

同調的対話を実現するプロトタイプシステムの開発

柏岡 秀紀 翠 輝久 大竹 清敬 堀 智織 中村 哲

独立行政法人 情報通信研究機構
ATR 音声コミュニケーション研究所
Email: hideki.kashioka@nict.go.jp

現在、我々は京都観光案内をタスクとする同調的対話システムの研究開発を進めている。音声を中心とした対話システムのプロトタイプを構築しており、可搬性を重視したシステムと画像情報を利用した大画面システムの2つのシステムがある。本稿では、各システムについて紹介するとともに、同調的対話についても議論する。また、対話システム開発のために、収集している京都観光案内対話コーパスについても紹介する。京都観光についてのエキスパートガイドとの対話を収録したものであり、対面、非対面、WOZ形式の対話を収録している。収録している対話に対し、対話行為タグを付与し、対話制御モデルの学習に利用することを検討している。

[キーワード] 対話システム、同調的対話、コーパス

Development of dialog system keeping step with users

Hideki KASHIOKA Teruhisa MISU Kiyonori OHTAKE Chiori HORI Satoshi NAKAMURA
National Institute of Information and Communications Technology
ATR Spoken Language Translation Research Labs.

We are developing two types of dialog system for Kyoto sightseeing task. Our development systems are proto-type system that mainly used speech information. One of our system is mobile type system that works in any place. And the other of our system is multi-modal system that has 52-inch display. We describe our proto-taype systems in this paper. And also we show our collected dialog corpus about Kyoto sightseeing dialogs. We collected these dialogs in three ways: 1) dialog via face-to-face, 2) dialog through lines with speech and information display (without face image), 3) wizard of oz style.

[keyword] Dialog System, Dialog Corpus,

1 はじめに

ネットワークの発展により多種多様な情報が多様な環境で利用可能になってきている。現在、多種多様な情報を提供する端末とのコミュニケーション

は、キーボードを利用した入力とディスプレイへの情報の表示が、主なインターフェイスである。しかしながら、千差万別の利用者の要求を満たすには、いまだ利用者と情報端末とのインターフェイスが大きな障害となっている。いつでもどこでも誰とでもコミュニケーションを行える基盤技術の一つとして

現在、我々は、人間の最も自然なインターフェイスの一つである音声を中心とした同調的対話の実現を目指し、京都観光を対象にプロトタイプを構築している。開発中のプロトタイプシステムは、音声を中心とした入力と対話システム内部での処理、利用者への情報提供としての出力という主に3つの要素から構成したシステムとなっている。実際に構築しているシステムは、可搬型としての機能を重視したシステムと音声以外のインターフェイスとの統合を重視したシステムの2つのシステムである。

また、対話システムの構築のために、対象となる対話のデータ収録を行っている。一日の京都観光の計画を立てることをタスクとして、実際に観光業務を行っているプロのガイドと利用者である観光旅行者との対話を複数の環境で収録している。

本稿では、現在構築しているプロトタイプシステムについて、大画面ディスプレイ対話システム、および可搬型対話システムの概要を述べるとともに、収録している京都観光案内対話コーパスについて紹介する。

2 同調的対話

対話システムには、テキストでチャットするような対話システムや、音声入出力による音声対話システムが考えられる。音声対話システムでは、入力された音声発話を理解し、システムが適切に応答することが期待される[3]。しかし、現実の対話では、音声のみの入力で円滑に自然な対話が行われているのではない。様々な情報に反応することで、各話者が相手話者が理解してくれており、積極的に対話しようとする状況での対話を、ここでは同調的対話と呼ぶ。このような状況では、より円滑な対話が成り立つと考えている。ここで考えられる同調性は、多様な側面を持っている。相づちや身振りなどによる対話の自然さもその一面として捉えることができる[1, 2]。この同調性をいくつかの侧面で分けてみると、以下のようなものが考えられる。

- 動作タイミングによる同調性
「あいづち」や「うなずき」、発話のオーバラップなどに見られる同調性
- 表層表現による同調性
相手に応じて表現をかえことや表現のくだけ方がかわるなどの同調性

- 対話制御(戦略)による同調性
提示する情報の内容と順序などによる同調性
- 信念共有による同調性
共有知識、信頼性などによる同調性

これらの同調性は、独立なものではなく、相互に複雑に関連しており、また、いずれかの同調性が満たされたからといって、対話全体が円滑に自然になるものでもない。しかしながら、個々の特徴的な同調性を特定の状況下で実現し、統合的な処理機構を対話制御として行うこと、人同士の同調的な対話の一部をシステムとの間で模倣することが可能と考えられる。

3 プロトタイプシステム

現在開発しているプロトタイプシステムの概略を図2に示す。音声を中心としたインターフェイスによる対話システムとして開発しているが、画像・動画処理など、音声以外のインターフェイスによる入出力を含むシステム構成となっている。また、システムの知識として、Wikipediaに含まれる京都に関するページを処理しらかじめDB化し、処理できるようにしている。さらに、一般のWebPageの信憑性・信頼性を示す検索機構であるWISDOMを用いてBlogやSNSなどに書き込まれている評判情報を利用できるようにしている。旅行者の発話から抽出される検索キーワードに関連するキーワードの連想検索も行っている¹。この検索された連想キーワード列をリスト表示することにより、普段関連づけていないキーワードや、思いもよらぬキーワードを捉えることができ、対話における話題の広がりを得ることができる。あらかじめDB化した情報や評判情報以外の情報に関しては、一般的なWEB検索を利用して、何らかの情報を利用者に提供できるようしている。

システムを開発するにあたり、タスクは、京都観光に関する対話を取り上げた。²現在のプロトタイプシステムは、履歴の管理を行いつつ基本的には一問一答を行うものとなっている。しかしながら、実際に、観光計画をたてるには、様々な知識処理、

¹京都検定に関するテキストを利用している

²我々が旅行タスクでの音声認識技術を持つていること、観光情報としてのコンテンツの利用が可能と思われるところから対象とした。

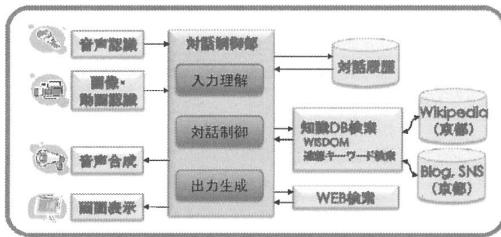


図 1: プロトタイプシステムの概略図

多様な対話行為を実現する基本的なシナリオの組み合わせによる状態遷移を制御する必要がある。このような制御機構を実現するため、図 2 に示す対話制御部を現在構築中である [5]。この図に示される Scenario を重み付き有限状態トランスデューサ (Weighted Finite-State Transducer: WFST) により記述し、システムの内部状態と入力信号による状態遷移によって対話を制御する。

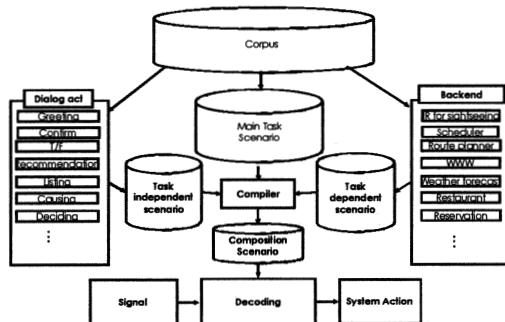


図 2: 対話制御機構

実際には、以下の 2 つのタイプの対話システムを構築している。

1. 可搬型システム

“いつどこでも”を実現するために小型で持ち運びできるシステムであり、主として音声によるインターフェイスを実現している。

2. マルチモーダルシステム

多様な利用者の発信している情報を利用し、音声以外の入出力情報を統合することを目的とした対話システムである。

以下各システムについて述べる。

可搬型システム

VAIO-U を利用したシステムであり (図 3)、システムへの入力は、音声を主としたもので、ディスプレイへの出力および音声による応答により、対話を実現する。音声入力は、Push-To-Talk のスタイルを利用している。システムからの情報提示に利用している画面情報は、画面左側に、利用者の発話に含まれていた検索キーワードおよびシステムの発話内容、検索キーワードから連想される連想キーワード列、評判情報がそれぞれ表示される。画面右側では、検索キーワードによって検索された WEB-Page を表示している。対話の履歴を管理することで、一部、検索の絞り込みをおこなっている。発話に含まれる検索キーワードに、事前に準備された DB の主要キーワードが含まれていなければ、それまでの主要キーワードを引き継ぎ、検索を行うようになっている。より適切な検索キーワードの管理を行うには、基本シナリオに従ったキーワード等の対話状況の管理が必要である。現在開発中の対話制御機構においては、WFST の状態に応じたキーワードの管理が実現できると考えている。

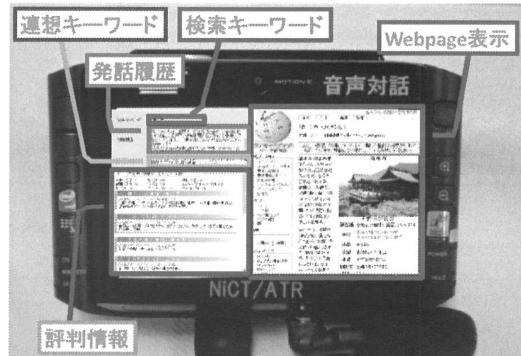


図 3: 可搬型システム

マルチモーダルシステム

マルチモーダルシステムでは、52 インチ大型ディスプレイを使用し、ディスプレイの左右および下部に計 3 台のカメラを設置し、正面に立った利用者を捉え顔や視線の方向を推定し、その情報を対話制御に利用している (図 4)。また、利用者に画面上の注

目すべき場所を明示し、システムとの対話をより円滑に進めることを目的として、ガイドエージェントを表示し、いくつかの動作をさせている。顔向き・視線の情報により、利用者が表示画面のどの部分に興味を持っているかを判断し、一定時間以上注視しているようであれば、システムから注視している画面に表示されている情報の詳細な情報を画面に表示し、音声でも詳細な説明を必要とするか問い合わせるようにしている。音声入力には、ディスプレイ上部の指向性マイクを利用している。

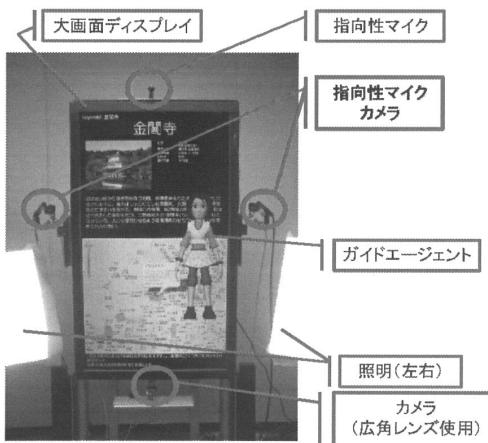


図 4: マルチモーダルシステム

4 コーパス収集

京都観光案内対話コーパスは、京都観光案内のエキスパートガイド3名（男性1名、女性2名）が模擬旅行者に対して京都市内一日観光の計画立案を行う2者による対話である[4]。今までに、対面での対話、非対面での対話、Wizard of Oz(WOZ)形式での対話を収録している。1対話は約30分である。

対面での対話では、ガイドが、ガイド自身の持つ知識、準備されているガイドブック、地図、WEB上の情報を利用し、旅行者に対して情報を提供、一日の旅程を作成していくものである。ガイドは、ヘッドセットマイクを使用し音声を収録していた。旅行者は、スタンダードマイクあるいはヘッドセットマイクのいずれかを用いて音声を収録している。旅行

者は、20歳代から50歳代の114名（男性57名、女性57名）を対象として収録した。

非対面での対話では、ガイドと旅行者の間での情報の授受は、音声およびディスプレイ上の表示に限定されている。音声は、ガイド、旅行者とともに、ヘッドセットマイクを使用して収録した。非対面の対話では、対面対話の収録に参加したガイド1名（女性）に限定して収録した。旅行者は、20名を対象としており、収録時間は約10時間となる。

WOZ形式での対話では、旅行者に情報を提示するディスプレイ上に、ガイドエージェントを表示した場合と表示しない場合の対話を収録している。ガイドエージェントを表示した画面のサンプルを図5に示す。旅行者は、表示／非表示のときともに20名を対象としている。非対面同様、収録時間は、表示／非表示、各々約10時間となる。WOZ形式での収録では、システム側に、ガイド、およびタイピストを用いて収録した。システムの発話は、一旦ガイドが発話したものをタイピストがテキストとして入力し、その入力テキストを音声合成したものである。そのため、ガイドが直接応答している対話に比べ、システムの応答までに時間がかかる。

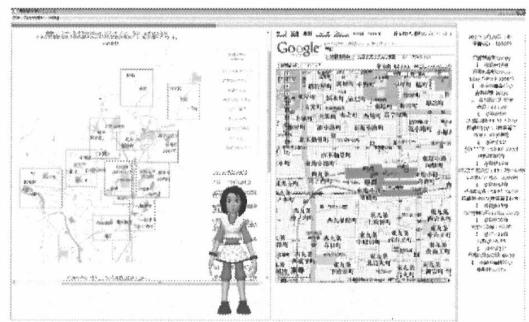


図 5: WOZにおける画面例

現在、これらデータの書き起しデータに対して、形態的なタグを始め、対話行為に対するタグ付けを行っており、対話コーパスからのWFSTの学習に利用する予定である。

5まとめ

本稿では、現在我々がプロトタイプとして開発・構築している可搬型システム、およびマルチモーダ

ルシステムについて述べた。これらのプロトタイプシステムは、個別に開発しているのではなく、開発している対話制御機構が、システムの利用できる多様な入力情報を統一的に適切に処理できることを示すとともに、多様なシステムとしての動作を確認し、様々な入出力情報を利用できるような実験環境を整えることを目的として開発している。

また、対話システムを開発するために収録している京都観光案内対話コーパスの概要について述べた。対面対話が約 50 時間、非対面対話が約 10 時間、WOZ 形式の対話が 20 時間（ガイドエージェント有: 約 10 時間、無: 約 10 時間）のコーパスとなっている。今後は、収録コーパスの分析を進め、頑健な対話を実現するための機構、および同調的対話を実現するための機構の研究開発を行う予定である。

6 謝辞

本稿で紹介しているプロトタイプシステムでは京都大学河原研究室の研究成果を、マルチモーダルシステムでは京都大学松山研究室の研究成果を移転・活用しています。また、ガイドエージェントでは、情報通信研究機構 ユニバーサルシティグループの研究成果を移転・活用しています。さらに、評判情報の検索 (WISDOM)、連想キーワードの処理は情報通信研究機構 知識処理グループの研究成果を移転・活用しています。

参考文献

- [1] Kawashima and Matsuyama. Interval-based hybrid dynamical system for modeling multimedia timing structures. In *First International Symposium on Universal Communication Proceedings*, pages 67–70, 2007.
- [2] Kitaoka. Liveliness of spoken dialog systems – considering response timing and prosodic synchrony. In *First International Symposium on Universal Communication Proceedings*, pages 63–66, 2007.
- [3] 河原 and 荒木. 音声対話システム. オーム社, 2006.
- [4] 大竹, 堀, 柏岡, and 中村. 京都観光案内対話コーパスにおける対話行為の分析. In 言語処理学会第 14 回年次大会発表論文集, pages 159–162, 2008.
- [5] 堀, 大竹, 柏岡, and 中村. 京都観光案内対話コーパスにおける対話行為に関する研究. In 日本音響学会講演論文集, pages 105–106, 2008.