

# 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ダンスのように、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回答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ヶ月の運用期間中、のべ 2500 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 MICHHIKO MINOH<sup>†,††</sup>

This paper addresses an interactive framework for information navigation with speech interface. In conventional audio guidance systems, such as those deployed in museums, the information flow is one-way and the content is fixed. In order to make an interactive guidance system, we propose the application of question-answering (QA) techniques. Since users tend to use anaphoric expressions in successive questions, we investigate appropriate handling of contextual information based on topic detection, together with the effect of using N-best information in ASR (automatic speech recognition) output. Moreover, we apply the QA technique to generation of system-initiative information recommendation. A navigation system on Kyoto city information was implemented. 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情報を探すのが困難になりつつある。とりわけ、出先などで速やかに情報が欲しい場合には、携帯電話などにより、場所を選ばずに利用可能な音声により検索可能であると

好都合である。また、音声は我々が人同士のコミュニケーションに用いる身近な存在であり、誰にでも使いやすいインターフェースとして注目を集めている。音声対話システムに関する従来の大抵の研究・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ほとんどであったが、近年の Web 検索に代表される情報検索技術の発達に伴い、より一般的な文書検索を対象とした研究が盛んに行われている<sup>4)</sup>。

しかし、これらの現状の音声による情報検索・案内システムに、人に何かを尋ねるような便利さがあるとは言いがたい。実際、一般的な自動電話応答サービスなどでは、基本的にユーザはシステムの示す手順に従っ

<sup>†</sup> 京都大学 情報学研究科  
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

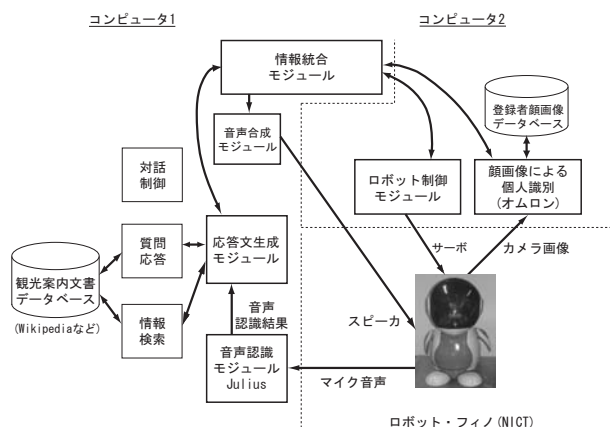


図 1 提案システムの構成

Fig. 1 System architecture

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、システムはユーザに一方的に情報を提示するだけであるため、聞き逃しをしたり、より詳細な説明を聞きたい場合であっても、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、システムを持つ知識にアクセスする仕組みを提案する。具体的なタスクとしては、ユーザに観光情報を案内するタスクを想定して、バックエンドとなる文書を効率的にユーザに伝えることを目指す。

しかし、このような情報案内を実現するためには、情報の提示方法が問題となる。音声により出力を行う場合には、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、知識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行うことを目指す。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づらい状況下において、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。

本研究ではさらに、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、ヒューマンロボット<sup>5)</sup>や、擬人化エージェント<sup>6)</sup>を利用したマルチモーダルなインター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、エージェントを用いない場合と比較して、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>。本研究ではこのようなエージェントの利用に加えて、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構築することを目指す。

以上の指針に基づいて、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し、約3ヶ月間の運用を行った。

以下に本論文の構成について述べる。まず、2章においてシステムの概要について述べる。次に3章において、提案システムの音声インターフェースの構成について説明する。続いて4章では、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、京都大学博物館において、来館者を対象に京都の観光案内を行うものである。また、テンプレートベースの応答を用意することにより、「こんにちは」といったエージェントに対するあいさつなどにも対処する。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。システムの構成を図1に、外観を図2に示す。

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

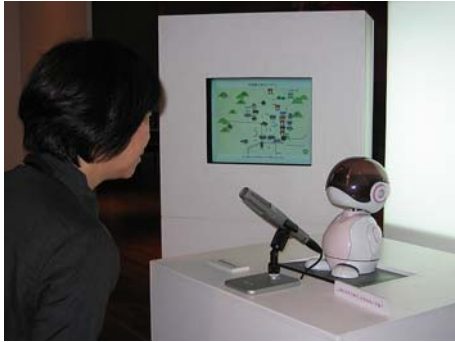


図 2 京都大学博物館でのシステムの利用風景  
Fig. 2 Field trial at the university museum

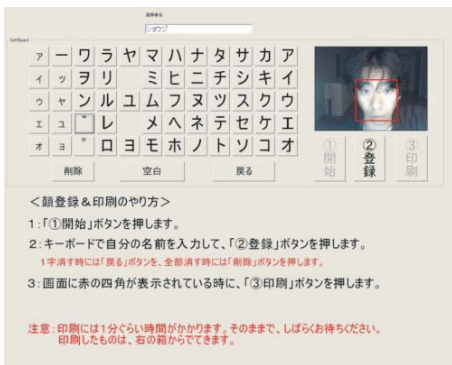


図 3 顔画像登録インターフェイス  
Fig. 3 Face Registration Interface

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 箇所を動作させることが可能であり、これらの部位を動作させることによる 21 種類の動作パターンをもつ。写真左のディスプレイは、提示中の情報に関係した (観光地などの) 写真の提示を行うためのものであり、初期状態では京都の観光マップを表示している。

2.1 画像情報を利用したユーザへの声かけ

本研究では、ユーザを認識するために顔画像認識を利用する。以下に、ユーザの顔登録から声かけまでの流れを示す。

ユーザは最初に企画展入り口に設置された登録端末を利用して、顔画像と名前をデータベースに登録する。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。この理由は、顔画像認証において、ユーザの表情や顔の向きが変化するため、より多くの顔画像を登録することにより、ユーザの顔の変化に対応するためである。しかしながら、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

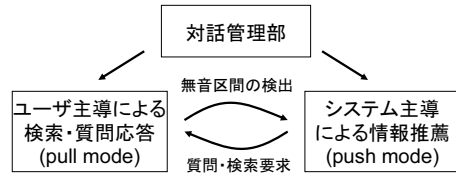


図 4 提案システムの対話の概要  
Fig. 4 Overview of dialogue

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、ユーザの顔が検出し易い入力インターフェイスを設計し、ユーザの名前と顔画像の登録を行う。

次に、ユーザが会場内のシステムの前に来た際に、ユーザの名前を呼び、声かけを行う。ユーザの認識はロボットに内蔵されているカメラを利用してユーザの撮影・顔検出を行い、顔画像データベースを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。これは、1 回のみでの認証では、認識精度が低下する恐れがあるためである。

なお、システムの顔検出・認識処理は、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) が明確に捉えられるように照明環境を整えた。

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り替えを行う。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、発話の意図にふさわしい応答を生成する。また、システム発話後から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す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で閲覧す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

テキストの提示は行わない

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

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 | 1219    | 約 22 万 |

### 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

げるのは不自然になる。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 により構文解析したものを知識ベースとして使用する。

### 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二つのカテゴリに分類して扱う。一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」という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。もう一つは質問であり、「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 に示す。

##### 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

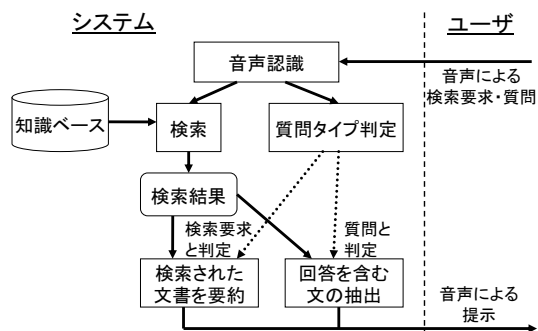


図 6 情報検索・質問応答処理の概要  
Fig.6 Overview of document retrieval and QA

するために、コンテキスト情報の補完は重要な課題である。質問文に対して照応解析を行う方法<sup>9)</sup>も提案されているが、これは容易ではなく、十分な精度は得られていない。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を考える。しかし、単純に全てのユーザの発話を利用すると、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与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。

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、応答を生成する基となった文書を記憶・追跡す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

本研究ではユーザの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、知識ベース中の文書の節を単位として、知識ベース中の自立語に対して、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)・現在のトピックに関するユーザの発話(コンテキスト情報)中の名詞に対して、事後確率に基づく音声認識の信頼度<sup>10)</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。

ユーザの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

音声認識の信頼度は 0~1 までの値をとり、本研究ではこの値をそのまま単語ベクトルの重みとして用いる

して、文書中での文の出現位置と文間のつながりを手がかりに重要文抽出による要約を行い、その結果を提示する。

なお、ユーザの発話の音声認識誤りに対処するために、音声認識の信頼度が低い場合には、文書のタイトルや見出しを利用して「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も、システムが質問に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 質問に対する回答の生成

ユーザの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)毎に以下の特徴量に対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、それが最大となるNEを質問への回答とする。

- ユーザ発話と文書間の類似度 (3.1.2 節)
  - 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、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、国語辞典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回答をユーザに説明する際に、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

インタラクティブな情報推薦を行うために、システム主導の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という形式を考える。

具体的な処理としては、事前に質問応答(3.1.4 節)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対して名詞を対象に  $tf*idf$  スコアを計算し、そのスコアが高いNE(=ユーザが興味を持ちそうな箇所)を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。
- (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。

原文: イギリスのエリザベス女王が石庭を絶賛したことが海外のマスコミでも報道されて、...  
 ↓(対象の名詞を対応する疑問詞に置き換え)  
 - イギリスのエリザベス女王が何を絶賛したことが海外のマスコミでも報道されて、...  
 ↓(対象の名詞に係る/係られる文節以外を削除)  
 - イギリスのエリザベス女王が何を絶賛した  
 ↓(疑問文に変換)  
 質問: イギリスのエリザベス女王が何を絶賛したか知っていますか?

図7 システム主導の質問の生成例

Fig. 7 Example of system-initiative question generation

### 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。

U 1 : 銀閣寺について教えてください。  
 S 1 : 銀閣寺は京都市左京区にある東山文化を代表する臨濟宗相国寺派の寺院です。通称は銀閣寺、山号は東山です。銀閣寺の歴史やみどころについて何か知りたいことはありませんか？  
 U 2 : いつ建てられましたか？  
 S 2 : 銀閣寺の建立は 1482 年です。  
 (沈黙を検出)  
 S 3 : では、銀閣寺は誰が建てたか知っていますか？  
 U 3 : わかりません。  
 S 4 : 銀閣寺の創立者は室町幕府 8 代将軍の足利義政、開山は夢窓疎石とされています。  
 U 4 : どうやって行けばいいですか？  
 .....

図 8 システムとの対話例 1  
 Fig.8 Example dialogue 1

- 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。
- 文書類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。

以上の対話戦略を実装したシステムによる対話の例を図 8 に示す。

#### 4. フィールドテストによるデータの収集

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を行った。実装に際して、奈良先端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・情報推薦の機能を拡張した。ユーザの発話を音声認識するための言語モデルとして、著者らが文献<sup>11)</sup>において提案した手法により大量の 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図を提示した。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 9 に示す。

(話者の顔認識に成功)  
 S 1 : 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？  
 U 1 : 龍安寺について教えてください。  
 S 2 : 龍安寺は、京都市右京区にある臨濟宗妙心寺派の山号を大雲山と称します。本尊は釈迦如来、創立者は細川勝本、開山は義天玄承です。  
 S 3 : 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？  
 U 2 : 教えてください。  
 S 4 : 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。  
 .....

図 9 システムとの対話例 2  
 Fig.9 Example dialogue 2

#### 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ヶ月間の運用を行い、2,564 対話を収集した。図 10 に一週間毎の対話数(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と考えられる。システムの利用者は一日平均で約 50 名であった。これは、博物館の入場者が一日平均 150 名程度であり、本企画展の他に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。さらに、顔画像認識を利用した声かけの回数を図 10 に併記する。システムが生成した呼びかけの回数は、対話数とほぼ同じであり、この機能が積極的に利用されていたことが確認された。

#### 4.2 発話内容の分析

次に 6 月 17 日から、7 月 7 日までの 11 開館日を対象に(無意味な入力なども含む)全ての発話を書き起こし、発話内容の分析を行った。この期間中に収集された対話数は 465 回で、この中の 52.7%にあたる 245 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 23.4%にあたる 109 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は 70.6%であった。

次に、発話内容の詳細な分析を行った。前述の対話中に含まれる発話数は合計 2930 回であった。この内訳を表 2 に示す。表中の「検索要求」は、本システムのメインピッ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

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

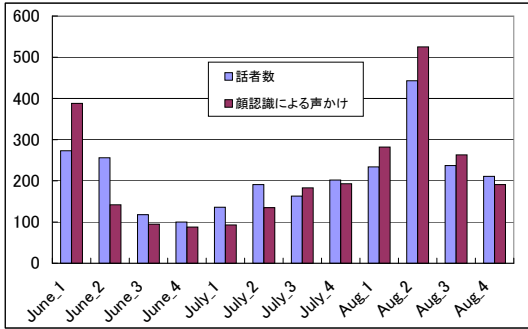


図 10 利用者数・顔画像認識利用回数  
Fig. 10 Number of collected dialogue

表 2 ユーザ発話の内訳  
Table 2 Specification of user utterances

| 発話内容    | 比率 (サンプル数)  |
|---------|-------------|
| 検索要求    | 17.5% (513) |
| 質問      | 8.0% (234)  |
| 受理/拒否   | 8.0% (672)  |
| あいさつ    | 23.3% (684) |
| ドメイン外発話 | 15.2% (444) |
| 非音声     | 13.1% (383) |

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、以上のど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、顔画像認識による名前の呼びかけを行うエージェントが、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。

## 5. 観光案内対話の性能評価

### 5.1 質問応答の性能評価

まず最初に、366回の質問により質問応答の成功率の評価を行った。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質問応答成功とした。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、60.7% (回答が存在する場合で 62.9%, 存在しない場合で 47.2%)であった。

次に、音声認識の信頼度を用いることによる効果を調べた。提案システムでは、ユーザの発話と知識ベー

質問応答は MRR(mean reciprocal rank) により評価が行われることが多いが、音声で複数候補を提示することは困難なため単純な成功率により評価を行った。

表 3 音声認識信頼度利用と QA 成功率の関係

Table 3 Effect of using speech recognition confidence measure

| 音声認識信頼度の有無 | QA 成功率 (%)        |
|------------|-------------------|
| あり         | 60.7 (62.9, 47.2) |
| なし         | 55.7 (54.0, 66.0) |

表 4 N-best 候補利用と QA 成功率の関係

Table 4 Effect of using N-best hypotheses

| N-best 候補の利用             | QA 成功率 (%)        |
|--------------------------|-------------------|
| 3-best 候補のマージ (proposed) | 60.7 (62.9, 47.2) |
| 1-best 候補のみ (baseline)   | 57.9 (61.0, 39.6) |
| 最適候補の選択 (reference)      | 63.1 (65.8, 47.2) |

表 5 コンテキスト利用と QA 成功率の関係

Table 5 Contextual effect for QA

| コンテキスト利用           | QA 成功率 (%)        |
|--------------------|-------------------|
| 現在のトピック (proposed) | 60.7 (62.9, 47.2) |
| コンテキストなし           | 36.9 (30.4, 75.5) |
| 直前の 1 発話のみ         | 54.6 (54.3, 56.6) |
| 直前の 5 発話           | 55.7 (56.9, 49.1) |
| ユーザの全発話            | 55.5 (56.5, 49.1) |

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 に示すとおり、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を調べた。比較対象として、1-best 候補のみを用いる場合 (baseline), 3-best 候補それぞれで回答抽出を行い多数決を行う場合、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、3-best 候補で別々に検索して多数決をとる場合と比較して、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で、さらに高い成功率が得られることが示された。この成功率は、3-best 候補の選択による上限値であり、著者が文献<sup>12)</sup>において提案した検索後の確認を生成することによって得られると考えられる。

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) の評価を行った。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成功率に悪影響を与えることが示された。従って、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質問の効果を確かめるため

現在のシステムには、この確認は実装されていない

表 6 システム主導の情報推薦の評価  
Table 6 Comparison of recommendation method

| 推薦手法                   | 受理率 (%) |
|------------------------|---------|
| 質問 (Proposed method 1) | 74.7    |
| 文書構造 (Method 2)        | 51.1    |
| 文書間類似度 (Method 3)      | 30.8    |

に、利用可能な推薦候補をランダムに生成した。システムが生成した情報推薦は 427 対話中で 319 回であった。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。推薦手法毎の受理率を表 6 に示す。システム主導による質問 (手法 1) 受理率が他の手法と比較して高い。これは、手法 1 がユーザに回答を答えさせるという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であると考えられる。

さらに、システム主導の検索・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、185 回の対話で行われ、のべ回数は 420 であった。逆方向の遷移は、122 回の対話において 297 回行われた。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

本稿では、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・顔画像認識による名前の呼びかけを用いてインタラクティブな情報提示を行う枠組みを提案した。観光案内システムを構築し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。

## 謝 辞

本研究に対して多大な協力を頂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、及び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. on 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, 情報学研報, 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, 伊藤克巨, 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. of HLT/NAACL* (2004).
- 9) Matsuda, M. and Fukumoto, J.: Answering Question of IAD Task using Reference Resolution of Follow-up Questions, *Proc. the Fifth NTCIR Workshop Meeting on Evaluation of Information Access Technologies*, pp.414-421 (2006).
- 10) 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).
- 11) 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).
- 12) 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).