

## フォーム型情報検索サービスのための 音声/ペンタッチ入力インタフェースの比較評価

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> 静岡大学工学部 システム工学科

<sup>†</sup> 〒 441-8580 愛知県豊橋市天伯町字雲雀ヶ丘 1-1

<sup>††</sup> 〒 432-8561 静岡県浜松市城北 3-5-1

### 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作成した。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。本システムにおいて音声入力とペンタッチ入力について比較実験を行ったところ、実際の擬似的な利用環境によって音声入力とペンタッチ入力はほぼ同等の評価が得られた。

### Comparison of pen-touch and speech user interfaces for form-based information retrieval services

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

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

1-1, Hibarigaoka, Tenpaku-cho , Toyohashi-shi, Aichi, 441-8580, Japan  
{nakano,nakagawa}@slp.tutics.tut.ac.jp

<sup>††</sup> Department of Systems Engineering , Faculty of Engineering , Shizuoka University  
3-5-1, Johoku, Hamamatsu-shi, Shizuoka, 432-8561, Japan  
kai@sys.eng.shizuoka.ac.jp

### Abstract

Recently, we can access a variety of information retrieval services on the WWW. However, they enforce the users to access via a WWW browser which is basically based on a graphical user interface. In this study, we implement 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. In the experiment evaluating a usability of both pen-touch or speech interface systems, we observed almost comparable usability results for both interfaces in typical user operating conditions.

## 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。また、携帯電話やPDA(携帯情報端末機器)の普及により、ペン入力やボタン操作などでWWWサービスが利用できるようになるなど、WWWを取り巻く環境が多様化している。このことから、WWWブラウザにおける機能拡張として、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。桂浦らのシステム[1]では、WWW上のリンク情報からキーワードを抽出し、キーワード発話を音声認識することによりネットサーフィンを実現するシステムを実現している。近藤らのシステム[2]は、WWWブラウザに音声コマンドや音声ブックマークなどの機能拡張を行い、かつHTMLの記述を拡張し、音声によるフォーム入力を容易にする仕組みを持ったシステムを提案している。また、文献[3]のシステムでは、WWW上のアプリケーションを、音声で利用できるように対話的インタフェースを開発している。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[4]。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。そのため、WWW上の情報検索サービスで使われているフォーム型情報検索ページに対する音声操作インタフェースを試作した[5][6]。

本研究では、Webページ内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システムを作成した。また、選択メニューのようなGUIベースの直接操作(direct manipulation)型インタフェースにおいて、単純な作業ではマウスと音声入力で、同等な操作効率を得られることが示されているが[7][8]、本研究では、より身近な既存のWWW上の情報検索サービスに対して、インタフェースの比較を行うことを目的としている。

作成したシステムでは、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。そのため、WWWブラウザで動作するJava applet および JavaScript 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

図1: 情報検索ページのフォーム表示部分の例

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。また、本システムにおける評価として、利用状況の違いとユーザインタフェースの違い(音声入力とペンタッチ入力)に関して比較実験を行った。

## 2 フォーム入力のための音声インタフェースシステムの概要

### 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。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システム[5]の概要とシステム応答に関して拡張された点について述べる。

### 2.2 システム構成

図2に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、本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サーバ及び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[9, 10] を利用している。また、形態素解析サーバには茶筌[11]を用いた。

音声合成には、合成音をサーバ側で作成し、クライアント側にデータを転送することで音声出力を行う方法と、クライアント側に出力文字列のみを転送し、クライアント側マシンにインストールされている音声合成プログラムによって音声出力を行う方法とが考えられる。サーバ側で合成音を作成する場合、合成音データを転送する際に、回線容量の太さによって応答時間に差が出る可能性がある。一方、クライアント側による合成音の出力では、音声合成プログラムを事前に用意しておく必要がある事や、クライアント側マシンの処理能力によって応答時間が左右される可能性がある。

### 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
- 音声入力のリ認識用文法・辞書の生成
- 選択リストの入力項目名や種類の認識・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 [4] で用いた方法とほぼ同じように、キーワードの形態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 3 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

```

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 3: フォーム記述と辞書の生成の例

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。

フォーム内に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際 [5]、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。一方、項目名情報を取得しにくい記述例としては、(1) いくつかのセレクトが項目名を共有している場合 (2) 項目の情報を示すために画像を利用している場合、(3) 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、などがあつた。

応答文の生成については、得られた項目名に対し、

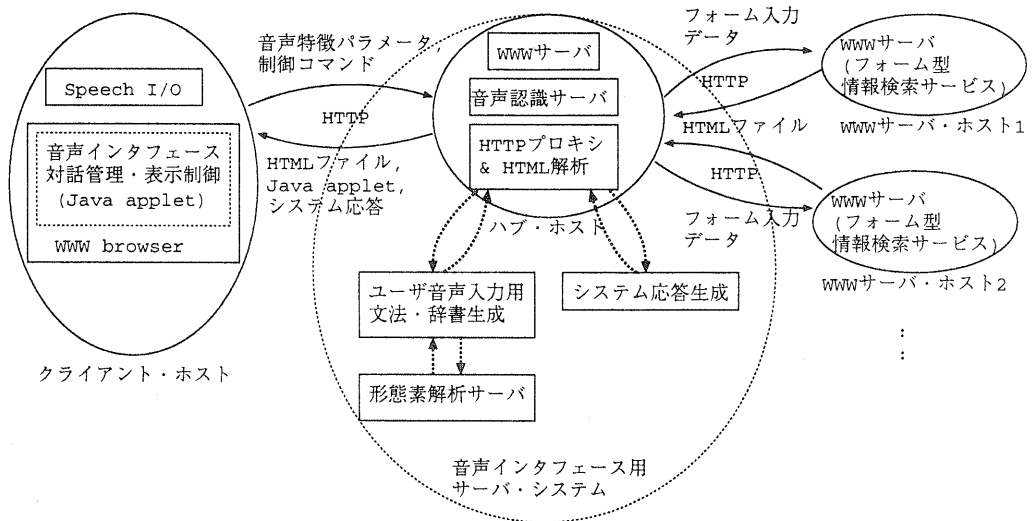


図 2: システムの構成

“～について入力して下さい”, “～について以下の候補から選択(発話)してください”, などというテンプレートと結びつけることによって対話応答文を生成することが可能である。また、項目名を更に解析することにより、より適した応答文を生成することも可能ではないかと考えられる。(例えば、項目名が出発日を表している場合、“出発日はいつですか?” という応答文を生成する)

## 2.5 システムの実行例

図 4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 applet 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、システムへの返答(はい、いいえ)等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。

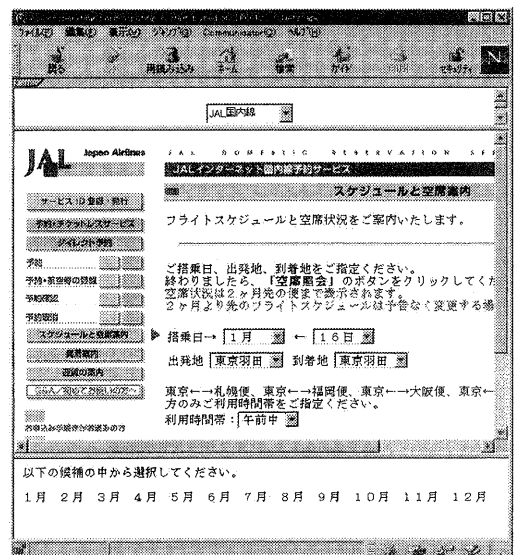


図 4: 音声インタフェースシステムにおける WWW ブラウザの表示例

図 5に、JAL のホームページ [12] 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。

また、本システムでは、WWW ブラウザからの音声認識サーバ、音声入力サーバとのインタフェースは、Java applet のネットワーク機能を用いて実

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 5: 音声インターフェースによる対話例

装している。このため、本システムを稼働させる際に、WWWブラウザでJava appletが動く環境であれば基本的に動作するため、クライアント側のOSに依存しにくい構造になっている。現在は、本システムのユーザ側アプリケーションのうち、音声入力サーバがWindows系OS(95/98/NT/2000等)またはUNIXで動作する為、両OS環境において利用できるようになっている。

### 3 評価実験

#### 3.1 タスク及び使用環境の設定

評価実験に用いたシステムは、クライアント側マシンとして、WWWブラウザの表示と音声の入力用に小型のノート型パソコン(Celeron 300MHz)を使用した。音声認識サーバ、形態素解析サーバ及びappletサーバなどのサーバ側マシンとして、PC-UNIXが動作しているデスクトップPC(PentiumII 300MHz)を使用した。また、クライアント側マシンで使用するWWWブラウザのソフトウェアとしてはWindows用のNetscape Communicator 4.61を利用した。

被験者には表1のような3種類のタスクを提示し、それぞれについて音声、ペンタッチ入力での実験を行った。

表 1: 実験タスク

タスク名	タスク内容
AT	飛行機の時刻表情報を検索
TT	列車の時刻表情報を検索
CR	料理のレシピ情報を検索

#### 3.2 被験者実験の手順

まず最初に被験者が処理の対象となるタスクを選択すると、システムはタスクの対象となるURLからHTMLテキストを取得し、フォームに関する情報を抽出した後、音声認識に必要な辞書や文法を一度に生成する。処理時間としては、一つのフォーム検索サイトの処理に必要な時間は平均で約13秒であった。

評価実験では、音声入力とペンタッチ入力について、立って利用した場合と座って利用した場合について、また文字の大きさがユーザに与える影響について評価する為に、2種類のフォントサイズを用い、6人の被験者に実際にシステムを利用してもらった。被験者のPC利用歴は平均2年で、6人のうち2人は音声入力を利用した事がなく、またWWWブラウザについても不慣れであった。そのため、全員に平均20分程度、音声入力方法やWWWブラウザの操作等について慣れてもらった後、実験を行った。

被験者を二つのグループに分け、最初のグループ(被験者ABC)は、ノート型PCを机に起き、椅子に座った状態で、文字の大きさが大きい状態で実験を行った。次のグループ(被験者DEF)は、ノート型PCを手に持ち、立った状態で、文字の大きさが小さい状態(モバイル環境)で実験を行った。その後、それぞれのグループについて実験条件を入れ替えて行った。

音声による検索の場合、ブラウザの操作は、マウスやキーボードを一切使わず、音声入力による手段に限定した。そのため、音声入力箇所への移動や認識誤りを修正するために、「つぎ」「戻る」などといったコマンド発話を受け入れるようになっている。

また、被験者実験終了後にユーザに対してアンケートを実施し、音声認識との関わり具合やシステム使用感、満足度などを調査した。

#### 3.3 評価実験結果

システムの評価については、実験終了後のアンケートで調査し、被験者が便利だ、あるいは満足したと感じたら5点、不便だ、使いづらいと感じたら1点という5段階方式を採用した。表2に被験者別のタスク完了までの所要時間(秒)とシステムの評価を示す。

各タスクにおける入力すべき項目数は3~5項目であり、各項目の語彙数(選択可能候補数)は最大で40であった。音声認識率は約90%、タスク達成率は100%であった。

動作時間については、音声入力の方がペンタッチ入力よりも時間がかかっている。これは、ペンタッチ入力の場合はペン入力を開始した時点からタスク終了(検索結果が出力される)までの時間であるの

表 2: 情報検索タスクの実験結果

座る, 文字サイズ大						
項目	音声			ペン		
	タスク	時間	評価	タスク	時間	評価
A	AT	201	3	TT	25	4
B	CR	123	2	AT	30	5
C	TT	149	4	CR	18	4
D	AT	154	4	TT	26	5
E	CR	71	5	AT	38	4
F	TT	115	4	CR	19	4
平均		135.5	3.6		26.0	4.3

立つ, 文字サイズ小						
項目	音声			ペン		
	タスク	時間	評価	タスク	時間	評価
A	AT	182	3	TT	32	4
B	CR	85	4	AT	35	4
C	TT	101	4	CR	16	4
D	AT	160	4	TT	30	4
E	CR	70	5	AT	40	3
F	TT	108	4	CR	25	4
平均		117.7	4.0		29.7	3.8

に対し、音声入力の場合は、HTML テキストの解析や辞書作成などによる前処理時間（約 13 秒）や、音声認識時間（約 3 秒）などが含まれること、また被験者が音声認識よりもペンタッチ入力の方が使い慣れていたことが原因ではないかと考えられる。

システムの評価については、音声入力については、座ろうが立とうが、あるいは文字の大きさによらず、それほど評価の差は見られなかったが、ペンタッチ入力については、「座る、文字サイズ大」の環境よりも「立つ、文字サイズ小」の環境になったとき評価が下がった。これから、ペンタッチ入力方式は、表示される文字が小さかったり、手に持って利用するというような環境では使いづらいということが言える。

また、「座る、文字サイズ大」の場合はペンタッチ入力の方が評価が高いが、「立つ、文字サイズ小」の場合は音声とペンタッチ入力はほぼ同等の評価が得られている。これらからも、音声入力は携帯情報端末にとって有効なインタフェースとなりうると考えられる。

## 4 むすび

本報告では、WWW 上のフォーム型情報検索における、音声インタフェースシステムの実装を行い、また音声とペンタッチ入力との比較実験を行った。本システムは、ペンタッチ入力を使用しづらい携帯型の移動端末などでの利用に有効だと考えられる。また、VoiceXML[13] などに見られる、電話や携帯情報端末での利用を想定した音声対話アプリケーションのために拡張されたマークアップ記述言語の

規格も提案されているが、既存の WWW 上の簡単なフォーム入力（選択）タスクに対して音声対話を実現することが可能である。今後の課題として、より多くのフォーム型情報検索システムに対応させる為に、テキスト入力ボックスへの音声入力を可能にすることが必要である。そのため、まずフォーム型情報検索システムにおいて、任意テキストの入力に対応できるように、固有名詞の入力について機能を追加していく予定である。

## 参考文献

- [1] 桂浦 誠, 中村 哲, 鹿野 清宏: 音声キーワードによる WWW のブラウジング, 情報処理学会論文誌, Vol.40, No.2, pp.443-452.
- [2] 近藤 和宏, チャールズ ヘンプヒル: 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.
- [3] Sunil Issar: "A speech interface for forms on WWW", Proc. of EUROSPEECH'97, pp.1343-1346.
- [4] 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).
- [5] 中野 崇広, 甲斐 充彦, 中川 聖一: 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, 情報処理学会研究会資料, SLP25-1 (1999.2).
- [6] 中野 崇広, 甲斐 充彦, 中川 聖一: WWW 上のフォーム型音声検索のための音声インタフェースの評価, 情報処理学会第 59 回全国大会講演論文集, 2-135 (1999).
- [7] Philip R. COHEN: "Natural Language Techniques for Multimodal Interaction", 電子情報通信学会論文誌, Vol.J77-D-II, No.8, pp.1403-1416 (1994).
- [8] Martin G. L.: "The utility of speech input in user-computer interfaces", International Journal of Man-Machine Studies, Vol.30, No.4, pp.355-375
- [9] 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).
- [10] <http://www.slp.tutics.tut.ac.jp/SPOJUS/>
- [11] <http://cactus.aist-nara.ac.jp/lab/nlt/chasen.html>
- [12] <http://www.jal.co.jp/>
- [13] <http://www.voicexml.org/>