

車載情報サービス用音声対話コーパス

立石 雅彦*¹, 赤堀 一郎*¹
スコット ジュディ*², 大淵 康成*³, 三田村 照子*², エリック ナイバーグ*², 畑岡 信夫*⁴

*1 デンソー基礎研究所
〒470-0111 愛知県日進市米野木町南山 500-1
TEL:05617-5-1064
mtatei@rlab.denso.co.jp

*2 カーネギーメロン大学 Language Technologies Institute
*3 日立製作所基礎研究所
*4 日立製作所中央研究所

あらまし 本稿は音声対話システム Conversational Agent for Multimedia Mobile Information Access の設計のために収集した音声対話コーパスについて述べる。CAMMIA は VoiceXML を拡張した新規対話シナリオ記述言語を使用し、柔軟な話題遷移を可能とするのが特長である。車載情報サービスを想定したコーパスを 20 代から 60 代の男性 137 名、女性 113 名から収集した。コーパス中の対話数、発話数はそれぞれ 450、67,123 となり、話者の語彙数は約 5,000 語、話者 1 名あたりの話題数は約 5.0 であった。

A SPOKEN DIALOG CORPUS FOR CAR TELEMATICS SERVICES

Masahiko Tateishi*¹, Ichiro Akahori*¹
Scott Judy*², Yasunari Obuchi*³, Teruko Mitamura*², Eric Nyberg*², Nobuo Hataoka*⁴

1* Research Laboratories, DENSO CORPORATION
500-1, Minamiyama, Nisshin city, Aichi, 470-0111, JAPAN
TEL: 05617-5-1064
mtatei@rlab.denso.co.jp

*2 Language Technologies Institute, Carnegie Mellon University
*3 Advanced Research Laboratory, Hitachi, Ltd.
*4 Central Research Laboratory, Hitachi, Ltd.

Abstract This paper describes a spoken dialog corpus used in the design of CAMMIA (Conversational Agent for Multimedia Mobile Information Access), which employs novel extensions to VoiceXML to allow users to switch dialog tasks flexibly. The corpus for car telematics services is collected from 137 male and 113 female speakers in the five age brackets of 20's, 30's, 40's, 50's, and 60's. The spoken corpus consists of 450 conversations comprising of 67,123 utterances. The speakers' vocabulary size is approximately 5,000 words. The average number of dialog tasks per speaker is 5.0.

¹ (株)デンソー 基礎研究所
DENSO CORPORATION, Research Laboratories

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、各方面で盛んに研究されている。^{6),7)}

筆者らが提案する Conversational Agent for Multimedia Mobile Information Access^{1),4)}は、車載情報サービス等への応用を目指したクライアント-サーバー型の音声対話システムである。CAMMIA は VoiceXML⁵⁾を拡張した新規対話シナリオ記述言語を使用し、柔軟な話題遷移を可能とする。この特長により、ドライバーは目的地設定対話の途中で目的地の天気を問い合わせ、また目的地設定対話に戻るといった対話を行うことが可能である。

ドライバーの主な発話に対応し、かつ要領の良い音声対話を実現する対話シナリオを設計するには、実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。そこで我々は、交通、飲食店、観光案内等を行う車載情報サービスを想定した、人対人の音声対話コーパスを収集/分析した。

本稿は、第2節で CAMMIA のシステム構成について説明する。第3節で音声対話コーパスの作成について、第4節で音声対話コーパスの分析結果について述べる。最後にまとめと今後の課題について述べる。

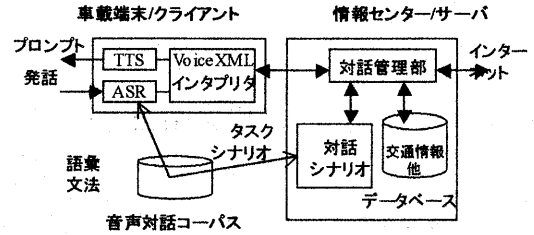


図1: CAMMIA のシステム構成

2. CAMMIA のシステム構成

図1に CAMMIA のシステム構成を示す。CAMMIA は無線ネットワークで接続された車載端末とサーバから構成される。

クライアントは ASR、VoiceXML インタプリタ、TTS を備えた車載端末であり、一方サーバは対話管理部、情報データベース、対話シナリオを備えたコンピュータである。クライアントはドライバー発話を ASR によって単語列に変換し、検索コマンドとしてサーバに送信する。サーバは受信した検索コマンドを用いてデータベースを検索し、その結果を組み込んだ VoiceXML 対話シナリオをオンラインで生成してクライアントに返す。クライアントはサーバから返された VoiceXML 対話シナリオを VoiceXML インタプリタにロードし、以後の音声対話処理を行う。

音声対話コーパスは ASR の語彙・文法、および対話シナリオの設計/評価に使用する。以下、その作成および分析結果について述べる。

3. 音声対話コーパスの作成

本節では、音声対話コーパスの話者の選定と収集期間、収集方法、及び書き起こし方法について説明する。

3.1 話者の選定と収集期間

話者の発話に出現する語彙や文法、および音声対話に出現する話題は話者の年齢や性別に大きく左右されると考えられる。したがって幅広い年代の男女ユーザに対応できる音声対話システムを構築するには、収集する音声対話コーパスの話者の性別と年代を均等化

することが望ましい。そこで我々は、男性話者 137 名、女性話者 113 名の計 250 名を話者とした。話者の年代も 20 代、30 代、40 代、50 代、60 代でほぼ均等とした。話者は首都圏在住者で、うち 235 名は運転免許所有者であった。また、50 名はナビゲーションの使用経験者であった。

次に我々は、上記 250 名の話者を 50 名ずつの 5 つのグループに分けた。そして各グループの音声対話を 1 ヶ月から 1 週間程度の時間間隔を空けて順に収集した。1 グループの収集が完了した時点で、音声対話収集方法を改善し、次の収集に反映した。全グループの音声対話収集には約 4 ヶ月を要した。

3.2 音声対話収集方法

本節では音声対話収集に用いた音声対話収録環境、及びタスクと教示について述べる。前節で述べたように、我々は 50 名の音声対話を収集する度に音声対話収集方法を改善した。我々は最初の 50 名の音声対話収集後に、最も大きな改善を行った。以下、その改善前、改善後の音声対話収集方法を説明する。

3.2.1 音声対話収集方法（改善前）

・音声対話収録環境

我々は本音声対話を東京都内のスタジオで収録した。該スタジオは 2 部屋から構成され、片方の部屋にドライブ情報を問い合わせる話者が、他方の部屋にドライブ情報を提供するオペレータが入った。話者の部屋は車載端末を備えた車に相当し、オペレータの部屋は情報センターに相当する。また、話者はドライバーに相当し、オペレータは音声対話システムに相当する。話者とオペレータはヘッドセットとマイクを使用して音声対話を行う。オペレータは以下の情報をもつドライバー情報サービス要員として、話者からの質問に回答した。

- オンライン交通情報
- 飲食店情報
- 観光情報
- ホテル、旅館情報

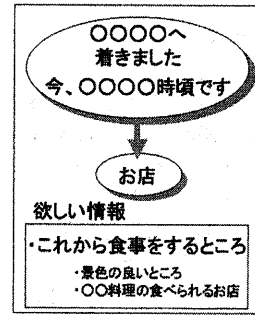


図 2：タスク教示書の例

オペレータには、正確な日本語を話すだけでなく、旅行案内という専門性の高いタスクを巧みにこなす能力が求められる。そこで我々は旅行代理店および民放ラジオ局の勤務経験があり、かつ旅行主任技術者の資格を有する熟練アナウンサーを起用した。

・タスクと教示

我々は話者に対し、車載端末を搭載した車に乗車し、東京から伊豆または箱根に旅行するという想定で、オペレータと対話し、ドライブ情報を得るよう指示した。

本音声対話コーパス収集に先立ち、302 名のドライバーに対して行った事前調査では、運転中に欲しいと思った案内の上位 3 つは交通案内、飲食店案内、観光案内であるという結果が得られていた。したがって、我々は最初に以下の 3 つのタスクを話者に与えて音声対話コーパスを収集した。

- タスク 1. 目的地までの交通情報を得る
- タスク 2. 目的地付近で昼食を食べる飲食店を探す
- タスク 3. 昼食後に観光する場所を探す

タスク教示は書面を用いて行い、状況と指示をそれぞれ文章で記述した。図 2 にタスク 2 の教示書の例を示す。

3.2.2 音声対話収集方法（改善後）

音声対話収集方法の改善内容を説明する。はじめにタスクと教示の改善について説明し、その後、音声対話収録環境の改善について述べる。

・タスクと教示

前節のようにあらかじめ決められたタスクを被験者に単に与える方法は幾つか問題点のあることが判明した。最も大きな問題は、話者の動機付けが十分にできないことである。被験者は旅行中であると十分思い込まず、与えられたタスクに関する質問ができない状況が発生した。その場合、オペレータは対話を中断し、どのような質問が可能なのか話者に教示し、場合によっては最初から対話をやり直す必要が生じた。また、教示を図2のように文章で記述すると、多くの話者は対話開始時に教示書を読んでしまい、自発話の収集に悪影響を与えることが判明した。以下にその例を示す。

話者： 山中湖へ着きました

オペレータ： はい

話者： えー今12時半ごろです

オペレータ： はい

そこで我々は残りの話者 200 名に対し、以下のように教示方法を変更した。200 名を5名のメンバーからなる 40 のグループに分けた。そして各グループに伊豆、箱根いずれかを目的地に選ばせ、その目的地に一泊二日のドライブ旅行をするという前提で、グループ内で議論をさせ、自由に計画を立てさせた。議論の後、各話者に各自の興味や関心事を反映した2つのタスクセット A、B を作成させた。タスクセット A は旅行1日目に自宅から旅行に出発する際に知りたい情報を列挙したリスト、タスクセット B は旅行2日目にホテルや旅館をチェックアウトする時点で知りたい情報を列挙したリストである。さらに、各話者の音声対話収録を2回に分け、最初の収録でタスクセット A に関する収録を行い、1時間ほど時間を置いた後にタスクセット B に関する収録を行った。

・音声対話収録環境

音声対話収録時には、旅行の日を事前に条件設定しておかなければならないことが分かった。日には設定は交通案内、ガソリンスタンド案内等に影響する。たとえば、交通状況は平日か週末かで混雑する道路や混雑の程度が大きく変化するので、オペレータの案内もそれに応じて変えなければならない。

しかし日にちを条件設定していないと、音声対話収録中に「今日は何曜日でしたっけ・・・」といった条件を確認する対話が入ってしまう。そこで、各被験者に対し、音声対話収録前に、旅行に出発する日を平日にするか週末にするかを選択させ、それに従い日にちを決定した。

これら日にち等の条件設定についても、書面に条件を文章で記述すると話者がその文章を読み上げてしまう可能性がある。そこで我々は、話者の机の上に卓上カレンダー、置時計を置いて日にちや時刻を示す等、条件設定を物で示す工夫をした。

また、次に収録を行う話者（以下、次話者）の緊張をほぐし、収録環境やタスクに慣れさせるため、ドライバー役の話者の隣に同乗者役として座らせた。次話者はヘッドセットを着用し、オペレータと話者の音声対話を聴くことができるが、音声対話に加わることはできない。

次話者を同乗者役として収録に立ち合わせたところ、次話者をタスクに慣れさせる効果だけでなく、以下のように話者の質問を引き出す効果もあることが分かった。

- (i) タスクセット中の項目で、話者が質問を忘れていた項目があると、次話者がその項目を示して質問を促す。
- (ii) オペレータが「（お問い合わせの情報を）お調べいたしますのでしばらくお待ちください」と対話を中断したとき、話者と次話者が会話して質問項目の確認などをする。

例) 話者と次話者の会話（太字の箇所）

オペレータ： 箱根湯本からえー芦の湖畔までの時間ということでお調べいたしますので

話者： はい

- ・お店の名前まで聞いていいのかな
- ・うんいんじゃない
- ・実際にあるお店でも
- ・いんじゃないいんじゃない

オペレータ： はい大変お待たせいたしましたえーまず中目黒から

L:はいドライブ情報センターです
 R:すいません
 L:はい
 R:えー{R東京駅}から
 L:はい
 R:えー{P箱根}までの行き方を教えていただきたいんですが
 L:はいえー{R東京駅}からえー{P箱根}ですねー
 R:はい

図3:書き起こしテキストの例

以上の改善の効果についてオペレータに尋ねたところ、改善後の話者は積極的に質問をするようになったと意見を述べた。また、オペレータが対話を中断し話者に教示を与えた回数は改善前では50名の話者に対して68回だったのに対し、改善後の200名の話者に対してはわずか1回と大幅に削減された。

3.3 書き起こし方法

収録した音声対話は手作業でテキストに書き起こした。音声対話は発話毎に分ち書きし、オペレータ、話者の発話の先頭に文字‘L.’、‘R.’を付けて区別した。

次にテキスト中の固有名詞を{X名前}といった形式でタグ付けした。ここでXはA, P, R, S, Wのいずれかであり、それぞれ観光施設名、地名、鉄道施設名、飲食店・商店名、道路施設名を示す。図3にタグ付け後のテキストの例を示す。なお、タグ付け後のテキストはChaSen⁸⁾により、形態素解析を行い、単語に分割した。

4. 音声対話コーパスの分析

本音声対話コーパスはCAMMIAの設計において、a)ASR文法の設計、b)ASR文法の評価、c)実際の音声対話に現れる語彙の把握、d)頻出単語・文章の特定、e)音声対話シナリオの設計等に利用する。本稿では話者の発話中に含まれる語彙、及び音声対話コーパス中に現れた話題の種類について述べる。

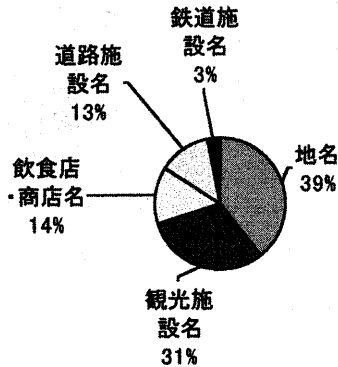


図4:固有名詞の語彙数の比率

4.1 話者発話中の語彙の分析

本音声対話コーパスは450の対話からなり発話数は67,123であった。

話者の語彙数は4,603個であり、内999個が固有名詞、残りの3,604個がそれ以外の単語であった。これより、話者の発話を理解するのに必要ASRの語彙数は5,000個程度であることが分かった。

図4に各固有名詞の語彙数の比率を示す。地名、観光施設名、飲食店・商店名、道路施設名の上位4種類の合計で割合は97%に達した。また、上位2種で割合は70%に達した。この割合は案内する観光地やオペレータの対話戦略により変わりうるが、この結果からASRの語彙には地名、観光施設名を優先的に組み込むのが妥当と考えられる。

4.2 音声対話コーパス中の話題の分析

本音声対話コーパス中に出現した話題数は話者250名で1253個、話者1名あたり平均5.0個となった。出現話題は上位から順に交通案内(28%)、飲食店案内(21%)、観光施設案内(15%)、景勝地案内(15%)となった。そして音声対話コーパス中の話題の例えば90%以上をカバーするには上位7種類の話題に対応する必要のあることが分かった。

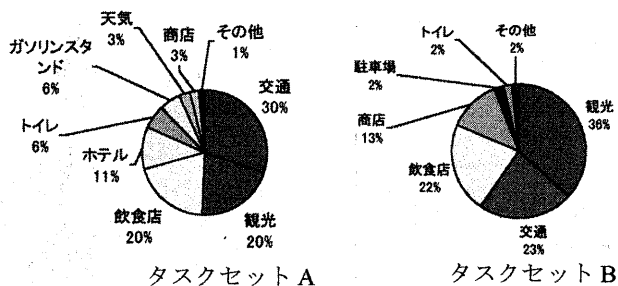


図5：話題出現頻度比率の比較

3. 2節で述べたように、我々は教示方法を改善し、200名の話者に対し、2つのタスクセットA,Bを作成させて音声対話を収集した。前者は旅行1日目の自宅出発時、後者は旅行2日目のホテル出発時を想定したタスクである。音声対話中に出現した話題数は前者が総計671個、後者が429個となり、話者1名あたりの平均はそれぞれ3.4個、2.1個となった。

図5にタスクセットA,Bそれぞれの各話題の出現頻度比率を示す。2日目ではホテル案内が消滅し、観光案内内の比率が上昇した。これはホテル案内がもはや不要であること、またホテル出発時に話者が最初に興味を持つ話題の多くが観光だったためである。さらに商店案内の比率が3%から13%に上昇した。これは何名かの話者が帰路に土産物、例えば干物専門店の案内を希望したためである。本音声対話コーパスはスタジオ内で収録した模擬音声対話であるが、上記話題出現頻度の比率変化は、話者のドライブ旅行の経験を反映した音声対話を収集できたことを示す。

5. まとめと今後の課題

車載情報サービス用の音声対話コーパスを250名の話者から収集した。話者の発話の語彙数は4,603個であった。これより、話者の発話認識に必要なASRの語彙数は5,000個程度であることが分かった。また、音声対話コーパス中のタスクの例えば90%以上をカバーするには上位7種類のタスクに対応する必要があることが分かった。今後、これらの結果を元にASRの語彙・文法、および対話シナリオの設計を行う。

謝辞

本音声対話コーパス開発にあたり貴重なご意見、ご指導をいただいた京都工芸繊維大学、荒木雅弘助教授に感謝します。また、本音声対話コーパスの分析に関し、貴重な助言をいただいた日立中央研究所、渡邊純一郎氏に感謝します。

参考文献

- 1) E.Nyberg et al., 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) Y.Obuchi et al., Robust Dialog Management Architecture Using VoiceXML for Car Telematics Systems, Proc. of Workshop on DSP in Vehicular and Mobile Systems(2003)
- 4) M.Tateishi et al., A Spoken Dialog Corpus for Car Telematics Services, Proc. of Workshop on DSP in Vehicular and Mobile Systems(2003)
- 5) VoiceXML forum, <http://www.voicexml.org/>
- 6) 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
- 7) 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)
- 8) ChaSen Home Page, NARA INSTITUTE of SCIENCE and TECHNOLOGY, "Morphological Analyzer ChaSen", <http://chasen.aist-nara.ac.jp/>