

## 音声質問応答システムのユーザ発話を用いた 質問応答データベース拡張支援

成松 宏美<sup>†1</sup> 中野 幹生<sup>†2</sup> 船越 孝太郎<sup>†2</sup>  
長谷川 雄二<sup>†2</sup> 辻野 広司<sup>†2</sup>

質問と応答のペアからなるデータベースを利用する質問応答システムの性能は、データベースの内容に依存する。できる限り多くの質問パターンを用意することにより、ユーザの質問のバリエーションに対応できる。本稿では、音声質問応答の初期システムを用いて収集したユーザの質問の中から、データベースに追加すべき質問を検出することにより、質問応答データベースの拡張を支援する方法を提案する。提案手法は、質問応答データベースから作成した言語モデルを用いた音声認識結果と、大語彙統計言語モデルを用いた音声認識結果を併用する。実験結果により、提案手法の有効性が示された。

### Supporting the Augmentation of the Question-Answer Database For Speech-Based Question-Answering System Using User Utterances

HIROMI NARIMATSU,<sup>†1</sup> MIKIO NAKANO,<sup>†2</sup>  
KOTARO FUNAKOSHI,<sup>†2</sup> YUJI HASEGAWA<sup>†2</sup>  
and HIROSHI TSUJINO<sup>†2</sup>

The performance of a system that answers to a user question using question-answer database depends on the contents of the database. Dealing with a variety of user questions requires preparing question-answer patterns as many as possible. This paper proposes a method for supporting the augmentation of the question-answer database. It uses user questions collected with the initial question-answer system, and detects questions that need to be added to the database. It uses two language models; one is built from the database and the other is a large-vocabulary statistical model. Experimental results showed the effectiveness of the proposed method.

#### 1. はじめに

近年のロボットやアニメーションエージェントの発展に伴い、これらのロボットやエージェントが人間の音声による質問に答えられるようなシステムが注目されている。さまざまなタイプの対話システムがつくられており、カーナビゲーションシステムのように、実際に一般に用いられているシステムもある。

本研究では、さまざまなタイプの対話システムの中で、質問文例と応答のペア（質問応答ペア）からなるデータベースを持ち、ユーザの質問から質問応答ペアを検索して応答する質問応答システムを扱う。フレームに基づく対話管理を行うシステム<sup>1)</sup>や、VoiceXML<sup>10)</sup>で記述されるような対話システムに比べると単純ではあるが、逆にシステムの動きが予測しやすく、音声言語処理の非専門家でも容易にデザインすることができる。西村ら<sup>7),8)</sup>やCincarekら<sup>2)</sup>は、「たけまるくん」や「キタちゃん」などの質問応答システムを開発し、実環境で運用してデータ収集に用いている。収集されたデータは、応答選択の向上手法の研究に用いられている<sup>9),12)</sup>。

本稿では、我々がこれまで構築してきた質問応答システム<sup>6)</sup>を利用する。我々のシステムでは、簡易な記述言語のみで質問応答システムの実現が可能である。しかし、データベースのコンテンツ作成者が、あらゆる質問をカバーする質問応答データベースを作ることは困難である。ユーザは、コンテンツ作成者の予想から外れる質問をすることも多い。そのような、データベースに対応する応答が存在しないような質問をユーザがした時、システムは、適切な質問応答ペアを見つけられず、質問の繰り返しを要求したり、ユーザの発話をユーザの意図と異なる質問応答ペアにマッチさせてしまい、間違った応答をしたりし、対話がうまく進まない。したがって、なるべく多くの種類の質問に対して質問応答ペアを用意することが必要である。

本研究は、コンテンツ作成者に、質問応答ペアの作成支援を行うことを目的とする。実際にシステムを使ってもらい収集した発話データのうち、ある発話がデータベースにない質問であることをシステムが自動で判断出来れば、全部の発話を書き起こしたり、聞いたりする

†1 電気通信大学大学院情報システム学研究所

Graduate School of Information Systems, The University of Electro-Communications

†2 (株)ホンダ・リサーチ・インスティテュート・ジャパン

Honda Research Institute Japan Co., Ltd.

★1

ことなくデータベースを拡張することが可能となる。

本稿では、質問応答システムを使用して収集したデータを利用し、実際のユーザ発話データからデータベースに追加すべき質問（データベース外質問）を自動的に発見する方法を提案する。提案手法は、ユーザ発話の音声認識結果およびそれを用いた質問応答ペアの検索結果に関する特徴量に加え、大語彙言語モデルを使用した音声認識結果に関する特徴量を用いて、データベース外質問を自動検出する。

## 2. 質問応答データベースに基づく対話システム

本稿で扱う質問応答システムは、質問と応答のペアからなる質問応答データベースを用いる。ユーザの質問が入力されると、その質問にマッチする質問応答ペアを検索し、応答を行う。これは質問分類または応答選択と呼ばれるが、本稿では応答選択と呼ぶ。

このような質問応答システムのうち「たけまるくん」<sup>8)</sup>や「キタちゃん」<sup>2)</sup>は、大量のデータを用いてモデルを訓練することによりシステムを向上させている。しかし、初めから大量のデータがある状況でシステムを用いることはできないため、データ収集に用いるための初期システムを容易に構築する必要がある。我々は、誰でも簡単に質問応答システムが作成できるよう、簡易な記述言語を用いて質問応答を実現するシステムを構築した。

我々のシステムは、ASKA<sup>7)</sup>と同じく、キーフレーズベースのマッチングにより応答選択を行う。実際に用いた質問応答データベースの例を図1に示す。質問応答データベースには、質問文例と応答をペアで記述する。質問文例では、予想される質問を文章で記述し、マッチングに用いるキーフレーズを{と}で囲んで指定する。例えば、「仙川に公園はあるの」は、図1に示す質問応答ペアの質問文例の「{仙川}にある{公園}について教えてください」にあるキーフレーズをすべて持っているため、この質問応答ペアにマッチする。他にも「仙川にどんな公園があるか教えて」や「仙川に大きい広場はありますか」などの質問がこの質問応答ペアにマッチする。1位の音声認識結果が質問応答ペアにマッチすると、その応答部分を実行する。マッチする質問応答ペアがなければ、データベース中にあるデフォルト応答を用いて「もう一度おっしゃってください」「その質問には答えられないかも知れません」などと発話する。応答部分には、システム発話、システムの動作、画面上に表示するスライドなども指定することが出来るようになっている。音声認識用の統計言語モデルは、記述した質問文例から学習する。

```
<qa id="sengawa-park">
<question>
<sentence>{仙川}にある{公園}について教えてください</sentence> (質問文例)
<sentence>{仙川}に{広場}はありますか</sentence>
</question>
<answer topic="sengawa-park">
<mdl>
<show-slide uri="file://$[SENGAWA_PPT_DIR]/wh.ppt#17" /> (スライド表示)
<utterance>仙川には武者小路実篤の公園があります。 </utterance> (応答)
</mdl>
</answer>
</qa>
```

図1 質問応答ペアの記述例

## 3. ユーザ発話とそのシステム理解の分析

### 3.1 ユーザ発話データの分析

データベースに追加すべきデータベース外質問を発見するために、ユーザの発話と、その時のシステムの応答選択について分析を行った。分析には、我々の質問応答システムを用いて収集したデータを使用した。用いたコンテンツは、京王線仙川駅周辺のタウン情報に関するものである<sup>6)</sup>。232個の質問応答ペアを作成した。質問文例数は890、言語モデルの語彙サイズは460である。

データ収集は、画面上のキャラクターエージェントに向かって質問し、ユーザが知りたい情報を聞き出すという形態で行った。答えられる内容は、対話中のシステムの応答やスライド内に示したが、質問例の提示は行わなかった。男性12名、女性13名に、それぞれ約14分間、システムと対話をしてもらった。

分析では、ユーザ発話をテキストに書き起こしたものと、実際のユーザ発話を音声認識した結果の両方からキーフレーズマッチングにより応答選択を行い、その結果を比較した。その際、書き起こしは、2つのデータベースを用いて応答検索した結果を比較する。1つは、初期データベース (IniDB) を用いたもの、もう1つは、書き起こしから、初期データベースに質問応答ペアが存在しないコンテンツを追加することにより拡充したデータベースを用いたものである。拡充したデータベースを用いることで、ユーザの意図を得ることができ、システムがユーザの意図通りに理解できたかを分析できる。以降、後者のデータベースを拡張データベース (ExpDB) と呼ぶことにする。

分析は、収集したデータの中で、1話者につき、それぞれ最初の25発話分を使用して行っ

た．25 発話 × 25 人分=625 発話のデータを書き起こし，データベースに該当する質問応答ペアがなければ，データベースに質問と応答のペアを追加することによって，拡張データベースを作成した．

ここで，分類に利用する用語の定義する．

- Trans+IniDB: 発話の書き起こしを用い，初期データベースに基づいて応答選択した結果
- Trans+ExpDB: 発話の書き起こしを，拡張データベースに基づいて応答選択した結果（ユーザ意図に相当）
- ASR+IniDB: 音声認識結果を，初期データベースに基づいて応答選択した結果

Trans+ExpDB はユーザの意図を表していると考えられるため，Trans+IniDB と Trans+ExpDB が同じかどうかを調べることにより，その発話にマッチする質問応答ペアが存在するかどうかを調べることができる．存在するものをデータベース内質問，しないものをデータベース外質問と呼ぶ．また，ASR+IniDB が Trans+IniDB，Trans+ExpDB と同じかどうかを調べることにより，音声認識を用いた場合にどのくらいうまく理解できたかを調べることができる．ASR+IniDB で用いた言語モデルは，初期データベースに記述された例文を用いて作成し，一般的に使用頻度の高い 5000 語を追加したものである．5000 語は，未登録語彙がキーフレーズに誤認識され，理解誤りを起こす可能性を減らすために追加した．追加した語彙は未知語クラスとして扱われる．

表 1 に分類結果を示す．全発話数 625 に対して，Trans+IniDB と Trans+ExpDB が一致するものが (T1) に当たり，(T1) に加えて認識結果による理解 ASR+IniDB も一致する発話が 248 発話である．(T2) はシステムによる応答はあったが，ユーザ発話の意図と異なるものが 78 発話あったことを示す．発話書き起こしを元のデータベースで応答選択を行ったときに質問応答ペアが見つからなかったものが (T3) にあたり，135 発話あった．また，拡張データベースを用いても応答選択でシステムの応答が得られなかったものやノイズや区間検出誤りを含む発話が (T4) であり，122 発話あった．

### 3.2 データベース拡張支援に向けた分析結果の検討

データベース拡張のために，表 1 の結果の中でどの発話をデータベースに入れるべきかを検討する．Trans+IniDB と Trans+ExpDB の結果が異なるが，理解結果が正解と一致する場合には，データベース内にその質問に対しての質問応答ペアが存在するため応答はする．しかし，ユーザの意図とは異なるため，キーフレーズの指定方法に修正の必要があるか，若しくは，ユーザがその内容についてより詳しい内容を要求している場合が考えられ

		音声認識結果の分類結果					合計
		ASR+IniDBがDB内				ASR+IniDB がDB外 (A5)	
		ASR+IniDB =Trans+IniDB =Trans+ExpDB (A1)	ASR+IniDB = Trans+IniDB (≠Trans+ExpDB) (A2)	ASR+IniDB = Trans+ExpDB (≠Trans+IniDB) (A3)	その他 (誤認識) (A4)		
書き 起こし の 発話 分類 結果	Trans+IniDB =Trans+ExpDB (T1)	(1) 246			(2) 15	(3) 31	292
	Trans+IniDB (≠Trans+ExpDB) がDB内 (T2)		(4) 44	(5) 2	(6) 26	(7) 4	76
	Trans+IniDB がDB外 (T3)			(8) 6	(9) 38	(10) 91	135
	ノイズなど (T4)			(11) 2	(11) 12	(12) 108	122
合計		246	44	10	91	234	625

表 1 発話の分類

る．また，Trans+IniDB でデータベース外であるのに，応答選択結果でデータベース内のものに一致している場合には，ユーザ発話がデータベース外であってもデータベース内に類似した発話があることを示す．

ユーザ発話の意図とシステムの理解が異なる場合，またユーザの発話がもともとデータベースに存在しない場合には，データベースのコンテンツに追加すべきと考えられる．従って，表 1 の中でデータベースに追加すべき発話は，(T2) の DB 外発話が DB 内と誤って分類された場合と，と (T3) の実際に DB 外発話であったもの，ということができる．

## 4. 質問応答システムのデータベース拡張支援

### 4.1 大語彙言語モデルを用いた音声認識結果の利用

ユーザの質問のバリエーションに対応するためには，ユーザの実際の質問をもとにデータベースを拡張する必要があるが，システムが自動でデータベース外発話を検出し，そのリストをコンテンツ作成者に提示することにより，コンテンツ作成者は，すべてのユーザ発話を書き起こしたり聞いたりすることなく，少ない労力でデータベースを拡張することができる．

我々のシステムでは，データベース中の例文から作成した言語モデルでの認識結果を用いている．ユーザ発話中に，データベース中の例文にない単語が含まれている場合には，誤認識がおこり，その発話がデータベース外質問かどうかを判定することが難しい．

そこで、本稿では Julius 付属の大語彙の言語モデル<sup>4)</sup>を用いて認識した結果を用いることで、データベースから作成した言語モデルで認識した結果と比較して、データベース外質問を検出する方法を提案する。大語彙の言語モデルを用いた認識結果の中から応答検索の鍵となるようなキーフレーズを抽出する事ができれば、データベース外質問であることを自動で検出できると考えられる。また、大語彙の言語モデルを用いた認識結果は、誤った認識もあるが、音韻は正しく認識している場合が多く、この情報を用いることにより、データベースから作成した言語モデルを用いた認識結果の確からしさを推定することが可能である。

#### 4.2 データベース外発話の検出のための判別器の学習

データベース外発話を自動で検出するための方法として、発話から得られる様々な特徴量を用いて学習した判別器を用いる。特徴量は、認識結果から得られるものや、応答選択の結果として得られるものなどである。例えば、応答選択の際に用いたキーフレーズが認識結果の候補に出現する割合などが特徴量となる。キーフレーズの出現頻度により、そのキーフレーズの認識の正しさを知ることができると考える。また、認識結果の中に、応答選択に用いたキーフレーズ以外の名詞などが存在しているかどうかとも特徴量となる。応答選択に利用していない名詞が存在すれば、その名詞がユーザ発話の意図を理解する上で必須となる場合には、システムの理解とユーザ発話の意図が異なることを示唆する。このように、様々な特徴量を用いて判別器を学習し、データベース外発話の自動検出を行う。

### 5. 実験

提案手法の有効性を評価するために実験を行った。この実験では、3章の分析に用いたデータベースと発話データを用いた。

#### 5.1 使用した特徴量

本節では、使用した特徴量について述べる。音声認識の言語モデルは、

- $LM_{db}$ : 初期データベースから作成した言語モデルに一般的に使用頻度の高い 5000 語を追加したもの
- $LM_{lv}$ : Julius ディクテーション実行キット付属の言語モデル<sup>4)</sup>

を用いた。使用した特徴量について、表 2 に記述する。 $F_{db}$  は  $LM_{db}$  を用いた認識結果から得られた特徴量、 $F_{lv}$  は  $LM_{lv}$  を用いた認識結果から得られた特徴量、 $F_{dbl}$  は、 $F_{db}$  と  $LM_{db}$  を用いて算出した特徴量である。

$F_{db1}$  は、システムがユーザ発話に対して応答を見つけれられたかどうかである。見つけれなかった場合、誤認識か、ユーザ発話がデータベース外である可能性が高い。 $F_{db2}$ 、 $F_{db3}$  の

記号	意味	記号	意味
$F_{db1}$	システムがDB外と判断したかどうか(選択された応答の有無)	$F_{db10}$	認識結果があいづちリストに含まれているか
$F_{db2}$	認識結果の第一候補の音響スコア/発話長	$F_{db11}$	認識結果の第一候補の単語数
$F_{db3}$	認識結果の第一候補の言語スコア	$F_{db12}$	認識結果全ての単語数
$F_{db4}$	$\sum_n$ (応答選択に利用したキーフレーズの認識結果の第n候補に出現する個数/第n候補の単語数) / 認識結果の候補数	$F_{db13}$	AVG(応答選択に用いたキーフレーズの第一候補の julius confidence)
$F_{db5}$	$\min_i$ (応答選択に用いたキーフレーズ $i$ の全認識結果での出現回数)	$F_{db14}$	$\max_i$ (応答選択に用いたキーフレーズ $i$ の第一候補の julius confidence)
$F_{db6}$	$\max_i$ (応答選択に用いたキーフレーズ $i$ の全認識結果での出現回数)	$F_{db15}$	$\min_{\mu}$ (応答選択に用いたキーフレーズ $\mu$ の第一候補の julius confidence)
$F_{db7}$	認識結果の第一候補に含まれるDB内のキーフレーズ数 - 応答選択に用いたキーフレーズ数	$F_{db16}$	認識結果の候補数
$F_{db8}$	$\min_i$ (応答選択に利用したキーフレーズ $i$ の認識結果の第n候補に出現する個数/第n候補の単語数)	$F_{db17}$	認識結果の第一候補に含まれる名詞の数 / 認識結果の第一候補の単語数
$F_{db9}$	$\max_i$ (応答選択に利用したキーフレーズ $i$ の認識結果の第n候補に出現する個数/第n候補の単語数)	$F_{db18}$	認識結果の全候補に含まれる名詞の数 / 認識結果の全候補の単語数
$F_{lv1}$	システムがDB外と判断したかどうか(選択された応答の有無)	$F_{lv6}$	認識結果全ての語彙数
$F_{lv2}$	認識結果の第一候補の音響スコア/発話長	$F_{lv7}$	認識結果の候補数
$F_{lv3}$	認識結果の第一候補の言語スコア	$F_{lv8}$	認識結果の第一候補に出現する名詞の数 / 認識結果の第一候補の単語数
$F_{lv4}$	認識結果があいづちリストに含まれているか	$F_{lv9}$	認識結果の全候補に出現する名詞の数 / 認識結果の全候補の単語数
$F_{lv5}$	認識結果の第一候補の語彙数		
$F_{dbl1}$	音響スコアの差	$F_{dbl2}$	言語スコアの差

表 2 実験に使用した特徴量

音響スコア、言語スコアは、認識結果の精度を示す。 $F_{db4}$ - $F_{db9}$  は、応答選択に用いたキーフレーズに関して、認識の精度を示唆する。データベースに入れるべき発話の検出は、あいづちや、ユーザがシステムの反応を期待しない発話のリストを作成し、そのリストに一致するかどうかで判断する。リストに一致すれば、データベース外であるということを示唆する。また、名詞の数や出現頻度を比較することで、トピックの量などを判断することができる。発話に含まれる単語に対して、名詞の数が多ければ、複数のトピックを含んだ発話であることを示唆する。 $LM_{lv}$  を用いた音声認識結果に対しても同様の特徴量を求める。

#### 5.2 分類器の学習

抽出した特徴量を学習にかけることで分類器を生成する。学習に使用したツールは、weka<sup>11)</sup>である。今回の実験では、いくつかの分類手法を試した結果、NaiveBayes<sup>3)</sup>を用いることとした。実験データには、表 1 のうち、書き起こしの発話分類でも該当する応答のなかったノイズなど (T4) を除いた 505 発話を用いる。ノイズなどの検出は、別の方法で行うことを今後検討する。

		DB外発話と分類された数	DB外発話と分類された中のDB外発話数	DB外発話数	Precision	Recall	F値
判定	全てのDB外発話と判断 (ベースライン1)	505	211	211	0.418	1.000	0.589
	iniDBの判定 (ベースライン2)	367	116	211	0.316	0.550	0.401
	判別器を用いて判断 (提案手法)	230	154	211	0.670	0.730	0.698

表 3 ベースラインの判別精度

検出精度を評価する指標として、precision, recall, F 値を用いる。それぞれの求め方を以下に示す。

- precision: データベース外発話と分類された中のデータベース外発話数/データベース外発話と分類された数
- recall: データベース外発話と分類された中のデータベース外発話数/データベース外発話の数
- F 値:  $2 / ((1/\text{precision}) + (1/\text{recall}))$

ベースラインと、提案手法を用いて判別を行った判別精度を表 3 に示す！「全てデータベース外発話と判断」とは、全ての発話をデータベース外と判別することである。ここでの F 値は高くなってはいるが、実際検出をしているわけではないので、良い結果とは言えない。また、IniDB の判定とは、LM<sub>db</sub> の結果を IniDB を用いて応答選択を行ったとき、その結果がデータベース外となっているかどうかという情報のみを用いたものであるが、分類精度は高くないことがわかる。そこで、特徴量を学習させた分類器を用いて分類した結果を提案手法として評価した。学習は、NaiveBayes を用い、10-fold cross validation で学習と評価を行った。結果より、提案手法は、ベースラインよりも高い判別精度を示すことがわかった。

次に、判別精度向上に最も効果のある特徴量について記述する。一つの特徴量を削除し評価値を求めるという手順をそれぞれの特徴量について行うことで、どの特徴量が最も寄与しているかを調べた。有効な特徴量の上位 5 個と、その特徴量を削除した場合の F 値は次の通りである。発話分類に使用したキーフレーズの音声認識信頼度の最小値 F<sub>db15</sub>, (F 値 0.670), LM<sub>db</sub> と LM<sub>lv</sub> の第一候補の音響モデルの差 F<sub>dbl<sub>v</sub>1</sub>, (F 値 0.685), LM<sub>db</sub> の認識結果の候補数 F<sub>db16</sub>, (F 値 0.685), LM<sub>db</sub> の認識結果第一候補の単語数 F<sub>db11</sub>, (F 値 0.688), LM<sub>db</sub> のキーフレーズの julius confidence<sup>5)</sup> の平均値 F<sub>db13</sub>, (F 値 0.688)。

F<sub>db15</sub> と F<sub>db13</sub> は、キーフレーズの音声認識結果がどれくらい正しいかを示唆する。F<sub>db11</sub>

分類器に隔選したデータ数	DB外発話と分類された数	DB外発話と分類された中のDB外発話数	DB外発話数	Precision	Recall	F値
話者1人分のデータ	267	144	211	0.539	0.682	0.602
話者2人分のデータ	265	158	211	0.596	0.749	0.664
話者3人分のデータ	206	132	211	0.641	0.626	0.633
話者4人分のデータ	217	137	211	0.631	0.649	0.640
話者5人分のデータ	234	140	211	0.598	0.664	0.629
話者10人分のデータ	226	135	211	0.597	0.640	0.618
話者15人分のデータ	237	144	211	0.608	0.682	0.643
話者20人分のデータ	243	151	211	0.621	0.716	0.665

表 4 学習データ量を変化させたときの判別精度

は、認識の精度を示し、F<sub>db16</sub> は、単語数が多ければ複数のトピックについての質問の可能性であるということを示唆する。また、差 F<sub>dbl<sub>v</sub>1</sub> は、LM<sub>db</sub> と LM<sub>lv</sub> の認識結果がどれくらい実際の音声発話とずれているかを示唆する。以上により、キーフレーズが正しく認識されたかの推定が有効な特徴量であることがわかった。

### 5.3 少ないデータ量での判別器の学習

前節ではすべてのデータを用いて 10-fold cross validation により提案手法の評価を行ったが、判別器の学習に大量のデータが必要だと、学習データを構築する際に書き起こし等が必要となるため、データベース拡張の労力削減にならない。

そこで、判別器の訓練データ数が少ない場合の判別精度を調べた。その結果を表 4 に示す。10-fold cross validation の場合に比べると F 値は低いですが、ベースラインよりは高く、提案手法が有効であることがわかった。

## 6. おわりに

質問応答システムに対して、質問応答のデータベース拡張を支援するための手法として、システムによるデータベース外質問の自動検出方法を提案した。提案手法はドメイン依存の言語モデルを用いた音声認識と大語彙の言語モデルを用いた音声認識を併用する。実験の結果から、提案手法は、初期システムのデータベース外発話検出よりも高い精度で判別ができることがわかった。

提案手法では、同じドメインの発話を用いて判別器の学習を行っているため、判別器の学習に書き起こしが必要になる。この労力を削減するため、今後、他のドメインの発話で学習した判別器を用いることを検討する。

また、今回は、データベース外質問と判定された発話をすべてコンテンツ作成者に提示する方法をとったが、データベース外質問である可能性が高いものから順にコンテンツ作成者に提示することにより、少ない労力でデータベースの拡張が可能になると考えられる。このような手法に関しても検討していく。

質問応答システムにとって、システムがユーザの要求と正しく理解したか、期待する応答が出来ているかどうかを知ることは非常に重要な課題である。今後、データベース外発話の検出だけでなく、データベース外発話をデータベース内と認識し、誤った情報をユーザに提供することを防ぐための手法についても検討する。

### 参 考 文 献

- 1) J.Chu-Carroll. MIMIC: An adaptive mixed initiative spoken dialogue system for information queries. In *Proc. 6th Applied NLP*, pp. 97–104, 2000.
- 2) T.Cincarek, H.Kawanami, R.Nisimura, A.Lee, H.Saruwatari, and K.Shikano. Development, long-term operation and portability of a real-environment speech-oriented guidance system. *IEICE Transactions*, 91-D(3):576–587, 2008.
- 3) G.H. John and P.Langley. Estimating continuous distributions in bayesian classifiers. In *Eleventh Conference on Uncertainty in Artificial Intelligence*, pp. 338–345, San Mateo, 1995. Morgan Kaufmann.
- 4) T.Kawahara, A.Lee, K.Takeda, K.Itou, and K.Shikano. Recent progress of open-source LVCSR engine Julius and Japanese model repository. In *Proc. Interspeech-2004 (ICSLP)*, pp. 3069–3072, 2004.
- 5) A.Lee, K.Shikano, and T.Kawahara. Real-time word confidence scoring using local posterior probabilities on tree trellis search. In *Proc. ICASSP-2004*, pp. 793–796, 2004.
- 6) 成松, 中野, 船越, 長谷川, 辻野. ロボット・エージェント対話行動制御部構築ツール RIME-TK を用いた質問応答機能の実現. 情報処理学会研究報告 SLP-74, 2008.
- 7) 西村, 内田, 李, 猿渡, 鹿野. Julius を用いた学内案内ロボット用音声対話システムの作成. 電子情報通信学会技術研究報告 SP2001-99/NLC2001-64, pp. 93–98, 2001.
- 8) 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, J87-D-II(3):789–798, 2004.
- 9) 竹内, ツィンツアレク, 川波, 猿渡, 鹿野. 音声案内システムにおける質問応答データベースの最適化—音声認識結果を用いた質問応答データベースによる応答生成と最適化—. 電子情報通信学会技術研究報告 SP2007-146, pp. 295–300, 2007.
- 10) W3C. Voice extensible markup language (voicexml) version 2.0. W3C Recommendation, 2004.
- 11) I.H. Witten and E.Frank. *Data Mining: Practical machine learning tools and*

*techniques, 2nd Edition*. Morgan Kaufmann, San Francisco, 2005.

- 12) 吉見, 南角, 李, 徳田. 音声対話システムにおける条件付確率場に基づく自動応答文選択. 日本音響学会 2008 年春季研究発表会論文集, 2008.