんからのポータビリティに関する検討を行う予定である。また、キタロボは、様々な視覚センサーも備えており、視覚情報も活用したマルチモーダルハンズフリー音声対話システムに拡張する予定である。

謝辞

この研究は，おもに文部科学省の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」によって行なわれたものである。これまでたけまるくん、キタちゃん/キタロボの構築や運用に多大な貢献をしてくれた研究室の学生に感謝する。また、音声対話システムの設置を快諾し、ご協力を頂いた生駒市や近鉄の各位にも感謝します。

参考文献

[1] A.L Gorin, G. Riccardi, J.H. Wright: “How May I Help You?,” Speech Communication, vol.23, pp.113-127, 1997.

[2] 西村 竜一 他，”実環境研究プラットホームとしての音声情報案内システムの運用”，信学論，Vol.J87-DII，No.3，pp.789-798，March 2004.

[3] 西原洋平，西村竜一，李晃伸，猿渡洋，鹿野清宏，"ユーザ層別並列モデルを用いた音声情報案内システム"，音響学会講演論文集，1-8-22, pp.49--50, March 2004

[4] Ryuichi Nisimura, Akinobu Lee,Masashi Yamada, Kiyohiro Shikano,"Operating A Public Spoken Guidance System In Real Environment", Interspeech'-Eurospeech, pp.845-848, Sep., 2005.

[5] 李晃伸，西村竜一，山田真士，鹿野清宏: "公共音声情報案内システム「たけまるくん」の運用および収集発話の分析"， 情報処理学会研究報告，2004-SLP-53-9, Oct. 2004.

[6] Akinobu Lee, Keisuke Nakamura, Ryuichi Nisimura, Saruwatari Hiroshi, Kiyohiro Shikano: "Noise Robust Real World Spoken Dialogue System using GMM Based Rejection of Unintended Inputs", Interspeech-(ICSLP) Vol.I, pp.173-176, Oct. 2004

[7] Ryuichi Nisimura, Akinobu Lee, Hiroshi Saruwatari, Kiyohiro Shikano: "Public Speech-Oriented Guidance System with Adult and Child Discrimination Capability", ICASSP2004, Vol.I, pp.433-436, May 2004.

[8] Akinobu Lee，Tetsuya Kawahara，Kiyohiro Shikano，”Julius –An Open Source Real-Time Large Vocabulary Recognition Engine”，Proc. of EuroSpeech，pp.1691-1694，Sep. 2001.

[9] Tatsuya Kawahara, Akinobu Lee, Kazuya Takeda, Katsunobu Itou, Kiyohiro Shikano, “Recent Progress of Open-Source LVCSR Engine Julius and Japanese Model Repository”, Proc. of ICSLP, pp.688-691, Oct. 2004

[10] 河原達也，李晃伸，"連続音声認識ソフトウェア Julius"，人工知能学会誌, Vol.20, No.1, pp.41-49, 2005

[11] 鹿野清宏，伊藤克亘，河原達也，武田一哉，山本幹雄 編著 "音声認識システム"，オーム社, 2001-05

[12] Akinobu Lee, Kiyohiro Shikano, Tatsuya Kawahara, "Real-Time Word Confidence Scoring using Local Posterior probabilities on Tree Trellis Search," Proceedings of ICASSP2004, Vol.I, pp.793-796, May 2004

[13] Tobias Cincarek, Tomoki Toda, Hiroshi Saruwatari, Kiyohiro 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, March 2006

[14] Randy Gomez, Akinobu Lee, Tomoki Toda, Hiroshi Saruwatari, Kiyohiro Shikano, ``Improving Rapid Unsupervised Speaker Adaptation based on HMM Sufficient Statistics in Noisy Environments using Multi-template Models," IEICE Trans. Information and Systems, Vol.E89-D, No.3, pp.998-1005, March 2006

[15] 鮫島充, Randy Gomez, 李晃伸, 猿渡洋, 鹿野清宏、“実環境における子供音声認識のための音韻モデルおよび教師なし話者適応の評価”、情報処理学会論文誌, Vol.47, No.7, pp.2295-2304, July 2006

[16] 川波弘道，杉本理香，藤本秋子，ツィンツァレク山本瑞穂，木田学，ツィンツァレク トビアス，鹿野清宏,”公共音声案内システムユーザ発話データベースの整備と予備的分析”，音響学会講演論文集, 3-2-6 pp.83--84, Sep. 2006

[17] 木田学，川波弘道，猿渡洋，鹿野清宏,” 音声情報案内システムにおける質問応答データベース設計手法の検討”，音響学会講演論文集，3-2-7 pp.85--86, Sep. 2006

[18] 早川直樹，木田学，ツィンツァレク トビアス，川波弘道，猿渡洋，鹿野清宏,” 音声情報案内システムの応答文選択におけるスコアリング手法の改善”，音響学会講演論文集，3-2-8 pp.87--88, Sep. 2006

[19] ツィンツァレク トビアス，川波 弘道，猿渡洋，鹿野 清宏, “「キタちゃん」実環境音声情報案内システムの音響モデル構築と予備評価”，音響学会講演論文集, 1-2-10 pp.19--20, Sep. 2006

[20] 進藤 泉，ツィンツァレク トビアス，戸田 智基，猿渡 洋，鹿野 清宏,” 幼児特有の発音変動を考慮した実環境音声認識の検討”，音響学会講演論文集, 1-2-9 pp.17--18, Sep. 2006