

特集 車載情報サービス用音声対話システム*

Speech Dialog System for In-vehicle Information Services

立石 雅彦

Masahiko TATEISHI

赤堀 一郎

Ichiro AKAHORI

畑岡 信夫

Nobuo HATAOKA

渡邊 順一郎

Junichiro WATANABE

Eric Nyberg

Teruko Mitamura

Yasunari Obuchi

DENSO CORPORATION and Hitachi, Ltd. are developing a next generation speech dialog system for in-vehicle information services, through joint research with Carnegie Mellon University. The speech dialog system employs a novel dialog manager based on an extension to the VoiceXML that allows a user to switch the dialog tasks flexibly. This paper describes the basic configuration of the proposed system and the task switch handling by the dialog manager, followed by the analysis of a speech corpus used for designing the dialog scenarios.

Key words : Speech dialog system, VoiceXML, Dialog manager, Speech recognition, In-vehicle information services

1. はじめに

近年、ドライバーが車の中に居ながらにしてインターネット等のネットワークにアクセスし、交通情報やニュース等のオンライン情報を入手する車載情報サービスが立ち上がりつつある。

カーナビゲーションに代表される車載機器の情報化・高度化は、車載情報サービス導入によりさらに進展する。そしてドライバーに対し、オンライン情報を含む豊富かつ多様な情報を提供することが可能となる。これらの情報をドライバーが十分に活用するためには、自分の望む情報を自由に引き出せるHMI（ヒューマンマシンインタフェース）が必須となる。スイッチでは検索に何回もの手操作を必要とする情報を直接呼び出すことができ、かつ検索情報や検索条件を自由に切り替えられる柔軟なHMIを実現する手段として、音声対話が注目され、各方面で研究されている¹⁾²⁾³⁾⁴⁾。

音声対話システム実現にはドライバーの発話を聞き取る音声認識技術だけでなく、ドライバーの発話に対して適切に回答する対話管理技術、複雑な問い合わせを理解する自然言語処理技術が必要となる。こうした課題に対処するため、(株)デンソーは(株)日立製作所と米国カーネギーメロン大学との共同研究により、情報センターと車載機（車載端末）から構成される車載情報サービスを想定した新しい音声対話システムの開発を進めている。本システムは情報センターと車載機の連携によるスムーズかつ高性能な音声対話の実現をコンセプトとし、(1)話題遷移に対応できる柔軟な対話管理の実現、(2)情報センターへの問い合わ

せ不要な対話を車載機だけで処理することによる対話処理の高速化、という2点を主な狙いとする。

一般に、音声対話システムでは、ユーザとシステム間の応答規則を記述した対話シナリオを用いて対話を制御する。対話システムの用途に応じ、様々な対話シナリオ記述言語が提案されている⁵⁾⁶⁾。我々は車載情報サービス用の音声対話シナリオを記述する基本的な枠組みとして、音声認識、音声合成を含んだコンテンツをWEB上で記述する標準言語であるVoiceXMLに注目した。VoiceXMLを用いることで、今後普及が見込まれるVoiceXML対応の様々なシステムで、提案する音声対話システムを動作させることが可能となる。

VoiceXMLで自由な話題遷移を実現するには、話題間の遷移規則を明示的に記述する必要がある。しかしすべての話題遷移を想定した対話シナリオを人間が手作業で記述するのは工数がかかるだけでなく、新たな話題を追加するといった将来の機能拡張への対応も困難である。そこで我々は、話題毎に個別のシナリオを生成し、次いで各話題シナリオ間に遷移規則を組み込んで話題遷移可能なVoiceXMLシナリオを自動生成する方法を考案した。

また、音声認識エンジンおよびVoiceXMLインタプリタを車載機に持たせ、例えばオーディオ操作など車内だけで処理できる対話を車載機だけで処理するようにした。これは情報センターへのアクセス回数を低減し、対話処理を高速化する。

なお、ドライバーの主な発話に対応し、かつ要領のよい対話を実現する対話シナリオを設計するには、実

* 2003年2月27日 原稿受理

際の音声対話事例を収集・分析し、得られた知見を対話シナリオに反映しなければならない。我々はその基礎データとなる大規模な音声対話事例（音声対話コーパス）の収集にも取り組んだ。

本システム開発では、デンスーが音声認識エンジンおよび音声対話コーパスを、日立がVoiceXMLインタプリタおよび情報センター構築技術を、カーネギーメロン大学が対話管理技術および自然言語処理技術を担当している。

本論文では、はじめに本システムの構成と特長を説明する。次に話題遷移に対応できる対話シナリオをVoiceXMLで実現する方法を説明する。最後に音声対話コーパスについて述べる。

2. システムの構成と特長

Fig. 1に車載機および情報センターの構成を示す。車載機は音声認識部、音声合成部、VoiceXMLインタプリタから構成される。音声認識部はドライバーの発話を単語列に変換する。VoiceXMLインタプリタはこの単語列を解釈実行し、その結果を音声合成部により音声に変換してドライバーに伝達する。

VoiceXMLインタプリタは、単語列がオーディオ操作など車内だけで処理できる内容であれば車載機だけで対応する。さもなければ単語列を情報検索コマンドと解釈し、URLのフォーマットに変換して情報センタ

ーに送信する。そして情報センターから返された情報を使用してドライバーと対話を継続する。このように車載機は必要な場合にのみ情報センターにアクセスするため、対話処理の高速化が可能となる。

情報センターは対話管理部、情報データベース、音声認識用語彙・文法データベースから構成される。車載機から単語列を受信すると、対話管理部は情報データベースおよびインターネットを検索する。そして次節で述べる手順により検索結果を組み込んだVoiceXML対話シナリオを自動生成する。またVoiceXML対話シナリオが想定するドライバー発話を音声認識するための用語彙、文法を音声認識用語彙・文法データベースから検索する。そしてこれらVoiceXML対話シナリオ、音声認識用語彙・文法を合わせて車載機に送信する。なお、音声認識用語彙・文法データは、音声対話コーパスを自然言語処理部によりオフライン処理して事前に生成する。

例えばドライバーが飲食店を目的地として設定すると、対話管理部は関連する話題、例えば目的地までの経路情報、および目的地周辺の駐車場情報を元に、経路案内、駐車場案内の二つの話題を扱うVoiceXML対話シナリオを生成する。また、ドライバーの発話を音声認識するための用語彙・文法を音声認識用語彙・文法データベースから検索する。そしてこれらの情報を合わせて車載機に送信する。以後、車載機のVoiceXMLイ

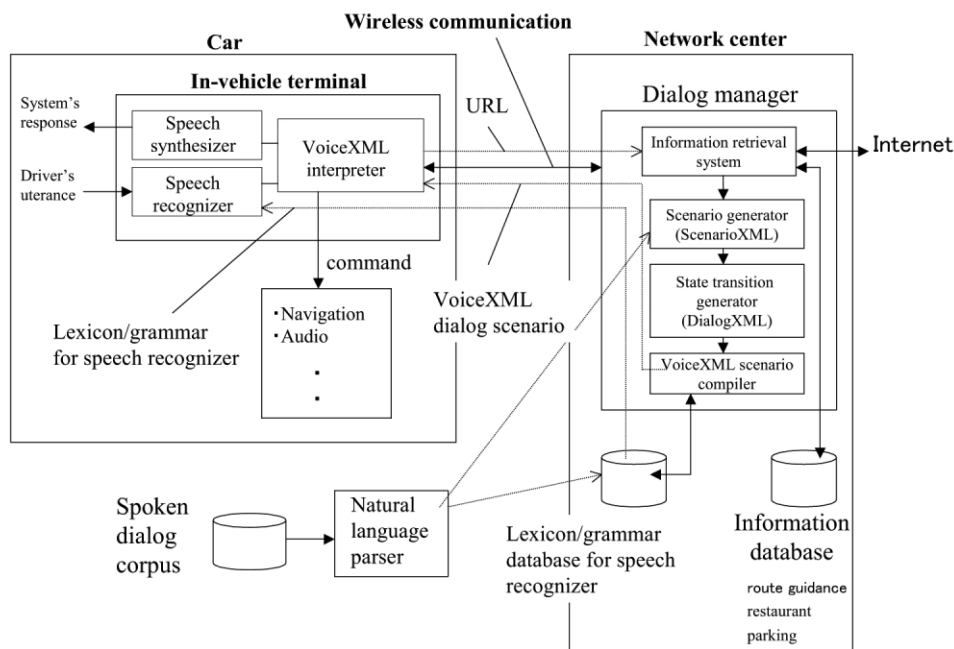


Fig. 1 System architecture

ンタブリタは対話管理部から受信した情報を用いてドライバーとの対話を処理する。

本システムの最大の特徴である柔軟な話題遷移は対話管理部が実現する。ドライバーは例えば目的地への経路案内の対話の最中に、駐車場情報を問い合わせ、再び元の対話に戻って経路案内対話を続けることができる。次節でその実現方法を述べる。

3. 対話管理方法

車載情報サービスでは、ドライバーの要求があった時点で、動的に対話シナリオを生成する必要がある。経路案内を例にとると、自車位置と目的地の組み合わせは無数にある。そのすべてを想定した経路案内シナリオを事前に用意することは不可能である。したがって、経路案内シナリオは、ドライバーが目的地を設定した時点で、自車位置情報も考慮して動的に生成しなければならない。

従来の対話システムは、階層的なメニューや状態遷移グラフが用いられている。さらに2, 3の話題に対処できるシステムも提案されている⁷⁾⁸⁾しかし、対話中に話題遷移が可能な対話シナリオを動的に生成する方法は提案されていない。

本システムは動的シナリオ生成および話題遷移に対応するため、VoiceXMLを拡張したScenarioXML, DialogXMLと呼ぶ言語を新たに開発し、VoiceXML対話シナリオを自動生成する方法を採用した。はじめに後者のDialogXMLについて説明する。通常、音声対話は状態遷移モデルを用いて記述することができる。例えば経路案内では、経路案内開始時、交差点右左折時、目的地到着時に案内音声を出力する。各案内音声を出力する時点をそれぞれ一つの状態として表現し、案内音声を出力する順に遷移規則をつけることで経路案内対話を記述できる。DialogXMLは状態と遷移規則を記述する構成要素を持つ対話シナリオ記述言語であり、音声対話を状態遷移モデルの形で記述することを可能とする。以後、DialogXMLで記述した音声対話シナリオをDialogXML対話シナリオと呼ぶ。

一方、前者のScenarioXMLはある一つの話題について、DialogXML対話シナリオを生成する一種の対話シナリオ生成言語である。これは情報検索結果を入力とし、その情報検索結果を組み込んだDialogXML対話シナリオを出力する。音声対話システムが扱う話題それぞれについて、ScenarioXMLで記述した対話シナリオ

生成プログラムを用意する。

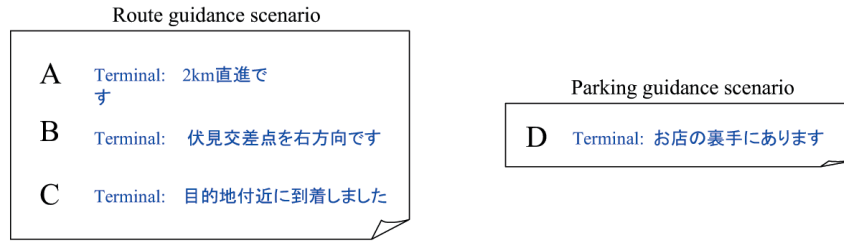
対話管理部は以下の3段階の手順で、VoiceXML対話シナリオを生成する。

- (1) ScenarioXMLを用い、情報検索結果をもと個々の話題のDialogXML対話シナリオを生成する
- (2) 話題シナリオの各状態から他の話題シナリオの先頭に飛ぶ遷移規則を組み込む
- (3) (2)のシナリオをVoiceXML対話シナリオに変換する

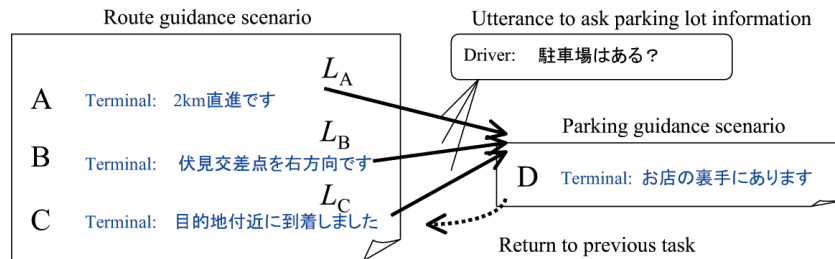
なお、(1) (3)の機能は、Fig. 1の対話管理部内の話題シナリオ生成部、状態遷移組み込み部、VoiceXML変換部がそれぞれ担当する。ここで、ScenarioXMLの設計には、音声対話コーパスから得られた対話規則を反映している。

ドライバーが飲食店を目的地に設定した場合を例に、生成される対話シナリオを説明する。簡単のため、出現する話題は飲食店までの経路案内および飲食店の駐車場案内の二つとする。対話管理部は、以下に示す3段階の処理で、VoiceXML対話シナリオを生成する。

- (1) 話題シナリオの生成
経路情報、駐車場情報をもとに、経路案内および駐車場案内のDialogXML対話シナリオを動的に生成する (Fig. 2 (1))
- (2) 話題遷移機能の組み込み
経路案内シナリオの各状態 (A, B, Cと表記) から、駐車場案内シナリオ先頭の状態 (Dと表記) への遷移規則 L_A, L_B, L_C を挿入する。各遷移規則はドライバーが駐車場を問い合わせる発話、例えば「駐車場はある?」をした場合に有効になり、対話が遷移先の話題に移るよう設定する。(Fig. 2 (2))
- (3) (2)の対話シナリオをVoiceXML対話シナリオに変換する



(1) Online generation of scenarios for each task



(2) Insertion of global transitions across a set of dialog scenarios

Fig. 2 Generation of dialog scenarios

この対話シナリオによれば、ドライバーが伏見交差点を右折後に駐車場を問い合わせると、VoiceXMLインタプリタは遷移規則 L_B をたどって、駐車場案内シナリオに制御を移す。駐車場案内が終了すると、経路案内シナリオに制御を戻して対話を継続する。なお、遷移先の対話から元の対話への復帰はVoiceXMLのsubdialog機能を用いる。この対話シナリオによる対話例をFig. 3に示す。

VoiceXML対話シナリオに自動変換する。したがって、話題遷移可能なVoiceXMLシナリオを動的に生成できる。また、各話題のシナリオを独立に開発できるため、シナリオの設計・保守、話題追加等の拡張も容易である。

4. 音声対話コーパスの開発

音声対話システムの対話管理部は、車載情報サービスで出現する主な話題に対応し、かつ要領の良い対話を実現しなければならない。また、音声認識部はユーザが問い合わせによく使用する単語や文型を極力受理しなければならない³⁾。

車載情報サービスでは、交通案内、観光案内、施設案内等が話題として出現する。対話管理部が主な話題に対応できるようにするためには、音声対話中に出現する話題の種類と出現頻度を分析し、出現確率の高い話題について対話シナリオを生成できるようにしなければならない。また、音声認識部がユーザの発話を認識できるようにするためには、音声対話中に出現するユーザ発話を分析し、出現確率の高い単語や文型を受理するよう音声認識で使用する語彙や文法を決めなければならない。これら出現確率を求めるには、実際の音声対話事例を大量に収集し、統計的に分析する必要がある⁹⁾。

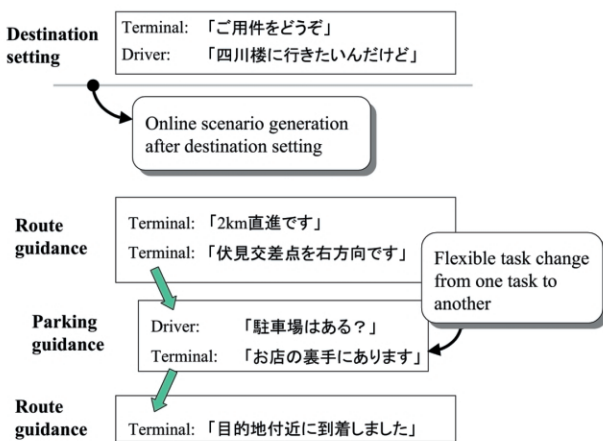


Fig. 3 An example of speech dialog

上述の対話シナリオ生成法は対話開始時に関連する話題のシナリオを生成し、それら話題シナリオ間に遷移規則を挿入して話題遷移機能を組み込んだ後、

音声対話システム設計や評価のために、大量に収集された音声対話事例を音声対話コーパスと呼ぶ。我々は交通情報、観光情報、施設情報等を案内する車載情報サービスを想定し、東京都内から伊豆または箱根に一泊二日のドライブ旅行に出発する、あるいは旅行中という設定で、情報センターを模擬するアナウンサーとドライバーを模擬する被験者との間で自由に対話をしてもらい、音声対話コーパスを収集した。Table 1 にその収集条件を示す。被験者数は20代から60代の男女それぞれ137名、113名の計250名とした。各被験者の対話時間は約15分、合計で約60時間とした。また音声対話コーパスは、東京都内のスタジオで収集した。ここで情報センターを模擬するアナウンサーには、観光案内という専門性の高い話題について対話を要領良くこなす技量が求められる。そこで我々は、旅行業務取扱主任者の資格を有し、かつ旅行代理店および放送局勤務経験がある熟練アナウンサーを起用した。

Table 1 Condition of speech corpus collection

Test subject	Number	137 male speakers 113 female speakers
	Age range	20's 30's 40's 50's 60's
Task setting	Habitation area	Tokyo metropolitan area
	Place of departure	Tokyo
	Destination	Izu or Hakone
Duration of dialog		15 min per test subject
Place of recording		A studio in Tokyo

次に音声対話コーパスの分析結果を述べる。音声対話コーパスに出現した話題は被験者あたり平均5.0件、全体で1,253件であった。各話題の出現回数の比率をFig. 4に示す。また、合計発話回数はオペレータ、被験者それぞれ34,830回、33,957回であった。さらに被験者発話を形態素解析ツールChaSen¹⁰⁾で形態素解析し、語彙数を調査した。その結果をTable 2に示す。

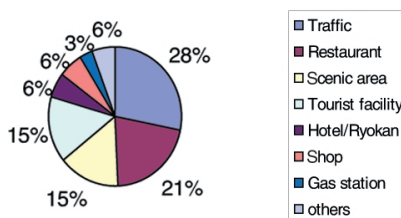


Fig. 4 Ratio of task frequency in the speech dialog corpus

Table 2 Size of vocabulary in the speech corpus

Category	Size of vocabulary
Place name	394
Tourist attraction name	307
Shop name	140
Traffic facility name	128
Railway facility name	30
Others	3,604
Total	4,603

Fig. 4に示すように交通に関する話題が最も多く、次いで飲食店、観光情報の順となった。対話に出現する話題全体の例えば90%以上をカバーするには交通（経路案内、渋滞情報等）から店（土産物店、コンビニエンスストア情報等）までの上位6種類の話題に対応しなければならないことが判明した。一方、Table 2に示すように、地名から鉄道施設名までの固有名詞が語彙数999個、その他の単語の語彙数が3,604個、合計4,603個であることが判明した。これより、音声認識に必要な語彙数は5,000個程度であることが分かった。現在、これらの情報を元に、対話シナリオの設計、音声認識用辞書・文法の設計を行っている。

5. まとめと今後の課題

車載情報サービス用の新しい音声対話システムを提案した。本システムは、(1)話題遷移に対応できる柔軟な対話管理の実現、(2)情報センターへの問い合わせ不要な対話を車載機だけで処理することによる対話処理の高速化、という2点を主な狙いとする。前者の課題に対応するため、我々はVoiceXMLを拡張した新規の対話シナリオ記述言語を導入し、個々の話題シナリオを生成し、各話題シナリオ間に話題遷移機能を組み込んだ後、VoiceXML対話シナリオに変換する方法を考案した。一方後者の課題に対応するため、音声認識エンジン、VoiceXMLインタプリタを車載機に持たせ、例えばオーディオ操作など車内だけで処理できる対話を車載機で処理し、対話処理を高速化する構成とした。

今後は情報センター、車載機双方について実用化の課題解決に取り組む。前者の課題として、音声対話シナリオおよび音声認識用辞書・文法の拡張、音声認識

結果に誤りがあっても対話を正しく続けられる頑強な対話管理、音声認識結果を自然言語処理システム¹¹⁾で解析し、複雑な問い合わせに対応する機能等がある。後者の課題としては、音声認識率向上、情報センターから送信された情報を車載機に適宜蓄積し再利用することによる対話処理のさらなる高速化等がある。

謝辞

音声対話コーパス開発にあたり、貴重なご助言、ご指導をいただいた京都工芸繊維大学工芸学部の荒木雅弘助教授に感謝いたします。

< 参考文献 >

- 1) E. Nyberg, T. Mitamura, P. PLaceway, M. Duggam and N. Hataoka, DialogXML: Extending VoiceHML for Dynamic Dialog Management, Proc. of HLT-2002 (2002)
- 2) 畑岡他, VoiceXMLをベースにした音声対話方式の開発, 日本音響学会2002年秋季研究発表会 (2002)
- 3) 伊藤他, ドライブプランニングシステムの自然言語インタフェース, 人工知能学会誌, Vol.17, No.3, pp. 285-290.
- 4) 河口他, 実走行車内における音声データベースの構築, 情報処理学会研究会, 音声言語情報処理, 99-SLP-30-12 (2000)
- 5) VoiceXML forum, <http://www.voicexml.org/>
- 6) SALTforum, <http://www.saltforum.org/>
- 7) P. Price, Evaluation of Spoken Language Systems: The ATIS domain, Proc. of 3rd DARPA Workshop on Speech and Natural Language (1990)
- 8) S. Seneff, et al., Organization, Communication, and Control in the Galaxy-II Conversational System, Proc. of Eurospeech99 (1999)
- 9) C. Cieri: " Resources for Robust Analyses of Natural Language ", Conference ROMAND 2000, Lausanne, 2000
- 10) ChaSen Home Page, <http://chasen.aist-nara.ac.jp/index.html.ja>
- 11) E. Nyberg, and T. Mitamura, The KANTOO Machine Translation Environment, Proc of AMTA (2000)



< 著 者 >



立石 雅彦
(たていし まさひこ)
基礎研究所
音声認識の研究に従事



赤堀 一郎
(あかほり いちろう)
基礎研究所
音声認識の研究に従事



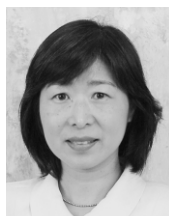
畑岡 信夫
(はたおか のぶお)
(株)日立製作所 中央研究所
工学博士
音声認識・音声合成を含む信号処理,
及びヒューマンインタフェース
用メディア処理の研究に従事



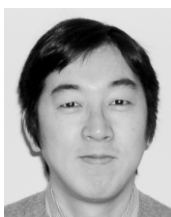
渡邊 順一郎
(わたなべ じゅんいちろう)
(株)日立製作所 中央研究所
ワードスポット技術, 騒音に頑健な
電話音声認識技術, 多言語音声通訳
の研究に従事



Eric NYBERG
(ナイバーグ エリック)
Carnegie Mellon University, Language
Technologies Institute
Ph. D.
Machine Translation, Controlled
Language, Dialog Systems, Open-
Domain Question Answering, Software
Architectureの研究に従事



Teruko MITAMURA
(みたむら てるこ)
Carnegie Mellon University,
Language Technologies Institute
Ph. D.
Machine Translation, Controlled
Language, Dialog Systems, Open-
Domain Question Answering,
Computer-Aided Language Learning,
Automatic Rewritingの研究に従事



Yasunari OBUCHI
(おおぶち やすなり)
(株)日立製作所 中央研究所
Carnegie Mellon University, Language
Technologies Institute
Robust speech recognition, Spoken
dialog system, Speech to speech
translationの研究に従事