

## WWW 上のフォーム型情報検索サービスのための 音声インタフェースの検討

中野 崇広<sup>†</sup> 甲斐 充彦<sup>††</sup> 中川 聖一<sup>†</sup>

<sup>†</sup> 豊橋技術科学大学 情報工学系

<sup>††</sup> 豊橋技術科学大学 情報処理センター

〒 441-8580 豊橋市天伯町字雲雀ヶ丘 1-1

### あらまし

最近、WWW 上で利用できる様々な情報検索サービスが提供されているが、それらは一般に WWW ブラウザの GUI(Graphical User Interface) 環境での利用を想定している。本研究では、既存の WWW ブラウザの利用者向けの情報検索サービスに対して音声によるユーザインタフェースを提供するシステムを検討している。特に、ブラウザでメニュー選択の操作を必要とするフォームを含んだ Web ページを対象として、音声言語による対話的なインタフェースを提供する方法について検討を行なった。音声インタフェースに基づく情報検索システムは、代理サーバを介する仕組みに基づいて、汎用的な WWW ブラウザを利用する形で実装した。本システムは、既存の情報検索の Web ページへの適用を想定しており、主要な Web ページを対象として音声インタフェースとしての利用を実現した。

### Investigation on speech interface for form-based information retrieval services on the WWW

Takahiro Nakano<sup>†</sup> Atsuhiko Kai<sup>††</sup> Seiichi Nakagawa<sup>†</sup>

<sup>†</sup> Department of Information and Computer Sciences, <sup>††</sup> Computer Center,  
Toyohashi University of Technology

1-1, Hibarigaoka, Tenpaku-cho, Toyohashi-shi, Aichi, 441-8580, Japan

E-mail: {nakano,kai,nakagawa}@slp.tutics.tut.ac.jp

### Abstract

Recently, we can access a variety of information retrieval services on the WWW. However, they enforce the users to access via an WWW browser which is basically based on a graphical user interface. In this study, we investigate a general speech interface system which can be applied for many form-based information retrieval services on the WWW. The system is implemented based on a client-server configuration and a client system consists of a general WWW browser, speech I/O components and a Java applet. The proposed system aims at an architecture which is independent of the target Web pages for information retrieval services and it achieved the access to the various information retrieval services using our developed speech interface.

# 1 はじめに

WWW上の情報を閲覧したり情報検索サービスを利用するために、一般にマウスによる操作が可能なGUIを備えたWWWブラウザが利用されている。近年、WWWの利用が広がり、用途や利用環境(例えば携帯情報端末機器)も多様化していることから、WWWブラウザにおける音声インタフェースの拡張がいくつか検討されてきた[1, 2, 3, 4, 5, 6, 7, 8]。我々は以前、音声入力を用いたWWWブラウザ操作システムを試作し、音声による効率的な操作の実現方法について検討した[9]。このシステムでは、Webページ中のリンクに対応しているキーワードやその一部の発話により、リンク先へのジャンプ操作を可能にした。しかし、情報検索を目的とした利用においては、リンクの選択操作への対応だけでは不十分であった。

本研究では、特にWebページ中に選択メニューを持った情報検索サービスを対象として、特定の情報検索タスクに依存した対話モデルや言語モデルなどを使わず、汎用的な音声インタフェースを実現する方法を検討する。これまでも同様なシステム例[4, 10]や、HTMLの記述の段階でページ内で必要な音声認識文法を指定するタスク依存の方法によるシステム例[8]などの報告がある。提案するシステムは、対話のためのシステムのプロンプトを生成する機能を持たせると共に、Webブラウザへの表示を併用して汎用性を持たせていること、クライアント側のシステムを汎用のWebブラウザをメインとして軽量な実装を実現している点が主な特徴である。評価のために試作したシステムは、WWWブラウザで動作するJava appletを利用するとともに、クライアント・サーバ構成によって音声認識やフォームの解析等をリモート・サーバで実行するように実装を行なった。以下の節では、初めに音声インタフェースの設計について述べ、次に試作したシステムの構成と実行例および課題等について述べる。

## 2 フォーム入力のための音声インタフェース設計

### 2.1 フォーム型情報検索ページ

情報検索サービスを目的として一般に利用できるWebページには、図1の例のように、選択メニューやボタンなどのマウスで操作可能なGraphical User Interface(GUI)を用意したものが多く見られる。これらのGUIは、一般的なWWWブラウザ上で機能

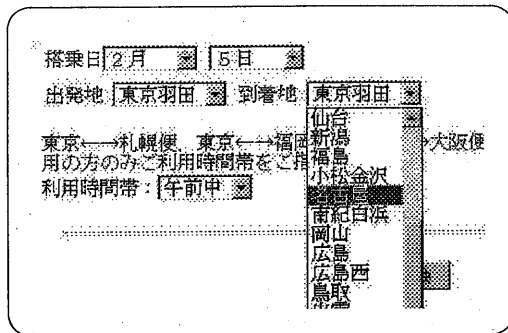


図1: 情報検索ページのフォーム表示部分の例

することを想定しており、情報検索のための効果的なユーザインタフェースを提供している。

HTML(Hyper Text Markup Language)の規格では、これらのGUIをWebページ上に配置して機能させるために、フォームと呼ばれる書式を用意している。フォームの記述では、テキスト入力フィールドや、セレクトボックス(選択メニュー)、チェックボックス、ラジオボタンなどの種類がある。ここでは、フォームを利用してユーザが情報検索を行なえるように記述されたWebページを対象として、汎用的な音声インタフェースを提供する方法を検討する。

### 2.2 フォーム入力への音声利用

フォームの記述では、一般に、ユーザが指定すべき複数の項目についての記述から構成されている。従って、フォームの記述を含むページ上でユーザが行なう情報検索の作業の流れは、情報検索において指定可能な属性に対する穴埋め(メニュー選択)のタスクとして見ることが出来る。そこで、この穴埋めのタスクを簡単な音声対話によって実現することを考える。WWW上での情報検索を対象とした音声対話型のシステムとしては、天気予報、航空案内などある程度対象を絞ったものが報告されている[3]。しかし、フォームを利用するものではなく、自然言語による問い合わせを可能とするために、特定のドメインに対する対話モデルを必要としている。

ここでは、前述のようなフォーム入力のGUIを備えた情報検索ページにおいて、通常はマウスやキーボードを用いて行う一連のフォーム上の操作を、音声のみによって代替するための汎用的なインタフェースを提供することを検討する。また、可能な範囲で、ユーザが入力する必要がある項目について、システムからの音声による問い合わせを可能にすることを考える。

音声だけでブラウザ上のフォームに対する入力を行うには、フォーム内のセレクトボックスなど各入力毎に、システムから入力を促す方法が考えられる。このとき、ユーザには以下のような情報を提示する必要がある。

- フォーム中のどの部分を入力しているのか
- 音声入力できるキーワードの一覧

このキーワードは、入力するフォームの種類によって変化する。チェックボックスであればチェックボックスに対応する項目名、セレクトボックスであればセレクトボックス内で選択可能な項目名、テキストボックスならばキーワードは自由入力となる。しかし、チェックボックスやラジオボタンの場合、HTMLにおいて項目名の指定に関する明確な定義がないため、項目名を自動的に抽出することは困難である。また、テキストボックスについては、音声による自由入力の実装は容易ではない。但し、音声インタフェース用に文法を指定できるなど、マークアップ記述言語の規格の拡張を考えれば [11]、将来的にはより実用的になると考える。セレクトボックスの場合は、現在の HTML の書式においても、項目名を自動的に抽出することは比較的容易である。そこで、特にフォームの中にセレクトボックスを含む Web ページをここで扱う対象とする。

### 2.3 選択型メニューのための音声インタフェース

任意のフォームに対して音声インタフェースを実現するには、フォームの記述内容に応じて対話に必要なユーザ発話の文法やシステムの発話を自動的に生成する必要がある。選択型メニューの GUI であるセレクトボックスについて音声インタフェースを作成するには、具体的に次のような処理が必要である。

- 選択リストに含まれるキーワード名の抽出
- 音声入力の認識用文法・辞書の生成
- 選択リストの入力項目の名前や種類の認識・抽出
- システムからの問合わせ・応答文の生成
- 複数の選択リストに対する入力の対話管理

一番目のキーワード名については、HTML のタグの情報をを用いてほぼ正確に対象ページの記述内容から抽出することができる。また、二番目の文法・辞書の生成については、以前に開発したシステム [9] で用いた方法とほぼ同じように、キーワードの形態

```

搭乘日<SELECT NAME="MONTH">
<OPTION VALUE="1"> 1 月
<OPTION VALUE="2"> 2 月
<OPTION VALUE="3"> 3 月
:
</SELECT>
:
出発地 <SELECT NAME="DPORT">
<OPTION VALUE="HND">東京羽田
<OPTION VALUE="NRT">東京成田
<OPTION VALUE="ITM">大阪伊丹
:
</SELECT>

```

(a) セレクトボックスの記述例

1 月	いちがつ	i ti ga tu
2 月	にがつ	ni ga tu
3 月	さんがつ	sa N ga tu
:		
東京	とうきょう	to o kyo o
成田	はねだ	ha ne da
東京羽田	とうきょうはねだ	to o kyo o ha ne da
:		

(b) 生成辞書ファイルの例 (ローマ字部分は音節表記)

図 2: フォーム記述と辞書の生成の例

素解析を行なった後、名詞の連続を主な単位とする一定の規則で、元のキーワードを形態素単位の系列からなる断片に分割したものを辞書に登録する。文法については、それらのキーワードの断片の単語発話の他、対話的な入力において必要となるユーザのコマンド発話等を定義する。図 2 に、HTML でのフォームの記述と、辞書に登録される単語リストの例を示す。

最後の三つの処理については、フォームの記述内容に応じて適当な方法を考える必要がある。次の節で、実際のフォームの記述内容例について述べる。

### 2.4 フォームの情報抽出と応答生成

HTML は文書構造に関しての記述の定義が十分ではなく、ブラウザ上では視覚的にセレクトボックスの種類や項目名の情報が得られても、HTML の記述から容易に抽出できない場合がある。しかし、一般的にはセレクトボックスのタグの前後のテキストからそれらの情報を抽出できる。抽出できれば、その情報をシステムの発話等に利用することが可能と考えられる。

実際に、フォームのセレクトボックスを持つ WWW 上の情報検索サービス用のページの HTML テキストの記述例を調査した。その結果、項目名の情報はセレクトボックスのタグを目印とすると、(1) タグの前にある文字列、(2) タグの後ろにある文字

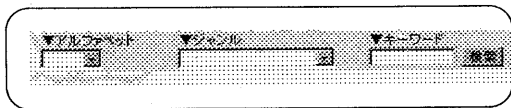


図 3: テーブルタグによるセレクトボックス配置の例

列、(3) タグの前および後にある文字列、として得られるケースが多かった。七つのホームページにおいて調査した結果、セレクトボックスの総数が 21 個で、そのうち 15 個 (71.4%) は上記 3 つのパターンで項目名が取得できた。ページ内の全ての項目名が取得できたものは 4/7 である。但し、セレクトボックス以外にボタンなどのフォームを同時に含んでいるページも二つあり、マウスやキー操作を行わないで音声のみで扱うには、特別な選択操作のための対話的な処理が必要となる。

その他のパターンとしては、(1) いくつかのセレクトタグが項目名を共有している場合 (例えば、「月」と「日」について、出発日という項目名が先頭についている)、(2) 項目の情報を示すために画像 (テキストを画像で表現していることもある) を利用している場合、(3) 入力に Javascript を用いている場合、(4) HTML のテーブルタグを含んでおり、セレクトと項目名の対応を得るにはテーブルの構造の解析が必要な場合 (図 3)、などがあつた。

抽出されたテキストによる項目の情報には、おもに次のような種類が存在する。

- 箇条書き (例: 「年齢:」、「出発地」)
- 疑問調の文 (例: 「出発地はどちらですか?」)
- 依頼調の文 (例: 「選択してください」)

これらの情報は、対話的にシステムからの問い合わせを行なうために利用できると思われる。図 2(a) の例では、「搭乗日」、「出発地」などが項目名とみなすことができるので、項目名を問合わせるシステムの発話文のテンプレートに基づいて、「搭乗日を入力してください」や「出発地を入力してください」というように文を生成し、音声合成による問い合わせを行なう。

### 3 音声インタフェースシステムの試作

#### 3.1 システム構成

図 4 に試作したシステムの全体の構成を示す。ユーザの計算機に相当するクライアント側では、WWW

ブラウザと音声入出力を受け持つ部分のみから構成される。また、試作システムでの主な処理は、リモートのハブ・ホスト上で行なう。このハブ・ホストでは、WWW サーバ、音声認識サーバとして機能する他、クライアント側で指定する Web ページの HTML ファイルの取得および解析、WWW サーバの CGI(Common Gateway Interface) 機能を用いて文法・辞書生成、システム応答 Web ページ、システム応答発話文章の生成の処理などを行なう。生成されたシステムの発話文は、クライアント側にテキストまたは音声として送られ、クライアント側で再生される。

音声認識サーバとしては、ネットワークベースで利用可能な音声認識システム SPOJUS[12, 13] を利用している。音声認識サーバは、音節単位の HMM をベースとした不特定話者のシステムで、音声の取り込み及び分析を行なう音声入力サーバから音声データを受信しながら時間同期的に処理を行なう。音声入力サーバが音声認識サーバに送る音声データは、マイクから入力された音声信号を 8msec 周期で 14 次の LPC 分析を行ない、10 次のメルケプストラム係数に変換したもので、4 バイトの float 型を 2 バイトに近似している。また、形態素解析サーバには茶筌 [14] を用いた。

#### 3.2 システムの実行例

図 5 に、試作したシステムでの WWW ブラウザの画面表示の例を示す。本システムの GUI は、全て一般的な WWW ブラウザ上で実装されている。図のように、WWW ブラウザは縦三段のフレームで構成され、最上段のフレームは、フォーム入力を提供している Web ページを選択するためのインタフェースとして Java アプレットが動作している。この部分が、音声入出力、音声認識部との通信、他のフレームの表示制御などを行う。中段のフレームは、最上段のフレームで指定されている、フォーム入力を含んだ Web ページが表示される。下段のフレームには、音声入力できるキーワードの一覧が表示される。

ユーザは、システムからの音声による問い合わせとブラウザ上に示されたキーワード一覧に基づいて、キーワードの一部または、入力項目を訂正・変更するためのコマンドを音声で入力できる。用意されている音声コマンドは、現在のところ、入力対象のセレクトタの変更のためのもので、「戻る」、「つぎ」の 2 つである。

図 6 に、JAL のホームページ [?] での音声インタフェースによる対話例を示す。WWW ブラウザ上

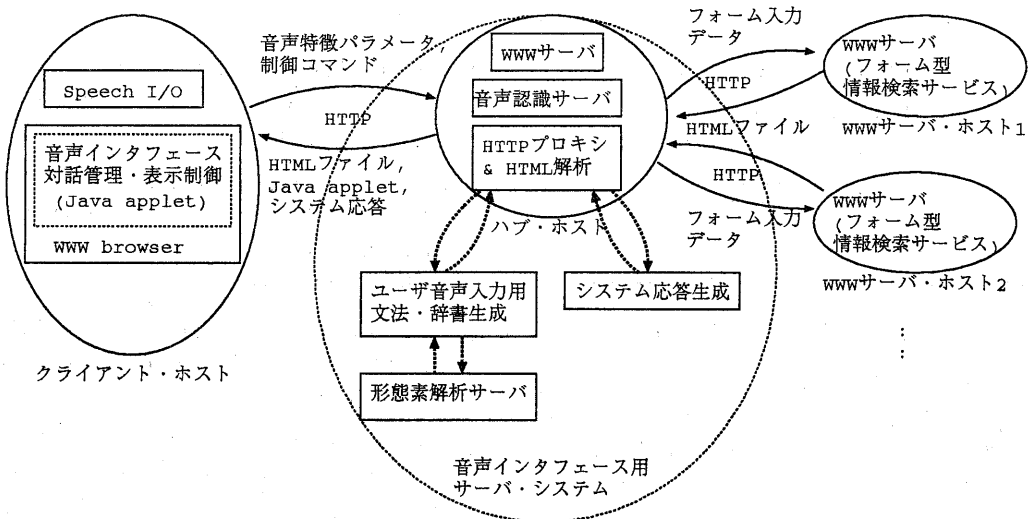


図 4: システムの構成

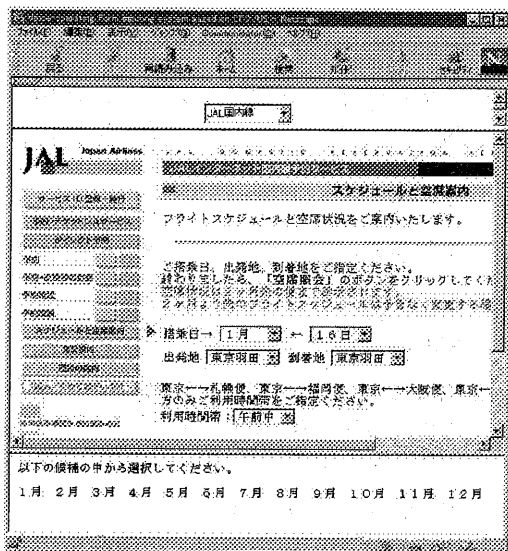


図 5: 音声インタフェースシステムにおける WWW ブラウザの表示例

の表示では、ユーザの発話毎に検索ページ内の入力箇所を指し示す点滅のポインタが移動し、キーワード一覧の表示も切り替わる。

### 3.3 システムの課題

既存の Web ページを対象として音声インタフェースの適用を考えたとき、HTML テキストが基本的

システム：「搭乗日を入力してください。」  
 ユーザ：「2月」  
 システム：（注：システムからのプロンプトなし）  
 ユーザ：「16日」  
 システム：「出発地を入力してください。」  
 ユーザ：「東京です」（注：‘東京羽田’を省略）  
 システム：「到着地を入力してください。」  
 ユーザ：「えーと、名古屋です」  
 システム：「利用時間帯を入力してください。」  
 ユーザ：「もどる」（注：到着地を訂正）  
 システム：「到着地を入力してください。」

図 6: 音声インタフェースによる対話例

に GUI を想定して記述されているため、特別な扱いが必要な場合や扱いが困難な場合がある。その例をいくつか述べる。

- セレクタの項目名の取得

項目名が取得できない場合、キーワード一覧を読み上げるような対応が必要 [10]。但し、ブラウザにキーワード一覧を表示されており、画面表示からの確認は可能。

- キーワードの読みの付与

キーワードが英語や記号などで、読みが与えられない場合、前の例 [9] と同じく、画面表示においてはキーワード一覧に番号を付与し、番号の音声入力によるキーワードの選択が可能。

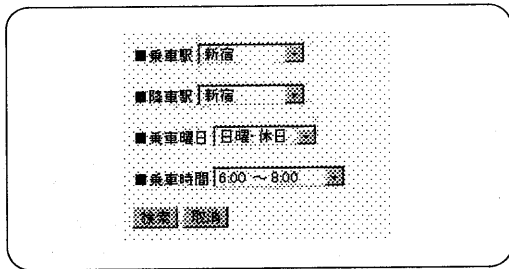


図 7: セレクトボックスの例

- キーワードの読み以外での入力

図 7 の「6:00~8:00」のような例において、音声入力では、「7 時」の発話で指定できることも必要と考えられる。

- セレクトボックス以外のフォーム入力

情報検索のページには、セレクトボックス以外に、ボタン、チェックボックス、テキストボックス等を含む場合がある。特に前者 2 つについては、項目名についての情報の抽出が困難であり、特別な音声操作インタフェースが必要である。従って、マウス、キーボード等との併用によるユーザインタフェースも考える必要がある。音声インタフェースとブラウザをより融合するには、音声入力や携帯端末でのユーザインタフェースを想定し拡張されたマークアップ記述言語が望まれる。

## 4 むすび

WWW 上のフォーム型情報検索サービスを対象として、特に、フォームのメニュー選択操作の GUI に関して、汎用的な音声インタフェースを提供するシステムを試作した。試作したシステムは、調査を行なった 7 種類の情報検索向けのページのうちフレーム構造を含まない 5 つのページにおいて、メニュー選択操作の GUI に対して音声インタフェースとして処理できることを確認した。今後、携帯情報端末およびデスクトップ型 PC の画面情報量、操作環境、操作方法などの違いを考慮して評価実験を行ない、従来の GUI による操作との操作性や操作効率の違いについて調査する予定である。

## 参考文献

- [1] 淵 武志, 加藤 恒昭: WWW ブラウザの音声による制御, 情報処理学会研究会資料, SLP-16-7

(1997.5).

- [2] Alex Rudnicky, et. al.: "Speechwear: A mobile speech systems," *Proc. of ICSLP'96*, Philadelphia (1996).
- [3] R. Lau, G. Flammia, C. Pao, and V. Zue: "WEBGALAXY - Integrating spoken language and hypertext navigation," *Proc. of EUROSPEECH'97*, pp.883-886 (1997).
- [4] Sunil Issar: "A speech interface for forms on WWW," *Proc. of EUROSPEECH'97*, pp.1343-1346 (1997).
- [5] 木村 貞弘, 中村 哲, 鹿野 清宏: MOSAIC ブラウザを用いた音声対話システム, 情報処理学会第 52 回全国大会講演論文集, 2-407 (1996).
- [6] <http://www.ibm.co.jp/pspjinfo/voice30/>
- [7] 西本 卓也, 小林 豊, 新美 康永: ネットサーフィンにおける音声コマンド候補の生成について, 信学技報, SP97-59 (1997.11).
- [8] 近藤 和宏, チャールズ ヘンブル: 音声認識を用いた WWW ブラウザとその評価, 電子情報通信学会論文誌, Vol.J81-D-II, No.2, pp.257-267 (1998).
- [9] 甲斐 充彦, 中野 崇広, 中川 聖一: 音声認識サーバ-SPOJUS-を利用した WWW ブラウザの音声操作システム, 情報処理学会研究会資料, SLP20-14 (1998.2).
- [10] 西本 卓也, 小林 豊, 新美 康永: WWW 上のデータベース検索のための汎用音声インタフェース, 音講論集, 2-Q-20 (1997.3).
- [11] <http://voxml.mot.com/>
- [12] 甲斐 充彦, 伊藤 敏彦, 山本 一公, 中川 聖一: 自然な発話を対象としたパソコン/ワークステーション用連続音声認識ソフトウェア, 音講論集, 2-Q-30 (1997.9).
- [13] <http://www.slp.tutics.tut.ac.jp/SPOJUS/>
- [14] <http://cactus.aist-nara.ac.jp/lab/nlt/chasen.html>