

音声ガイドシステムにおける全文検索結果の利用法

酒井 桂一 山田 雅章 伊藤 史朗
小森 康弘 池田 裕治 藤田 稔

キャノン (株) 情報システム研究所
〒 211 神奈川県川崎市幸区鹿島田 890-12
Tel: (044)549-5111

あらまし

本稿では、全文検索の結果を音声ガイドシステムに利用する新しい手法について提案する。音声認識で利用できる語彙数は、実用規模のデータを対象としてガイダンスを行なうには現状では不十分である。しかし、対話の局面に応じて認識語彙を動的に変更すれば、見かけ上の認識語彙を増大させることができる。我々が作成した「たびの電子ガイドシステム:TARSAN」は、CD-ROMに収録された約8万件の観光情報を対象としている。対話処理部がその全文検索結果を利用して、名称、地名、立地条件など認識語彙を動的に変更することによって音声対話を実現した。また、本システムを用いてガイダンスを行なった結果に基づき、本手法の有効性を検証する。

和文キーワード

音声対話システム, ガイダンス, 対話管理, ユーザ次発話予測, 全文検索

A Usage of Full-Text Retrieval in a Speech Guidance System

Keiichi SAKAI Masayuki YAMADA Fumiaki ITOH
Yasuhiro KOMORI Yuji IKEDA Minoru FUJITA

Information Systems Research Center, **Canon** Inc.
890-12, Kashimada, Saiwai-ku, Kawasaki-shi, Kanagawa, 211, Japan
Phone: +81-44-549-5111

Abstract

A new approach, which applies a full-text retrieval method to speech dialogue system, is discussed. TARSAN is a travel guidance system, which can retrieve data from practical scale text information (about 80,000 items in two CD-ROMs), using spoken dialogue. To deal with a large number of vocabulary in the speech recognition part, we developed a new technique which dynamically changes recognition-vocabulary according to the results retrieved from the CD-ROMs. In the current system, city and/or prefecture names, location conditions and subject-names are changed actively according to the situation in speech conversation. Experimental results using TARSAN are also discussed.

英文 key words spoken dialogue system, guidance, discourse management, prediction of next user utterance, full-text retrieval

1 はじめに

これまでの自然言語対話システムの研究は、音声入力を伴わない対話システムを対象としたものが主であった [1, 2, 3]。これらのシステムでは、キーボードからの入力がユーザにとって煩雑であるという欠点から、自然言語の持つ柔軟性が相殺されてきた。一方、近年の音声認識技術の進歩にともない、音声による対話システムの研究が盛んになってきている [4, 5, 6]。しかし、現在の音声認識システムで利用できる語彙数は、実用規模のデータを対象としてガイダンスを行うには不十分である。これに対処するために、我々は、対話の局面に応じて認識語彙を動的に変更することによって、見かけ上の認識語彙を増大させる音声対話システムの研究開発を行なった。

我々は、ガイダンスのドメインを「旅行」とし、その情報を収録した CD-ROM の全文検索結果から音声対話によりガイドする「たびの電子ガイドシステム:TARSAN¹」を作成した [7]。TARSAN では、旅行のガイダンスを、以下の 2 つのステップの繰り返しにより、目標に到るタスクとして考えた。

1. 検索条件の付加/変更により候補を変更する。
2. 現在の候補の情報を得る。

このように考えると、それぞれのステップ (以下、「対話の状態」と呼ぶ) において用いられる語彙は限られるので、それを予測して認識語彙とすれば、現在の音声認識でも対処できる。またこのとき、対話の状態をいかにユーザに提示するかも対話システムとして重要な要素となろう。

今回、ドメインを旅行としたので、候補を絞り込む条件として特に地名に注目した。また、地名以外の条件についても適宜ユーザに情報を提供することによって、上記のタスクを実現した。さらに、条件が不足していたり、曖昧であったりする場合には、システムが入力を要求したり、確認したりする枠組みを作成した。

以下、2章では、TARSAN の概要について述べ、3章で上記のタスクを実現した対話管理方法について述べる。4章では、全文検索結果から認識語彙を生成し、ユーザに提示する方法を述べ、5章では、実際に TARSAN を用いてガイダンスを行った結果に基づき、本手法の有効性を検証する。

2 システム概要

TARSAN のシステム構成を図 1 に示す。

¹Travel Assistant system composed of information-Retrieval, Speech conversation Access and Natural-language processing

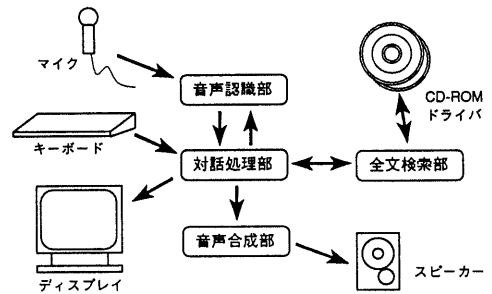


図 1: システム構成

TARSAN は、大きく 4 つのモジュール: 1) 音声認識部、2) 対話処理部、3) 全文検索部、4) 音声合成部、から構成される。

全文検索の対象として用いている 2 枚の CD-ROM には、日本全国にある観光地 [13] と宿泊施設 [14] に関する情報、約 8 万件が収録されている。全文検索部では 24 種類のジャンル (温泉、ゴルフ場、ホテル、神社・仏閣、美術館・遊園地など) および各ジャンルに依存する項目 (所在地、行き方、説明、電話番号、立地条件、効能、御利益など) を検索条件として、検索条件を満たす文書データの名称やその各項目の内容を検索することができる [10]。

入力は、マイクによる音声入力とキーボードによるタイプ入力を切り替えることができる。マイクから入力された音声は、音声認識部で認識され、文字列にされて対話処理部に送られる [8, 11]。一方、タイプ入力では、音声認識部で対処していない検索条件 (任意の単語や数値など) も直接入力することができる。

対話処理部は、入力された文を解析して「検索要求」を作成し、全文検索部に送る。また、全文検索部から返される検索結果から次発話に用いられる語彙を予測し、音声認識部に送るとともに、ディスプレイを通してユーザに提示する。さらに、音声合成部に送る読み上げ文を生成する。

出力は、検索要求、認識語彙、対話履歴などを出力するディスプレイ出力と、読み上げ文を音声合成部で規則合成した音声スピーカーから出力する音声出力とがある。

本システムは全て NeXT 上に実現されている。対話処理部、全文検索部、音声合成部に 1 台の NeXTstation Turbo Color を用い、音声認識部に NeXTcube 1 台、NeXTstation Turbo Color 2 台の 3 台を用いた。

音声入力後、4 つのモジュールで処理を行ない、音声出力し始めるまでの時間は、検索条件として都道府県名を指定した場合、10 ~ 20 秒である。

3 対話管理

3.1 発話意図

ムード解析[12]の結果に基づき、以下のようにユーザーの発話意図を決定する。

要求: 以下の場合に発話意図を要求とする。

- 「教えてほしい」、「示せ」など、主動詞が“教示動詞”か“調査動詞”で伝達ムードが“命令”、“依頼”、“相手希望”の場合。
- 「知りたい」など、主動詞が“知覚動詞”か“調査動詞”で伝達ムードが“動作主希望”の場合。
- 「何ですか」、「どこ」など、疑問詞を含んだ伝達ムードが問いかけの文の場合。

提示: 伝達ムードが述べたての名詞述語文の場合。

挨拶: 「おはよう」、「ありがとう」など。

肯否: 「はい」、「いいえ」を含む文。

条件変更: 「(<項目名>が)<項目値>では。」の形式の問いかけ。

対象変更: 「その<項目名>は。」あるいは「<名称>の<項目名>は。」の形式の問いかけ。

3.2 検索要求の変更/追加

3.2.1 履歴情報の保持

ジャンル(温泉やゴルフ場など)毎に“履歴テーブル”を作成する。履歴テーブルには、ジャンル名と文番号および“検索要求リスト”を保持する。また、検索要求リストの各要素は、「条件(C)」あるいは「対象(R)」を示すフラグ、項目名、項目値、文番号からなる。検索要求リストは、作成時に項目名を名称とする「対象」を要素を持つ。例えば、図2のusr1から(温泉,1,(C,所在地,東京都,1)(R,名称,(null),1))という履歴テーブルが作成される。

3.2.2 条件変更/追加の判定

発話意図が要求あるいは条件変更の文が入力された場合、入力文および履歴テーブルから、以下の手順にしたがって条件の変更/追加を行なう。

1. 発話意図が要求の文で与えられた条件を“初期条件”とする(以降の処理では履歴テーブルの文番号と比較して判定する)。
2. 発話意図が条件変更の文中の<項目名>が直前の条件と一致する場合、あるいは「<項目名>が」が省略された場合には、直前の条件の項目値を<項目値>に変更した検索要求を全文検索部に送る。

3. それ以外の場合、

- (a) <項目名>が検索要求リスト中の条件にある場合には項目値を<項目値>に変更する。
- (b) <項目名>が検索要求リスト中の条件にない場合にはその条件を追加する。

さらに、変更した条件と初期条件以外に条件がなければ、検索要求を全文検索部に送り、それ以外に条件があれば、その検索条件を残すか否かを確認する。

3.2.3 対象変更/追加の判定

発話意図が対象変更の文が入力された場合、入力文および履歴テーブルから、以下の手順にしたがって対象の変更/追加を行なう。

1. 「その<項目名>は。」の場合、

- (a) 検索要求リスト中の対象が名称のみの場合には<項目名>を対象として追加する。
- (b) 検索要求リスト中に名称以外に対象がある場合にはその対象を<項目名>に変更する。

2. 「<名称>の<項目名>は。」の場合、検索要求リストに(C,名称,<名称>,n)と(R,<項目名>,(null),n)を追加する(nは文番号)。

3.2.4 対話例

上記、検索要求の変更/追加の対話例を図2に示す。

```
usr1: 東京都にある温泉を教えてください。
sys1: 港区に2件、神津島村に2件、八丈町に2件、新島本村に2件など、合計で16件あります。
usr2: その効能は。
sys2: リューマチ・神経痛が1件、貧血症が1件、胃腸病、婦人病が1件、データなしが13件です。
usr3: 効能が胃腸病では。
sys3: 間々下温泉です。
usr4: 神経痛では。
sys4: 神津島温泉です。
usr5: 所在地が神奈川県では。
sys5: 他の条件、効能が神経痛、を残しますか。
usr6: はい。
sys6: 阿部倉温泉、大楠温泉、奥湯河原温泉です。
```

図2: 検索要求の変更/追加の対話例

まず、usr1により温泉の履歴テーブルを作成し、検索要求リストに(C,所在地,東京都,1)という初期条件を格納する。usr2は、対象追加であり、検索要求リストに(R,効能,(null),2)を追加する。usr3では、初期条件以外に条件がないので、検索要求リストに(C,効能,胃腸病,3)を追加する。usr4は、直前の条件の変更なので、確認せずに検索要求リスト中の効能を(C,効能,神

経痛,4)に変更して検索を行なうが,usr5は,初期条件の変更であり,それ以外に検索条件があるので,所在地を(C,所在地,神奈川県,5)とした後,検索要求リスト中の(C,効能,神経痛,4)を残すかどうか確認を行なう。

3.3 対話対管理

要求-提示,挨拶-挨拶,確認-肯否など,典型的な発話意図の対を対話の単位(対話対)とし,図3に示すように,対話対が入れ子になるような対話に対して,「対話対スタック」を用いて処理を行なう[12]。

TARSANでは,ジャンル毎に対話対スタックを作成する。前節で述べた「条件変更/追加時の確認」のほかに,以下の場合に対話対スタックを用いて「問い返し」を行なう。

1. 何も条件が指定されていない場合(sys1)
2. 必須条件としている所在地が指定されていない場合(sys3)
3. 名称に対し,検索対象がない場合(sys5)

また,ジャンルや項目そのものに対するユーザの質問(メタ質問,usr2)は,さらに対話対スタックに積んで,処理を行なう。

```

usr1: ゴルフ場について教えて。
sys1: 何か条件を指定して下さい。
usr2: どんな条件がありますか。
sys2: 名称,所在地,立地条件,種別,行き方,電話番号,パー,ホール数,面積があります。
usr3: 立地条件が河川数です。

usr3: 所在地を指定して下さい。
usr4: 千葉県。
sys4: 大利根チサンカントリークラブ,野田市パブリックゴルフ場,東我孫子カントリークラブです。
usr5: 大利根チサンカントリークラブについて知りたい。
sys5: 何について知りたいのですか。
usr6: 電話番号です。
sys6: 大利根チサンカントリークラブの電話番号は,(0471)38-2121です。
  
```

図3: 対話対が入れ子になる対話例

4 全文検索結果の利用

本システムが扱うようなガイダンスタイプの対話では,ユーザに提示する情報量によって,効率良く進められるかどうか左右される。TARSANでは,ドメインを「旅行」としたので,検索条件としてまず,地名に注目した。以下では,全文検索結果から生成する認識語彙(ディスプレイに表示)と読み上げ文の内容について述べる。

4.1 認識語彙の動的生成

対話処理部は,動的に変更する認識語彙(表記とその読み²の対,以降,動的語彙と呼ぶ)を検索結果から生成して,音声認識部に送る。なお,音声認識用に静的な文法に記述されている語彙を静的語彙と呼ぶ。具体的には,検索条件として候補を絞り込んでいくための「地名」を始めとする項目値,さらに,絞り込まれて候補の件数が少なくなると,名所や宿泊施設の名称を動的語彙とする。

音声認識部は,常時入力可能な「静的な文法」と,上記,動的語彙を含む「動的な文法」の2つの文法を持ち,両者の認識結果のスコアが高い方を対話処理部に送る。なお,動的な文法の作成時間は数秒であるので,音声出力の間に終了し,対話の進行には支障をきたさない[8,11]。

認識語彙の変化を示す対話例を図4に,静的語彙および対話例によって生成される動的語彙を表1に示す。なお,生成した動的語彙はディスプレイに出力することによって,ユーザに提示する。

```

usr1: おはよう。
sys1: おはようございます。TARSANです。
usr2: 東北地方にある温泉を知りたい。
sys2: 福島県に133件,秋田県に94件,山形県に73件,青森県に72件など,合計で470件あります。
usr3: 宮城県では。
sys3: 仙台市に8件,川崎町に5件,鳴子町に5件,白石市に3件など,合計で36件あります。
usr4: 仙台市では。
sys4: 秋保温泉,神々根温泉,作並温泉,広瀬川温泉など,合計で8件あります。
usr5: 作並温泉の行き方は。
sys5: 作並温泉の行き方は作並駅バス5分です。
usr6: 仙台市にあるホテルの立地条件は。
sys6: 市街が36件,駅前が13件,湊谷が12件,山間が9件,湯街が7件,盛場が1件,高台が1件,河畔が1件,丘陵が1件,田園が1件,山上が1件あります。
usr7: 立地条件が湊谷では。
sys7: ホテル奥仙台,ゆづくしの宿一の坊,ホテル佐勘,ホテルクレセントなど,合計で12件あります。
  
```

図4: 認識語彙の変化を示す対話例

4.1.1 地名

東北地方,関東地方などの「地方名」を静的語彙とする(表1の静的語彙)。また,全国の都道府県名をシステム起動時の動的語彙とする(表1の起動時)。これにより,システム起動時に検索条件にできる地名は地方名と都道府県名である。

その後,検索条件とした地名の項目値によって,以下のように動的語彙を生成する。

²旅蔵[13]には名称の読み情報が付与されているので,それを用いる。その他の読みは音声合成用辞書の読みを用いる。

表 1: 図 4 の対話例の音声認識語彙 (地名, 立地条件, 名称の表記のみ)

静的語彙	東北地方, 関東地方, 東海地方, 甲信越地方, 北陸地方, 近畿地方, 中国地方, 四国地方, 九州地方 (地名) 駅前, 丘陵, 市街, 林間 (立地条件)
動的	起動時 北海道, 青森県, 秋田県, 岩手県, 宮城県, 山形県, 福島県, 茨城県, 栃木県, 群馬県, ..., 沖縄県 (地名)
	sys2 青森県, 秋田県, 岩手県, 宮城県, 山形県, 福島県 (地名)
	sys3 青森県, 秋田県, 岩手県, 宮城県, 山形県, 福島県 仙台市, 川崎町, 鳴子町, 白石市, 花山村, 利府町, 小野田町, 宮崎町, 鹿島台町, 名取市, 栗駒町, 角田市, 岩沼市, 村田町, 蔵王町 (地名)
	sys4 青森県, 秋田県, 岩手県, 宮城県, 山形県, 福島県, ..., 村田町, 蔵王町 (地名) 秋保温泉, 泉温泉藤沢の湯, 神ヶ根温泉, 鴻ノ巣温泉, 作並温泉, 名取川温泉, 広瀬川温泉, 二口温泉 (名称)
語彙	sys6 青森県, 秋田県, 岩手県, 宮城県, 山形県, 福島県, ..., 村田町, 蔵王町 (地名) 市街, 駅前, 渓谷, 山間, 湯街, 盛場, 高台, 河畔, 丘陵, 田園, 山上 (立地条件) 秋保温泉, 泉温泉藤沢の湯, 神ヶ根温泉, 鴻ノ巣温泉, 作並温泉, 名取川温泉, 広瀬川温泉, 二口温泉 (名称)
	sys7 青森県, 秋田県, 岩手県, 宮城県, 山形県, 福島県, ..., 村田町, 蔵王町 (地名) 市街, 駅前, 渓谷, 山間, 湯街, 盛場, 高台, 河畔, 丘陵, 田園, 山上 (立地条件) ホテル奥仙台, 河鹿荘, 鷹泉閣若松旅館, ゆづくしの宿一の坊, 神の湯作並ホテル, 湯の原ホテル, ホテル佐勘, 茶寮宗園, ホテル秋保, ホテルクレセント, 秋保グランドホテル, ホテル華乃湯 (名称)

1. 地名を検索条件として検索した場合 (図 4 の usr2) には, 検索結果が存在する都道府県名 (とその読み) を動的語彙とする (表 1 の sys2). このように, あらかじめ地名を指定すれば, 例えば, 「福島県」と「徳島県」, 「宮城県」と「宮崎県」のように互いに誤認識しやすい似通った県名の他方を認識語彙から排除することができる.
2. 都道府県名を検索条件として検索した場合 (図 4 の usr3) には, 指定した都道府県と同じ地方にある都道府県名と, 検索結果が存在する市町村名 (とその読み) を動的語彙とする (表 1 の sys3). なお, 音声認識させたい都道府県名が動的語彙からはずれてしまった場合には, 地名を検索条件にして条件変更すると, その地方の都道府県名を動的語彙にすることができる.
3. 上記以外の検索条件で検索した場合 (図 4 の usr4, 5, 6, 7) には, 地名の認識語彙はそのまま残す. これは, 直前に認識語彙となっている地名を条件変更の候補として保持するためである.

なお現在, 動的語彙とする地名数は, 認識語彙数の制約により, グループ化して検索結果が多いものから, 最大 50 としている.

4.1.2 その他の項目値

宿泊施設やゴルフ場の立地条件, 温泉の効能, 神社の御利益なども候補の絞り込みを行なうのに有効な項目である. これらの項目値のうち, 典型的なものは静的語彙としている (表 1 の静的語彙). しかし, 宿泊施設の立地条件は全国で 68 種類あり, そのすべてを静的語彙とすることはできない. そこで, これらの項目を検索対象

として検索を行なった場合 (図 4 の usr6) に, その結果を動的語彙とする (表 1 の sys6). なお, 動的語彙とする項目値の数は, グループ化して検索結果が多いものから, 最大 16 としている.

4.1.3 名称

検索条件の追加/変更を繰り返して, 検索結果の件数が少なくなる (図 4 の usr4, 7) と, 名所や宿泊施設の名称を動的語彙とする (表 1 の sys4, 7). 具体的には,

1. 検索結果が 16 件以下の場合には, 全名称 (とその読み) を動的語彙とする.
2. 検索結果が 17 件以上の場合にも, 検索結果の市町村名をグループ化して単一の場合には, $1 + \text{int}(\text{件数}/16) * n$ 番目 ($n = 0, 1, 2, \dots, 15$) の名称 (とその読み) を動的語彙とする.

4.2 読み上げ文の生成

音声認識部に送るすべての動的語彙を内容とする読み上げ文を生成し, 合成音声で出力すると, 検索結果が多い場合に冗長になる. そこで, 検索結果を簡略化して読み上げ文を生成する.

今回特に, 地名に注目して, その範囲を狭めていくことによって絞り込みを行なう対話モデルを想定した. ここで, 検索対象が明示されない (すなわち, 名称を検索対象とする) 場合に, 検索結果が 5 件以上で, なおかつ複数の地名に存在する場合には, 名称ではなく, 検索結果が存在する地名および件数を読み上げ文の内容とすることにした.

具体的には以下のように読み上げ文を生成する.

4.2.1 検索対象が明示されない場合

1. 検索結果が4件以下の場合、全名称を出力する(図5の sys1)。
2. 検索結果が5件以上の場合、地名の検索条件が地名である場合には都道府県名で、都道府県名である場合には市町村名でグループ化した結果をさらに各件数で降べき順にソートを行ない(グループ化する単位を行政名と呼ぶ)、
 - (a) 行政名が単一の場合、 $1 + \text{int}(\text{件数}/4) * n$ 番目 ($n = 0, 1, 2, 3$) の名称と件数を出力する(図5の sys2)。
 - (b) 行政名が2以上5未満の場合、全行政名と各件数を出力する(図5の sys3)。
 - (c) 行政名が5以上の場合、件数上位4つの行政名と各件数および総件数を出力する(図5の sys4)。

```
usr1: 関東にあるお城を知りたい。
sys1: 館山城(市立博物館分館)、江戸城、小田原城です。
usr2: 箱根町にある温泉を教えてください。
sys2: 芦ノ湖温泉、木質温泉、底倉温泉、宮城野温泉など、合計で16件あります。
usr3: 茨城県にある美術館を教えてください。
sys3: 笠間市に3件、茨城県大和村に1件、水戸市に1件、合計で5件あります。
usr4: 千葉県にあるゴルフ場を教えてください。
sys4: 市原市に23件、野田市に7件、君津市に5件、千葉市に5件など、合計で86件あります。
```

図5: 検索対象が明示されない場合の対話例

但し、前節で述べたように、検索結果が16件以下であれば、全名称を動的語彙とするとともに、ウィンドウに表示する。したがって、読み上げ文では行政名を示していても、名称は音声入力可能である。

4.2.2 検索対象が明示された場合

1. 検索結果が4件以下の場合、全名称と項目値を出力する(図6の sys1)。
2. 検索結果が5件以上の場合、項目値でグループ化した結果をさらに各件数で降べき順にソートを行ない³、
 - (a) 項目値が16種類以下の場合、全項目値と各件数を出力する(図6の sys2)。
 - (b) 項目値が17種類以上の場合、総件数と種類数を出力する(図6の sys3)。

³但し、項目値が存在しないもの(項目値を「データなし」とする)は最後に置く。

```
usr1: 関東地方にあるお城の入場料を知りたい。
sys1: 館山城(市立博物館分館)の入場料は150円、江戸城の入場料はデータなし、小田原城の入場料は250円です。
usr2: 神奈川県にある温泉の効能を教えてください。
sys2: リューマチ・神経痛が2件、病後回復・ストレス解消が1件、リューマチ・神経痛、外傷・骨折・やけど、痔、皮膚病、婦人病が1件、外傷・骨折・やけど、痔、皮膚病、糖尿・肥満が1件、皮膚病が1件、データなしが25件あります。
usr3: 神奈川県にあるホテルの立地条件を教えてください。
sys3: 条件を満たす406件のうち、立地条件が24種類あります。
```

図6: 検索対象が明示された場合の対話例

5 評価

5.1 実験内容

今回設定したタスクを評価するために、以下の実験を行なった。

被験者 研究所員計20名

実験1: 音声認識に対する頑健さ 被験者が、図7に示すような対話例(3種類:合計24対話)を音声入力し、音声認識部で誤認識した結果が対話処理部に送られた場合の対話の状態の変化を調べ、対話システムとしての頑健さを評価した。但し、ここでいう誤認識とは、実際に発声した言葉と音声認識結果である文字列とが異なっている、結果として変わらないものの⁴は含まない。

```
usr1: 東京都にある神社を教えてください。
sys1: 新宿区に24件、中野区に20件、葛飾区に16件、八王子市に14件など、合計で291件あります。
usr2: 御利益が厄除けでは。
sys2: 港区に2件、目黒区に2件、文京区に1件、豊島区に1件など、合計で9件あります。
usr3: 所在地が目黒区では。
sys3: 他の条件、御利益が厄除け、を残しますか。
usr4: はい。
sys4: 祐天寺、滝泉寺(目黒不動)です。
usr5: 祐天寺の行き方を教えてください。
sys5: 祐天寺の行き方は祐天寺駅徒歩15分です。
usr6: 所在地が文京区では。
sys6: 源覚寺(こんにゃくえんま)です。
usr7: その行き方を知りたい。
sys7: 源覚寺(こんにゃくえんま)の行き方は都営三田線春日駅徒歩1分です。
usr8: ありがとう。
sys8: まだどうぞ。
```

図7: 神社・仏閣の対話例

実験2: 目標検索 被験者が次に示す目標を音声入力により検索し、与えられた目標に効率良く辿りつけるかどうかを評価した。

⁴「知りたい」、「教えて」、「示せ」間の誤認識など。

1. 彫刻の森美術館の行き方 (所在地は関東地方)
2. 鎌倉時代に作られた埼玉県にある寺院と同じ市町村にあるゴルフ場で、種別がパブリックのものの電話番号

5.2 実験1の結果および評価

3種類の対話例(合計480発話)に対し、51の誤認識があった。各誤認識に対して対話の状態を調べ、以下の3種類に分類して考察した。

1. 対話の状態に変化がなかったもの: 31件
2. 1対話で元の対話の状態に戻るもの: 7件
3. 元の対話の状態に戻すのに2対話以上かかるもの: 13件

1. の原因としては、

- 1-a) 誤認識した文が受理されて検索が行なわれたが、動的語彙に変化がなかったため、すぐに正しい文を入力することができるもの
- 1-b) 誤認識する前の状態では誤認識した文が受理不能な文だったため、システムが入力文を拒絶した⁵もの

に分けられた。

1-a) については、例えば、図7のusr3で、「目黒区」を「文京区」と誤認識した場合があげられる。この場合には、確認(usr4)に答えると、文京区の神社・仏閣を検索するが、動的語彙に変化がないので、その後、「所在地が目黒区では」という条件変更が入力可能である。

1-b) については、主に、文末表現である「です」、「は」、「は」間などの誤認識であった。

2. の原因としては、

- 2-a) 誤認識した文が受理されて検索が行なわれた結果、動的語彙に変化があったため、正しい文を入力するために別の入力文が必要なもの

があげられる。

2-a) については、例えば、図7のusr3で、「目黒区」を「群馬県」と誤認識した場合があげられる。この場合には、確認(usr4)に答えて、群馬県の神社・仏閣を検索すると、群馬県の市町村名が動的語彙となり、「目黒区」が認識語彙から外れてしまう。しかし、「所在地が東京都では。」という条件変更を入力すると、再び「目黒区」を認識語彙とすることができる。

3. の原因としては、

⁵この場合、システムは「別の言い方をして下さい。」と応答し、対話の状態は変化しない。

3-a) 誤認識した文が受理された結果、ジャンルが変わったため、元の対話の状態がなくなったもの

3-b) 誤認識を繰り返した結果、容易に元の対話の状態に戻れなくなったもの

に分けられた。

3-a) については、例えば、図7のusr5で、「祐天寺」を「遊園地」と誤認識した場合があげられる。この場合には、ジャンルが「神社・仏閣」から「遊園地」に変わり、システムは遊園地に対する条件を求める状態になるため、全く最初からやり直す必要はない⁶が、1対話では元の状態に戻れない。

3-b) については、最初からやり直す方が手っ取り早い場合が多かった。

今回作成したシステムでは、音声認識部での誤認識に対して特別な枠組を作成しなかったが、より円滑な音声対話を行なうには、例えば、誤認識入力をキャンセルし、元の状態に戻すなど、何らかの対策が必要であろう。

5.3 実験2の結果および評価

音声目標1: 彫刻の森美術館の行き方

本目標を求めるには、名称である「彫刻の森美術館」を認識語彙とするのに1対話必要なので、図8に示すように最短でも2対話が必要である。

usr1:	神奈川県にある美術館を示せ。
sys1:	横浜시에 4 件, 箱根市に 4 件, 鎌倉市に 3 件, 川崎市に 1 件など, 合計で 13 件あります。
usr2:	彫刻の森美術館の行き方は。
sys2:	彫刻の森美術館の行き方は彫刻ノ森駅徒歩 2 分です。

図8: 音声目標1に到る最短対話例

この結果、被験者20人が目標に到達するのに、最短2対話、最長10対話、平均5.4対話を要した。

本実験では、目標である彫刻の森美術館の所在地(神奈川県箱根町)を知っている被験者が多かったこと、また、被験者が詳しい所在地を知らない場合でも、所在地を関東地方に限定したので、上記の結果で目標に到達することができた。しかし、以下のことが考察される。

- 名称だけわかっていて地名がわからない場合には、その名称を認識語彙にするのは容易ではない。
- 地名がわかっている場合にも、単一市町村に同一ジャンルが17件以上存在する場合には、地名以外の検索条件をつけなければ、名称を認識語彙にすることができるとは限らない。

⁶この場合、東京都の市町村名は動的語彙からはずれないので、2対話で元の状態に戻すことができる。

音声目標 2: 鎌倉時代に作られた埼玉県にある寺院と同じ市町村にあるゴルフ場で、種別がパブリックのもの電話番号

本目標を求めるには、ゴルフ場がある市町村を求めるのに3対話必要なので、図9に示すように最短でも5対話が必要である。

usr1:	埼玉県にある寺院を知りたい。
sys1:	岩槻市に6件、川越市に6件、所沢市に6件、飯能市に6件など、合計で117件あります。
usr2:	作成年代が鎌倉時代では。
sys2:	清浄寺です。
usr3:	その住所は。
sys3:	清浄寺の住所は埼玉県吉川町木売です。
usr4:	埼玉県吉川町にあるゴルフ場の電話番号を知りたい。
sys4:	北越谷パブリックコースの電話番号は(0489)91-2011, 越谷ゴルフクラブの電話番号は(0489)82-2311です。
usr5:	種別がパブリックでは。
sys5:	北越谷パブリックコースの電話番号は(0489)91-2011です。

図 9: 音声目標 2 に到る最短対話例

この結果、被験者 20 人が目標に到達するのに、最短 6 対話、最長 34 対話、平均 12.5 対話を要した。

本実験で余分にかかった対話で特徴的なものを以下に示す。

音声誤認識: 検索条件とする市町村名「吉川町」が埼玉県以外にもあるため、「埼玉県吉川町」を動的語彙としていたので、被験者が単に「吉川町」と入力した場合には認識できなかった。また、「埼玉県吉田町」という似通った市町村名が認識語彙となっていたため、これと誤認識される場合があった。

ジャンルの変更: 最短の戦略は、まず寺院を特定してゴルフ場に移るのであるが、被験者が先にゴルフ場について検索した場合に、ジャンルを寺院に移す際にゴルフ場の検索条件が失われるため、再度入力する必要があった。

6 まとめ

本稿では、旅行に関する音声ガイドシステム TARSAN における全文検索結果の利用法について述べた。全文検索の結果から地名や立地条件などを動的な認識語彙として検索条件の追加/変更を繰り返し、ある程度まで候補が絞られると名称を認識語彙とすることによって、全国の観光地や宿泊施設のガイダンスを実現した。また、作成したシステムで実験を行ない、音声誤認識に対する頑健さと予め与えた目標に到る対話の過程の評価を行なった。

謝辞

本研究を行なう機会を与えて下さいました知能工学研究部 田村秀行部長に感謝いたします。

参考文献

- [1] 近藤他: 対話システム IDS における対話モデルと協調的応答生成, 情報処理学会, 知能工学と人工知能研究会 57-3, 1988.
- [2] 加藤他: 質問応答における意図の把握と話題の管理, 情報処理学会, 自然言語処理研究会 58-6, 1986.
- [3] 住田他: 質問応答システムにおける会話の自主性に関する考察, 信学会論文誌 D, Vol.J70-D No.11, pp.2287-2293, 1987.
- [4] 天野他: 音声対話システムの試作, 日本音響学会講演論文集, 1-1-20, 1992.
- [5] 竹林他: 音声対話システムにおける Spontaneous Interaction の実現, 日本音響学会講演論文集, 1-1-22, 1992.
- [6] 森屋他: 対話予測を利用した音声による観光案内対話システム, 信学技報, SP92-121, 1993.
- [7] 藤田他: 音声対話と全文検索を利用した電子ガイドシステム (1) -システム概要-, 情報処理学会 46 回全国大会 4E-4, 1993.
- [8] 山田他: 音声対話と全文検索を利用した電子ガイドシステム (2) -音声認識部-, 情報処理学会 46 回全国大会 4E-5, 1993.
- [9] 酒井他: 音声対話と全文検索を利用した電子ガイドシステム (3) -対話処理部-, 情報処理学会 46 回全国大会 4E-6, 1993.
- [10] 伊藤他: 音声対話と全文検索を利用した電子ガイドシステム (4) -全文検索部-, 情報処理学会 46 回全国大会 4E-7, 1993.
- [11] 小森他: 音声対話を用いた CD-ROM 情報検索システムの構築, 日本音響学会平成 5 年春季研究発表会, 3-4-14, pp. 77-78, 1993.
- [12] 酒井他: 文書整形システムを対象とする自然言語インタフェース, 「自然言語処理の新しい応用」シンポジウム論文集, pp.22-31, 1992.
- [13] 「旅蔵」電子ブック, 企画製作(株)廣濟堂, 発行(社)日本観光協会, 1990.
- [14] JTB の「宿泊情報」(電子ブック), 編集人 中嶋隆一, 発行人 岩田光正, 発行所 JTB 日本交通公社出版事業局, データ編集凸版印刷株式会社, 1992.