

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

1K-6

中野崇広[†] 甲斐充彦^{††} 中川聖一[†][†] 豊橋技術科学大学 情報工学系^{††} 静岡大学工学部 システム工学科

1 はじめに

WWW(World Wide Web)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の利用が広がるにつれ、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。

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システムを作成した。

本システムは、WWWブラウザで動作するJava applet及び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い、評価実験を行ったので報告する。

2 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、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。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。

フォームの記述では、一般に、ユーザが指定する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できる。

音声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、各入力毎に順次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。表1に示す各要素のうち、セレクトボックスの場合は、現在のHTMLの書式において、項目名を自動的に抽出することが比較的容易である。

表1: フォームの各要素におけるキーワード

チェックボックス	対応する項目名
セレクトボックス	ボックス内で選択可能な項目名
テキストフィールド	自由文入力
ラジオボタン	ボタンのオン/オフ

3 システム構成

本システムは、大きく分けて音声認識サーバ、音声入力・分析サーバ、HTML文書解析部、ブラウザ表示・音声入力制御部からなる[1]。ユーザの計算機上では、音声入力・分析サーバと汎用のWWWブラウザのみをあらかじめ用意する。ブラウザ表示・音声入力制御部はJava appletで実装されており、音声認識サーバ、HTML文書解析部と共に、残りの部分はappletサーバとなるホスト上に実体を持つ構成になっている。音声認識はSPOJUS[2]、形態素解析には茶筌[3]を用いた。

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装している。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Windows95/98またはUNIX上で動作するため、Javaが動作するWWWブラウザを用いることで両OS環境において利用できるようになっている。

システムの動作例を図1に示す。ユーザが入力した音声をシステムが誤認識した場合は、「戻る」などのコマンド発話によって、直前の入力を取り消すことができる。

4 評価実験

システムの実装では、WWWブラウザの表示と音声の入力用にノート型パソコン(Celeron 300MHz)を使用し、音声認識サーバ、形態素解析サーバ及びappletサーバとしてPC UNIXが動作しているデスクトップPC(Pentium-II 300MHz)を使用した。

Evaluation of speech interface for form-based speech retrieval on the WWW

Takahiro Nakano[†], Atsuhiko Kai^{††} and Seiichi Nakagawa[†]

Toyohashi University of Technology[†]
Shizuoka University^{††}

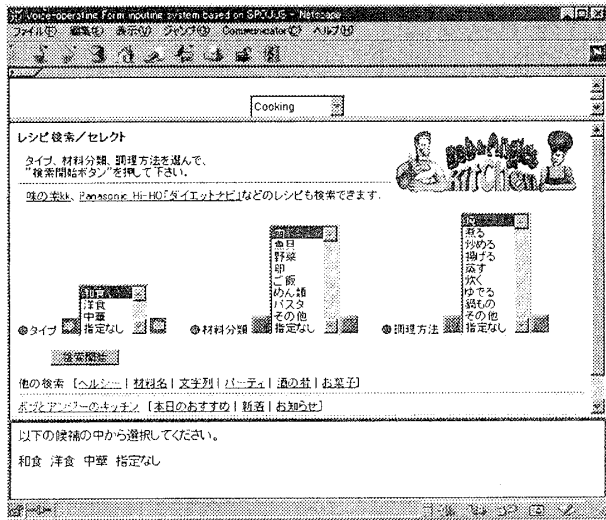


図 1: システムの動作例 (画面は 173mm × 130mm) WWW ブラウザのソフトウェアとしては Windows 用の Netscape Communicator 4.51 を使用した。

評価実験では、被験者が最初に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 文書を取得し、フォームに関する各項目を取得した後、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、後述の評価実験の結果では、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 8 秒であった。

実験では、システムの操作性や性能を評価する為に、四人の被験者に実際にシステムを使用してもらった。被験者の PC 利用歴は平均 2 年強で、四人のうち二人は音声入力を利用したことがなく、また WWW ブラウザの操作にも不慣れであった為、事前に 5 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。

被験者には表 2 の四つのタスクを提示した。被験者は二つのグループに分け、最初のグループ (被験者 A,B) は、まずペンタタッチ入力で行った後、音声による検索を行ってもらった。次のグループ (被験者 C,D) は、最初に音声による検索を行った後、ペンタタッチ入力による検索を行った。

表 2: タスクの内容

音声	
・小田急ロマンスカーのホームページを検索 (タスク VO)	新宿駅から御殿場駅までの料金を調べる
・料理のレシピのホームページを検索 (タスク VC)	洋食で、ご飯を炒めた料理のレシピの件数を調べる
ペンタタッチ	
・料理のレシピのホームページを検索 (タスク PC)	中華料理でめん類の料理レシピの件数を調べる
・小田急ロマンスカーのホームページを検索 (タスク PO)	小田原駅から向ヶ丘遊園駅までの料金を調べる

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。

表 3 に被験者別のタスク完了までの所要時間を示す。被験者 A,B のように、先にペンタタッチ入力を行い、後から音声入力を行ったグループの方が、先に音声入力を行ったグループよりもタスク完了に係る時間が短いことがわかる。これは被験者が音声入力に慣れていないためではないかと考えられる。また、音声入力方式よりもペンタタッチ入力の方がタスク完了までの所要時間が短くなっている。発話のタイミングの待ち時間や誤認識のための再入力 (音声認識率は 92%、なおタッチ入力ミスは 13.9%)、認識処理時間のためだと考えられる。

実験終了後、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。システムの評価については、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、遅い、使いづらいなどと感じたら 1 点という 5 段階方式を採用した。結果を表 4 に示す。この結果から、ペンタタッチ入力に比べて音声入力は、入力方式としてはそれなりに満足しているが、タスク終了までの時間が遅いことを不満に感じているようである。

表 3: タスク完了までの所要時間と発話/入力回数

被験者	VO	VC	PO	PC
A	84 秒 5 回	76 秒 4 回	15 秒 5 回	6 秒 4 回
B	79 5	85 4	14 6	8 4
C	129 6	113 5	18 6	10 4
D	112 6	89 4	20 8	7 4

表 4: アンケートによるシステムの使いやすさ

	音声	ペンタタッチ
システムの使いやすさ	3.2	4
入力のしやすさ	4	4.5
応答時間	2.1	-

5 むすび

本報告では、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システムとその評価実験の結果について述べた。本システムは、ペンタタッチ入力を使用しづらい携帯型の移動端末に有効だと考えている。今後の課題として、対話的なインタフェース機能の充実と、そのための拡張された汎用的なマークアップ言語の適用などが考えられる。

参考文献

- [1] 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, 情報処理学会研究報告 99-SLP-25-1, pp.1-6(1999-2)
- [2] <http://www.slp.tutics.tut.ac.jp/SPOJUS/>
- [3] <http://cactus.aist-nara.ac.jp/lab/nlt/chasen.html>