

# 質問応答・情報推薦機能を備えた音声による情報案内システム

翠 輝 久<sup>†</sup> 河原 達 也<sup>†,††</sup>  
正 司 哲 朗<sup>††</sup> 美 濃 導 彦<sup>†,††</sup>

音声が主要なモダリティである環境において、自然言語で記述された文書を知識源として、インタラクティブにユーザに情報を提供する枠組みを提案する。これは、現在のオーディオガイドシステムのように、システム側から一方的に情報を提示するのではなく、ユーザ・システム双方が対話の主導権をとりながら、インタラクティブに情報検索・提示するものである。そのために、ユーザ主導の検索・質問応答 (pull) モードと、システム主導の提示 (push) モードを用意して、ユーザの状態に応じてこれらを切り替える。検索・質問応答モードでは、漠然とした検索要求に対して文書を要約して提示したり、特定の情報・事実を求める質問に応答したりする機能を実装した。また、提示モードにおいては、システム側から、ユーザにとって有用な話題を動的に選択して、質問形式を用いて提示する方法を考えた。以上の枠組みを、顔認証機能を有するロボットエージェントに統合し、京都の観光案内システムとして、京都大学博物館の企画展示において運用を行った。3カ月の運用期間中、のべ2,500人のユーザの利用があった。収集された対話を分析・評価した結果、提案手法がおおむね有効に機能していることを確認した。

## Speech-based Interactive Information Guidance System Using Question-Answering and Information Recommendation

TERUHISA MISU,<sup>†</sup> TATSUYA KAWAHARA,<sup>†,††</sup> TETSUO SHOJI<sup>††</sup>  
and MICHIIHIKO MINOH<sup>†,††</sup>

This paper addresses an interactive framework for information navigation based on document knowledge base. In conventional audio guidance systems, such as those deployed in museums, the information flow is one-way and the content is fixed. To make the guidance interactive, we prepare two modes, a user-initiative retrieval/QA mode (pull-mode) and a system-initiative recommendation mode (push-mode), and switch between them according to the user's state. In the user-initiative retrieval/QA mode, the user can ask questions about specific facts in the documents in addition to general queries. In the system-initiative recommendation mode, the system actively provides the information the user would be interested in. We implemented a navigation system containing Kyoto city information with face recognition function. The effectiveness of the proposed techniques was confirmed through a field trial by a number of real novice users.

### 1. はじめに

情報技術の進展により、我々は世界中に存在するあらゆる情報にアクセス可能な状況にある。しかしその反面、膨大な情報の中から本当に必要なものを見つけるのが困難になりつつある。とりわけ、外出先などで速やかに情報が欲しい場合には、携帯電話やキオスク端末などにより、どこでも (ユビキタスに) 利用可能

な音声により検索できると便利である。また音声は、人間同士のコミュニケーションに用いる身近なメディアであり、誰にでも使いやすい (ユニバーサルな) インタフェースを実現すると考えられる。音声対話システムに関する従来研究・開発では、天気案内<sup>1),2)</sup> やバスの運行情報案内<sup>3)</sup> などに代表される関係データベース (RDB) 検索を扱ったものが大半であったが、近年の Web 検索に代表される情報検索技術の発展にともない、より一般的な文書検索を対象とした研究がさかんに行われている<sup>4)</sup>。

しかしながら、これらの現状の音声による情報検索・案内システムには、気軽に何かを尋ねるような便利さがあるわけではない。実際、一般的な自動電話応

<sup>†</sup> 京都大学大学院情報学研究所  
Graduate School of Informatics, Kyoto University

<sup>††</sup> 京都大学学術情報メディアセンター  
Academic Center for Computing and Media Studies,  
Kyoto University

答(IVR)サービスでは、基本的にユーザはシステムの示す手順に従って応答する必要があり、知りたい箇所に直接アクセスすることはできない。また、博物館や観光地などで利用されるオーディオガイドシステムでは、システムがユーザに一方的に情報を提示するのみであるため、聞き逃したり、より詳細な説明を聞きたかったりする場合でも、聞き返しや補足質問をすることはできない。また、音声対話による文書検索の研究の大半は、音声による検索方法のみを扱ったものであり、検索された文書の音声による提示方法を扱った研究はほとんど行われていない。

本研究では、このような現状を脱却し、ユーザフレンドリーな情報案内を実現するために、ユーザ・システムの双方が対話の主導権をとりながら、システムの持つ知識にアクセスする枠組みを提案する。具体的なタスクとして観光情報案内を想定して、バックエンドにある文書の情報を効率的にユーザに提示することを目指す。

ここで、音声により出力を行う場合には、情報の提示方法が問題となる。ユーザの理解しやすさを考慮すると長い文書を読み上げるのは適切でない。ただし、単純に短く要約して提示すると、ユーザが知りたい重要な箇所を説明できない可能性がある。そこで、本研究では質問応答技術を利用して、文書中のあらゆる情報に対してランダムアクセスを可能とすることにより、自由に聞き返し・質問が可能なインタラクティブな提示を実現する。さらに、システム側からもユーザに対して積極的に関連項目の情報推薦を行うことを考える。このようなシステム主導による提示機能は、ユーザがシステムの有している情報の一覧を把握しにくい状況において、スムーズに対話を続けるために必要であり、また、観光案内のように必ずしも明確なゴールが存在せず、漠然と興味のある情報を収集する場合においても有効であると考えられる。

これに加えて本研究では、ユーザをスムーズに対話に導入するために、ロボットエージェントを利用する。近年、ヒューマノイドロボット<sup>5)</sup>や擬人化エージェント<sup>6)</sup>などのマルチモーダルなインタフェースを備えた対話システムの開発が行われている。このようなシステムは、ユーザがシステムに対して親近感を持ちやすいために、多くのユーザに利用されることが報告されている<sup>7)</sup>。本研究ではさらに、顔画像認識を利用して名前の呼びかけを行うことにより、ユーザにとってより親しみやすいシステムを目指す。

ただし、従来のロボットエージェントを用いたシステムでは、対話は基本的に一問一答形式であり、シス

テムの応答もテンプレートを用いたものがほとんどであったのに対して、本研究では前述の指針に基づいて、本格的な音声対話により京都の観光名所を案内するシステムを構築する。

以下に本論文の構成を述べる。まず、2章においてシステムの概要を述べる。次に3章で、提案システムの音声インタフェースの構成を説明する。続いて4章で、フィールドテストにおける発話データの収集とその分析結果について述べ、5章において音声対話部分の評価を行う。6章は結論である。

## 2. システムの構成

本研究において開発した音声対話システムは、基本的に京都の観光案内を行うものである。ただし、テンプレートベースの応答を用意することにより、簡単なあいさつやエージェント自身に関する質問などに対応できるようにする。システムは最初にユーザの顔画像認識を行い、名前を呼びかけるところから対話を始める。システムの外観を図1に、構成を図2に示す。本システムは、音声対話部と顔画像認識部から構成される。各構成モジュールは、情報統合モジュール(サーバ)を介して他のモジュールの動作状態を取得することにより、自身の動作を決定する。

ユーザが対話を行う相手として、情報通信研究機構(NICT)で開発されたロボット・フィノ(モモコと命名)を用いる。このロボットは、頭、手、腰の3カ所の部位の動作が可能であり、これらの部位を動作させることにより21種類の動作パターンを持つ。図1の写真左のディスプレイは、現在のトピックに関係した(観光地などの)写真の提示を行うためのものであり、テキストの提示は行わない。初期状態では京都の観光マップを表示している。

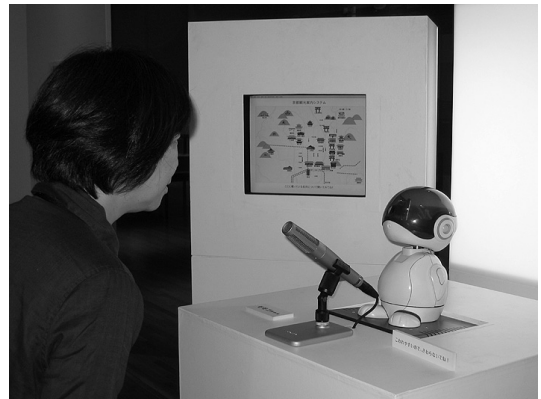


図1 システムの外観

Fig. 1 Outlook of the system.

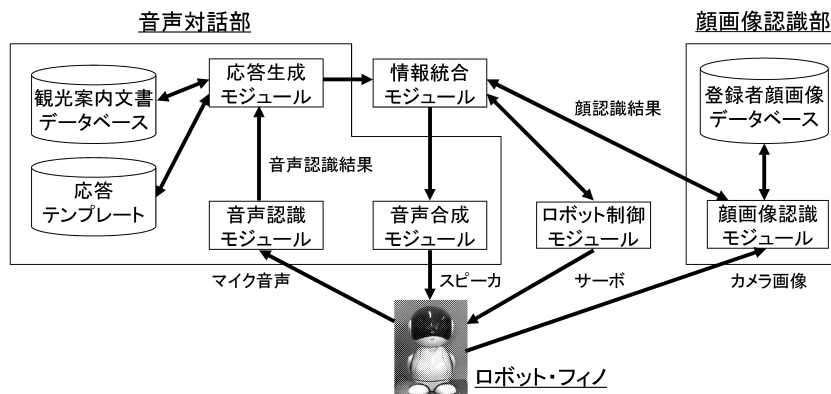


図 2 提案システムの構成

Fig. 2 System architecture.

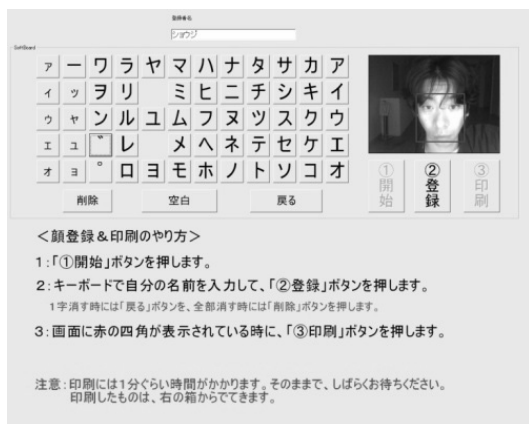


図 3 顔画像登録インタフェース

Fig. 3 Face registration interface.

## 2.1 顔画像情報を利用したユーザへの声かけ

本システムでは、ユーザを顔画像認識により声かけする。以下に、ユーザの顔登録から声かけまでの流れを示す。

ユーザは最初に別途設置された登録端末を用いて、顔画像と名前を登録する。顔画像の撮影は、ユーザが名前を入力している最中に連続して行う。これは、顔画像認識において、ユーザの表情や顔の向きの変化を考慮した多くのサンプルを登録することにより、頑健性を高めるためである。ただし、キーボードを使って名前を入力した場合、ユーザの顔が頻繁に下を向くため顔が検出できないことがある。そこで、本研究では、タッチモニタとソフトキーボードを利用して、図 3 に示すように、ユーザの顔を検出しやすい入力インタ

フェースを設計した。

次に、ユーザがロボットエージェントの前に来た際に、ユーザの名前を呼び、声かけを行う。これは、ロボットに内蔵されているカメラを用いてユーザの撮影・顔検出を行い、顔画像データベースと照合することにより実現される。この際に、3回連続して同じユーザが認証された場合のみ、認識成功とした。1回だけの認証では、認識誤りが頻出する恐れがあるためである。

なお、顔検出・認識モジュールには、オムロン社の「OKAO Vision ライブラリ」を使用した。また、顔画像の撮影は、Webカメラ（解像度 640×480）を用い、さらにユーザの顔の特徴（眼、鼻、口など）が明確にとらえられるように照明環境を調整した。

## 2.2 インタラクティブな情報案内

ユーザ・システムが相互に主導権をとるインタラクティブな情報提示を実現するために、本研究ではユーザ主導の検索・質問応答モードと、システム主導の情報推薦モードを用意して、ユーザの状態に応じて（対話管理部が）これらの切替えを行う。ユーザから検索要求や質問があった場合に、検索・質問応答モードに遷移し、発話の意図に沿った応答を生成する。また、システムの発話後、一定時間ユーザの発話がない場合には、システム主導の情報推薦モードに遷移して、現在のトピックに関連した情報を提示する。この概要を図 4 に示す。

本研究では、システムが検索・提示する文書として、Wikipedia の京都に関する文書と、京都市産業観光局が提供している京都情報データベースを使用する。それぞれの文書の概要を表 1 に、Wikipedia の

本システムを運用した博物館の企画展では、顔画像を処理して、印刷する別のデモも行われていた。

<http://ja.wikipedia.org/>

<http://raku.city.kyoto.jp/sight.phtml>

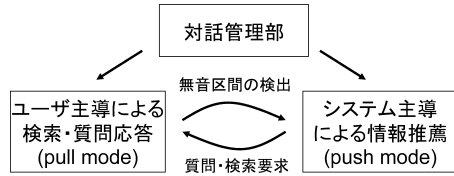


図 4 提案システムの対話戦略の概要  
Fig. 4 Overview of dialogue strategy.

表 1 システムが利用する知識ベース  
Table 1 Specification of the knowledge base.

知識ベースの種類	件数	見出し (節) 数	単語数
Wikipedia の 京都に関する文書	269	678	約 15 万
京都市観光局・ 京都情報データベース	541	541	約 7 万
合計	810	1,219	約 22 万

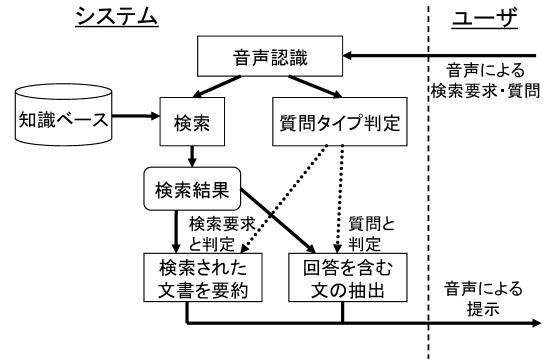


図 6 情報検索・質問応答処理の概要  
Fig. 6 Overview of document retrieval and QA.

### 3. 音声対話部の構成

#### 3.1 ユーザ主導の検索・質問応答

本研究ではユーザの発話を大きく 2 つのカテゴリに分類して扱う。1 つは検索要求であり、たとえば「金閣寺について教えてください」のような発話がこれに該当する。このような発話に対しては、知識ベース中の文書の見出し (節) を単位として検索し、結果をユーザに提示する。もう 1 つは質問であり、「建てられたのはいつですか」といった発話がこれに該当する。このような質問に対しては、システムは知識ベースから回答を含む文を抽出してユーザに提示する。以上の手順を図 6 に示し、以下詳しく説明する。

##### 3.1.1 トピック同定に基づく文脈解析

対話システムにおいて意味のある検索クエリを作成するために、コンテキスト情報を補完する必要がある。質問文に対して照応解析を行う方法<sup>9),10)</sup>も提案されているが、誤りを含む音声認識に対してこのような決定的な処理を行うのは容易でない。そのため、本研究ではユーザの過去の発話を用いて、検索に必要なキーワードを補うこと<sup>11)</sup>を考える。しかし、単純にすべてのユーザ発話を利用すると、対話中にトピックが変化した場合に、不要なキーワードを付加してしまう可能性がある。そこで、対話のトピックを同定することで、補うべきコンテキスト長 (= 利用する過去の発話数) を決定する。

トピックの境界を検出するために、質問文の類似度を利用する手法<sup>12)</sup>も提案されているが、音声入力は質問文が短くなりがちであるため、このような手法を適用することは困難である。そこで本研究ではトピックとして、文書知識ベースにおいて一般的に利用可能なメタデータであるタイトル情報を利用する。すなわち、応答を生成するもととなった文書を記憶・追跡す

#### 慈照寺

##### (概要)

慈照寺は、京都府京都市左京区にあり、東山文化を代表する臨済宗相国寺派の寺院。通称銀閣寺、山号は東山。開基は、室町幕府 8 代将軍の足利義政、...

##### 沿革

室町幕府 8 代将軍足利義政は、1473 年、嗣子足利義尚に将軍職を譲り、1482 年から、東山の月待山麓に東山山荘の造営を始めた。この地は、...

##### 境内

錦鏡池を中心とする池泉回遊式庭園「苔寺」の通称で知られる西芳寺庭園を模して造られたとされるが、江戸時代に改修されており、...

図 5 Wikipedia の文書の例

Fig. 5 Example of Wikipedia document.

文書の例を図 5 に示す。

これらの文書は、Web ブラウザによりテキストの形式で閲覧されることを前提に作成されており、文体は書き言葉調であるため、音声合成によりそのまま読み上げると不自然になる。そのため、文末の助詞の変更、書き言葉特有の語彙の平易な表現への言い換えを行う<sup>8)</sup>ことで話し言葉調に変換した。本研究では、これらの文書を構文解析器 CaboCha により解析したものを知識ベースとして使用する。

ることにより文脈解析を実現する。

### 3.1.2 知識ベースの検索

ユーザ発話と知識ベース中の文書との類似度を計算するために、単語ベースのベクトル空間モデルを採用する。具体的には、知識ベースの文書の節を単位として、その自立語に対して、タイトルに重み付けをした出現回数に基づく文書ベクトルを作成する。たとえば、図5の例においては、概要、沿革、境内の説明に対して、それぞれベクトルが作られる。ユーザ発話に対しては、音声認識結果(3-best候補まで)・現在のトピックに関する以前のユーザ発話(コンテキスト情報)中の名詞に対して、事後確率で正規化された音声認識の信頼度<sup>13)</sup>を各単語の重みとして、検索クエリベクトルを作成する。これらのベクトルの内積により類似度を計算する。

ユーザ発話が検索要求である場合には、検索された文書の中から、ユーザ発話との類似度が最大の見出し(節)を提示する。ユーザの理解のしやすさを考慮して、文書中における文の出現位置と文間のつながりを手がかりとして重要文抽出による要約を行い、その結果を提示する。

なお、ユーザ発話の音声認識誤りに対処するために、文書のタイトルや見出しに出現する名詞の音声認識の信頼度<sup>13)</sup>が低い場合には、「金閣寺でよろしいでしょうか」のような確認を行う。

### 3.1.3 ユーザ発話の質問タイプ判定

ユーザ発話の質問タイプの判定には、人手によるヒューリスティックなルールを用いる。たとえば、認識結果中に「誰ですか」という表現が含まれる場合には、人名をたずねる質問であると判定し、「いくらですか」という表現が含まれる場合には金額をたずねる質問であると判定する。このように用意したルールにより、上記の質問タイプに加えて「場所」「日時」「長さ・高さ」「アクセス方法」の合計6種類の質問に対応する。また、「なぜですか」といったシステムが回答できない質問に対しても判定ルールを用意し、ユーザがこのようなタイプの質問をした場合には、システムが回答できないことをユーザに伝えて対処する。用意したルールのいずれにも該当しない場合には、発話は検索要求であると判定する。

### 3.1.4 質問に対する応答の生成

ユーザ発話が質問である場合には、知識ベースの中から回答箇所を抽出してユーザに回答する。本研究では、テキストベースの質問応答システムで利用される一般的な手法を実装した。具体的には検索された文書の中に含まれる質問タイプに対応する固有表現(NE)

ごとに以下の特徴量に関するスコアを求める。これらの値の積により回答抽出スコアを計算し、これが最大となるNEを質問への応答とする。

- ユーザ発話と文書間の類似度(3.1.2項)
- NEを含む文にユーザの発話した名詞が含まれる個数
- NEを含む文節に係る文節、およびNEを含む文節に係る文節にユーザの発話した名詞が含まれる個数

なお、「創立」と「創建」のような表現の揺れの問題に対処するために、岩波国語辞典第六版から作成した同義語表現リストにより語彙拡張を行っている。

このようにして得られた応答をユーザに提示する際に、テキストベースのシステムのように、回答となる箇所(人名や年など)のみを答えることも考えられるが、対話の継続性をなどを考慮して回答を含む文全体を読み上げる。

## 3.2 システム主導の情報推薦

ユーザからの検索要求・質問に答えるだけでなく、インタラクティブな情報推薦を行うために、システム主導で質問を生成することを提案する。質問集合は、知識ベースの文書から半自動的に生成する。また、これに加えて従来の情報推薦手法として、文書構造を利用した情報推薦と文書間類似度による情報推薦を考える。

### 3.2.1 システム主導の質問(手法1)

本システムではユーザの検索要求にマッチした文書を要約して提示しているが、要約の際に重要文として選ばれなかった箇所にも、有用な情報が含まれている可能性がある。このような部分をユーザに提示する方法として、たとえば「より詳細な説明をしましょうか?」といったプロンプトを出すことも考えられるが、ユーザの現在の話題に対する興味を引き出すために、システム主導でユーザに質問をする形式を考える。

具体的な処理としては、事前に質問応答(3.1.4項)の回答候補となる固有表現(NE)を含む文を対象に、以下の手順で質問文に変換しておく。

- (1) 知識ベースの文書に含まれるNEを対象に  $tf * idf$  スコアを計算し、そのスコアが高いNEをユーザが興味を持ちそうな箇所と考えて、提示の対象語句として抽出する。
- (2) 文に含まれるNEを、対応する疑問詞に置き換える。
- (3) 構文解析の情報を利用して、NEを含む文節に係る文節、NEに係る動詞を含む文節と、その動詞に係る文節以外の文節を削除する。

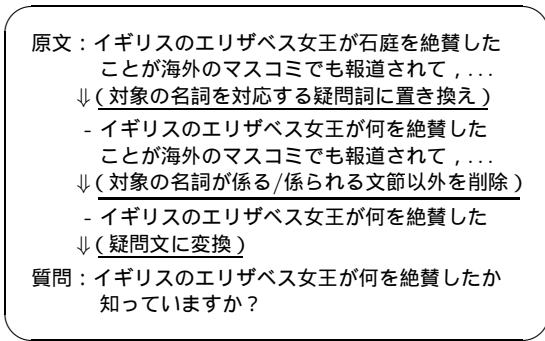


図 7 システム主導の質問の生成例

Fig. 7 Example of system-initiative question generation.

#### (4) 文末表現を質問形にする．

この手法により知識ベース中の文を質問に変換した例を図 7 に示す．このようにして生成した質問から不自然なものを人手で修正・除去し，合計 495 個の質問を用意した．この中から，現在のトピックに対応する質問をユーザへの推薦の候補とする．

#### 3.2.2 文書の階層構造・文書間類似度を利用した情報推薦

システム主導の質問に加えて，一般的に用いられる情報推薦手法として，以下の 2 つの手法を実装した．

- 文書の階層構造を利用した情報推薦 (手法 2)  
 一般的に文書は，節などの見出しを用いて階層的に記述されていることが多い．本研究で扱う Wikipedia においても，図 5 に示すように，見出し語を用いて階層的に記述されている．このような文書の階層構造を利用して現在提示中の文書の未提示の節を推薦する．
- 文書間類似度を利用した情報推薦 (手法 3)  
 Web などの情報推薦においては，ユーザの表示履歴情報を利用して，ユーザの興味に近い文書を推薦することが一般的に行われている．このような情報推薦における知見を利用して，現在提示中の文書と類似度が高い文書を推薦する．

### 4. フィールドテストとデータの分析

提案する対話の枠組みを京都大学総合博物館企画展「コンピュータに感覚を」の展示システムとして実装した．実装に際して，奈良先端科学技術大学院大学で開発された音声対話システム「たけまるくん」<sup>7)</sup>をベースに情報検索・質問応答・情報推薦の機能を追加・拡張した．ユーザ発話を音声認識するための言語モデルとして，文献 14) の手法により大量の Web テキストから N-gram 言語モデルを作成した．システムの利用者

(話者の顔認識に成功)

- S1: さんですよ！こんにちは，私は京都の観光案内をしているモモコと申します．京都の観光地について何か知りたいことはありませんか？  
 U1: 龍安寺について教えてください．  
 S2: 龍安寺は，京都市右京区にある臨済宗妙心寺派の寺院です．山号を大雲山と称します．本尊は釈迦如来，創立者は細川勝元，開山は義天玄承です．  
 (沈黙を検出)  
 S3: では，イギリスのエリザベス女王がこの寺の何を絶賛したか知っていますか？  
 U2: 教えて．  
 S4: イギリスのエリザベス女王が 1975 年に公式訪問した際に，石庭を絶賛したことが海外のマスコミに取り上げられて，一躍有名になりました．  
 U3: どうやって行くの？  
 S5: 龍安寺は市バスの龍安寺前からすぐです．  
 U4: ありがとう．  
 S6: どういたしまして．

図 8 システムとの対話例

Fig. 8 Example dialogue.

は，企画展期間中に博物館を訪れた一般の来館者であり，子供から高齢者までの幅広い年齢層の話者である．利用に際して，ユーザには顔画像認識を行うために，しばらくロボットの目(カメラ)を見る必要があることを記した注意書きを提示した．音声対話部に関しては，発話例(「          について教えて」、「          はいつですか」のみ)を提示した以外には，特別な教示は与えていない．顔画像認識による名前の呼びかけを含むシステムによる対話例を図 8 に示す．

#### 4.1 データの収集結果

特別展開催期間の 2006 年 6 月 7 日から 8 月 27 日までの約 3 カ月間運用を行い，2,564 対話を収集した．図 9 に 1 週間ごとの対話数(のべ利用者数)を示す．8 月に入り，利用者が増加しているのは，夏休みに入り，入館者が増加したためである．システムの利用者は 1 日平均で約 60 名であった．これは，博物館の入場者が 1 日平均 150 名程度であり，本企画展のほかにも常設展やイベントが行われていたことを考慮すると，企画展を訪れた入場者の多くが本システムを利用したと考えられる．

このうち，6 月 21 日から 7 月 21 日までの 23 開館日を対象に，顔画像認識を用いた声かけの効果に関する分析を行った．声かけのみが行われたユーザを含むこの期間中のシステムののべ利用者は 797 名であった．ユーザが利用したシステムの機能の詳細を図 10 に示す．多くのユーザが声かけ，対話の両方の機能を利用していることが分かる．また，声かけをされたユーザのおよそ 3 名に 2 名が引き続きシステムとの対話を行っており，「対話のみ」には顔認識に失敗して声かけ

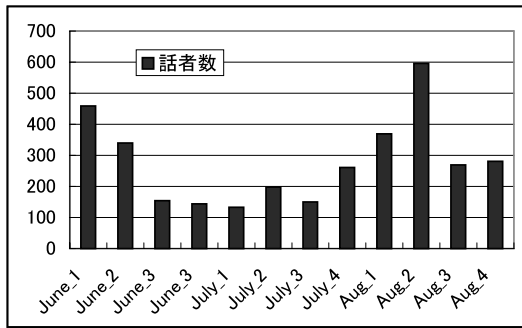


図 9 1 週間ごとの対話数  
Fig. 9 Number of collected dialogue sessions.

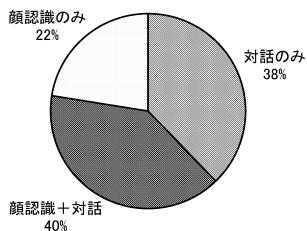


図 10 ユーザが利用したシステムの機能  
Fig. 10 Breakdown of system functions used by users.

を行っていない場合も含まれることを考慮すると、顔認識機能が対話への導入として有効に機能していると考えられる。

#### 4.2 発話内容の分析

次に、6月7日から7月21日までの33開館日を対象に（無意味な入力なども含む）すべての発話を書き起こし、発話内容の分析を行った。収集された対話数は1,497である。

これらの対話中に含まれる発話数は合計9,020であった。この内訳を表2に示す。表中の“検索要求”は、本システムの主なタスクである京都の観光案内に関する検索要求（質問を除く）である。“質問”は「金閣寺はいつできましたか?」といった、京都の観光地に関する具体的な質問である。“受理/拒否”は、システムの生成する確認や情報推薦に対する「はい」「お願いします」といったユーザの回答である。“あいさつ”には「こんにちは」といったロボットに対するあいさつや「何歳ですか」といった、ロボット自身に関する質問が含まれる。このような発話に対しては、あらかじめ用意したテンプレートにより対処を行う。“ドメイン外発話”は、以上のいずれにも該当しない発話であり、人が聞いても意味不明な発話も多い。“非音声”は、ロボットのサーボ音や周囲の雑音を拾ったもので、本来棄却することが望ましいものである。“あいさつ”に分類される発話が多いのは、本システ

表 2 ユーザ発話の内訳

Table 2 Breakdown of user utterances.

発話内容	比率
検索要求	18.8%
質問	7.0%
受理/拒否	23.8%
あいさつ	22.4%
ドメイン外発話	17.4%
非音声	10.6%

ムで用いたロボットがユーザにとって親しみやすいものであったためと考えられる。

これらの対話中の48.2%にあたる722対話に京都観光案内に関する発話が含まれていた。また、質問を含む対話は全体の19.0%にあたる285対話であった。なお、京都観光案内に関する発話に対する音声認識精度（単語認識精度）は70.6%であった。

#### 5. 情報案内対話の性能評価

前節で分析したデータの中で、京都観光案内に関する対話の部分について、詳細な評価を行った。

##### 5.1 質問応答の性能評価

まず最初に、ユーザの質問に対する応答の成功率の評価を行った。質問応答はMRR (mean reciprocal rank) により評価されることが多いが、音声で複数候補を提示することが困難なため、本研究では単純に第1候補の成功率により評価を行った。知識ベース中にユーザの質問の回答となる箇所が存在する場合には、該当箇所を含む文を提示できた場合に応答成功とした。また、回答箇所が存在しない場合には、質問の回答がないことをユーザに伝えられた場合に質問応答成功とした。評価の結果、質問応答の成功率は、61.4%（回答が存在する場合で64.0%、存在しない場合で46.8%）であった。なお、失敗の主な原因は、固有表現（NE）や質問タイプ判定に用いる表現の音声認識誤りである。

次に、音声認識の信頼度を用いることの効果も調べた。提案システムでは、ユーザ発話と知識ベースとのマッチングの際に信頼度を重みとして利用している。表3に示すように、信頼度を利用することによる成功率の改善が確認された。

また、音声認識結果の3-best候補をマージして用いることによる効果も調べた。比較対象として、1-best候補のみを用いる場合（baseline）、3-bestの中から最適な候補を選択する場合との比較を行った。この結果を表4に示す。1-bestのみ候補のみを用いる場合と比較して、3-best候補をマージしてクエリを作成することによる成功率の改善が見られた。ただし、3-best候補で別々に検索を行い、最適なものを選択すること

表 3 音声認識信頼度利用と質問応答成功率の関係

Table 3 Effect of using confidence measure of speech recognition.

音声認識信頼度の有無	成功率 (%)
あり	61.4 (64.0, 46.8)
なし	59.2 (61.0, 49.4)

表 4 N-best 候補利用と質問応答成功率の関係

Table 4 Effect of using N-best hypotheses.

N-best 候補の利用	成功率 (%)
3-best 候補のマージ (proposed)	61.4 (64.0, 46.8)
1-best 候補のみ (baseline)	58.2 (61.7, 39.0)
最適候補の選択 (reference)	65.1 (68.4, 46.8)

表 5 コンテキスト利用と質問応答成功率の関係

Table 5 Contextual effect on QA.

コンテキスト利用	成功率 (%)
現在のトピック (proposed)	61.4 (64.0, 46.8)
コンテキストなし	39.2 (33.3, 72.7)
直前の 1 発話のみ	55.9 (56.1, 54.5)
直前の 5 発話	57.5 (58.7, 50.6)
ユーザの全発話	50.0 (50.3, 48.1)

で、さらに高い成功率が得られることも分かった。この成功率は、3-best 候補の選択による上限値であり、著者らが文献 15) において提案している検索後の確認を生成することによって得られるものと考えられる。

さらに、検索において利用するコンテキスト長 (= 以前のユーザの発話数) に関する評価を行った。この結果を表 5 に示す。コンテキストの付与を行わない場合には、成功率が著しく低下している。一方、ユーザの全発話を利用した場合にも、成功率が低下しているため、必要以上の長さを付与することが悪影響を与えることが示された。したがって、提案するトピック同定に基づくコンテキスト付与の有効性が示された。

## 5.2 システム主導の推薦の評価

提案するシステム主導の情報推薦における各手法の効果を確かめるために、推薦に対するユーザの応答を調べた。3.2 節で述べた 3 つの手法による推薦候補をランダムに生成した 7 月 8 日以降の対話において、システムによる情報推薦は 268 回行われた。ユーザがシステムの提案に沿って情報提供を受けた場合に、推薦内容を受理したものとして評価した。すなわち、システム主導による質問 (手法 1) に関しては、ユーザが質問の回答を答えるか回答の提示を要求する発話を行った場合に受理されたものとし、他の手法ではユーザが推薦された文書の提示を求めた場合に受理と判定

表 6 システム主導の情報推薦の評価

Table 6 Comparison of recommendation methods.

推薦手法	受理率 (%)
質問 (proposed method 1)	75.6
文書構造 (method 2)	51.7
文書間類似度 (method 3)	37.3

した。この判定は人手により行った。推薦手法ごとの受理率を表 6 に示す。システム主導による質問 (手法 1) の受理率が他の手法と比較して有意に高い。これは、手法 1 がユーザに回答を答えさせるという点でインタラクティブ性が高く、ユーザの興味を引くものであったためと考えられる。

さらに、6 月 7 日から 7 月 21 日までの対話を対象にユーザ主導の検索・質問応答モードと、システム主導の情報推薦モードの遷移回数に関する分析を行った。検索・質問応答モードから情報推薦モードへの遷移は 581 回の対話で行われ、のべ回数は 1,635 であった。このうち、さらに、逆方向の遷移が行われたのが 248 対話で、のべ回数は 516 であった。これにより、システム主導の情報推薦を行うことで、ユーザの興味が喚起されて対話が長く継続していることが確かめられた。

## 6. おわりに

本論文では、音声为主要なモダリティである環境において、インタラクティブに文書の検索・提示を行う音声対話システムの提案を行った。インタラクティブに情報を提示するために、ユーザ主導の検索・質問応答モードと、システム主導の情報推薦モードを用意した。また、質問応答技術と顔画像認識による名前の呼びかけを用いて対話に引き込む仕組みを提案した。京都観光案内システムを構築し、博物館において長期間にわたる運用を行い、提案手法が有効に機能していることを確認した。

本論文では、ロボットエージェントを用いることがユーザの親しみやすさに与える影響に関する詳細な分析は行っていない。アンケートには「モモコと会話できて楽しかった」といった、ロボットの存在を意識した記述が多く見られたことから導入の効果はあったと思われるが、エージェントを利用しない場合との比較・評価は今後の課題である。

謝辞 本研究に対して多大な協力をいただいた奈良先端科学技術大学院大学教授鹿野清宏先生、京都大学教授黒橋禎夫先生、笹野遼平氏に深く感謝します。また、ロボット・フィノを提供していただいた独立行政法人情報通信研究機構、顔画像認識を提供していただいたオムロン株式会社に深く感謝します。



## 参 考 文 献

- 1) Zue, V., Seneff, S., Glass, J., Polifroni, J., Pao, C., Hazen, T.J. and Hetherington, L.: Jupiter: A telephone-based conversational interface for waeather information, *IEEE Trans. Speech and Audio Processing*, Vol.8, No.1, pp.85–96 (2000).
- 2) 堂坂浩二, 安田宜仁, 相川清明: システム知識制限下での効率的対話制御法, 自然言語処理, Vol.9, No.1, pp.43–63 (2002).
- 3) 安達史博, 河原達也, 奥乃 博, 岡本隆志, 中嶋 宏: Voice XML の動的生成に基づく自然言語音声対話システム, 情報処理学会研究報告, 2002-SLP-40-23 (2002).
- 4) Akiba, T. and Abe, H.: Exploiting Passage Retrieval for N-Best Rescoring of Spoken Questions, *Proc. Interspeech* (2005).
- 5) 神田崇行, 佐藤留美, 才脇直樹, 石黒 浩: 対話型ロボットによる小学校での長期相互作用の試み, ヒューマンインタフェース学会論文誌 (ソーシャルインタフェース特集号), Vol.7, No.1, pp.27–37 (2005).
- 6) 川本真一, 下平 博, 新田恒雄, 西本卓也, 中村 哲, 伊藤克亘, 森島繁生, 四倉達夫, 甲斐充彦, 李 晃伸, 山下洋一, 小林隆夫, 徳田恵一, 広瀬啓吉, 峯松信明, 山田 篤, 伝 康晴, 宇津呂武仁, 嵯峨山茂樹: カスタマイズ性を考慮した擬人化音声対話エージェントツールキットの設計, 情報処理学会研究報告, Vol.43, No.7, pp.2249–2263 (2002).
- 7) 西村竜一, 西原洋平, 鶴身玲典, 李 晃伸, 猿渡 洋, 鹿野清宏: 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-DII, No.3, pp.789–798 (2004).
- 8) Kaji, N., Okamoto, M. and Kurohashi, S.: Paraphrasing Predicates from Written Language to Spoken Language Using the Web, *Proc. HLT/NAACL* (2004).
- 9) Matsuda, M. and Fukumoto, J.: Answering Question of IAD Task using Reference Resolution of Follow-up Questions, *Proc. 5th NTCIR Workshop Meeting on Evaluation of Information Access Technologies*, pp.414–421 (2006).
- 10) Mori, T., Kawaguchi, S. and Ishioroshi, M.: Answering Contextual Questions Based on the Cohesion with Knowledge, *Proc. Computer Processing of Oriental Languages. Beyond the Orient: The Research Challenges Ahead (Proc. 21st International Conference on the Computer Processing of Oriental Languages)*, pp.1–12 (2006).
- 11) Murata, M., Utiyama, M. and Isahara, H.: Japanese Question-Answering System For Contextual Questions Using Simple Connection Method, Decreased Adding with Multiple Answers, and Selection by Ratio, *Proc. Asian Information Retrieval Symposium*, pp.601–607 (2006).
- 12) Boni, M.D. and Manandhar, S.: Implementing clarification dialogues in open domain question answering, *Natural Language Engineering*, Vol.11, No.4, pp.343–361 (2005).
- 13) 李 晃伸, 河原達也: 2パス探索アルゴリズムにおける高速な単語事後確率に基づく信頼度算出法, 電子情報通信学会技術研究報告, SP2003-160 (2003).
- 14) Misu, T. and Kawahara, T.: A Bootstrapping Approach for Developing Language Model of New Spoken Dialogue Systems by Selecting Web Texts, *Proc. Interspeech*, pp.9–12 (2006).
- 15) 翠 輝久, 駒谷和範, 清田陽司, 河原達也: 音声対話によるソフトウェアサポートタスクのための効率的な確認戦略, 電子情報通信学会論文誌, Vol.J88-DII, No.3, pp.499–508 (2005).

(平成 19 年 4 月 2 日受付)

(平成 19 年 7 月 3 日採録)



翠 輝久 (学生会員)

2003 年京都大学工学部情報学科卒業。2005 年同大学院情報学研究科知能情報学専攻修士課程修了。現在、同大学院情報学研究科知能情報学専攻博士後期課程在学中。2005 年から日本学術振興会特別研究員 (DC1)。音声対話システムの研究に従事。電子情報通信学会, 日本音響学会, 言語処理学会各学生会員。



河原 達也（正会員）

1987年京都大学工学部情報工学科卒業。1989年同大学院修士課程修了。1990年同博士後期課程退学。同年京都大学工学部助手。1995年同助教授。1998年同大学院情報学研究科助教授。2003年同大学学術情報メディアセンター教授。現在に至る。この間、1995年から1996年まで米国ベル研究所客員研究員。1998年からATR客員研究員。1999年から2004年まで国立国語研究所非常勤研究員。2001年から2005年まで科学技術振興事業団さきがけ研究21研究者。音声言語処理、特に音声認識・理解に関する研究に従事。京大博士（工学）。1997年度日本音響学会粟屋潔学術奨励賞受賞。2000年度情報処理学会坂井記念特別賞受賞。情報処理学会連続音声認識コンソーシアム代表。IEEE SPS Speech TC委員。言語処理学会理事。日本音響学会、人工知能学会各評議員。電子情報通信学会、IEEE各会員。



正司 哲朗

1997年龍谷大学理工学部電子情報学科卒業。2002年同大学院理工学研究科博士課程後期研究指導認定退学。2003年龍谷大学古典籍デジタルアーカイブ研究センター博士研究員。2004年京都大学学術情報メディアセンター博士研究員（科学技術振興）。教育支援システム、画像処理、講義のアーカイブ化に関する研究に従事。博士（工学）。日本教育工学会、電子情報通信学会、教育システム情報学会各会員。



美濃 導彦（正会員）

1978年京都大学工学部情報工学科卒業。1983年同大学院博士課程修了。同年同大学工学部助手。1987年から1988年までマサチューセッツ州立大学客員研究員。1989年京都大学工学部附属高度情報開発実験施設助教授。1995年同教授。1997年京都大学総合情報メディアセンター教授。2002年京都大学学術情報メディアセンター教授。画像処理、人工知能、知的コミュニケーション関係の研究に従事。工学博士。IEEE、ACM、画像電子学会、日本ロボット学会各会員。