

WWW上のテキスト入力フォームのための 任意文字列入力の音声インタフェース

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。

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[1]、音声による効率的な操作の実現方法について検討した際、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[2][3]。このシステムは、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た、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ものである。

本研究では、インターネット上に存在する様々なWebページについて音声インタフェースを実現するために、特にテキスト入力フィールドへの音声による文字入力についてシステムの拡張を行った。

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、チェックボックス、ラジオボタンなどの種類がある。

WWW上で提供されている情報検索サービスは、フォームの形式により表1のように分類することが

できる。

表1: WWW上の情報検索サービスの分類

(1)	セレクトボックスなどによる条件選択型
(2)	テキスト入力フィールドによる条件入力型
(3)	(1)と(2)の併用型

これらのうち、(1)については情報検索における指定可能な属性に対する穴埋めタスクとして見ることができるため、従来のシステム[2][3]で対応可能である。しかし、(2)および(3)におけるテキスト入力フィールドの存在は、キーボードによる文字の入力を前提としているため、音声によって任意の文字を入力するためのシステムが必要になる。

3 任意文字列の音声入力

フォーム型情報検索システムにおいて、テキスト入力ボックスは検索条件を任意の文字列によって指定する場合に用いられる事が多い。例えば文献検索などの場合は、検索条件に著者名などの情報を入力することが多い。そこで、任意テキストの音声入力として、まず固有名詞の入力についての機能を追加することとした。

音声により固有名詞を入力する方法としては、n-bestを利用する方法を用いた。入力音声に対して認識を行う際に、

- (1) 音節単位ごとにn-bestを出力する
- (2) 音節列(文字列)単位のn-bestを出力する

という2種類の結果を出力し、認識結果のn-best出力を画面上に提示することにより、ユーザは音節や音節列の指定を行うことができる。

(1)(2)の出力候補の中に、目的の名前が含まれていなかった場合は、認識精度を上げるために

- (3) 音節ごとに区切って音声入力する

という入力方法を受け付ける。

また、認識精度を上げる為に、言語情報を利用する。この場合、

- (4) 出現頻度の高い名前については、登録語方式を用いて、事前に登録しておく
- (5) 音節のbigramやtrigramを用いる

という方法があり、これらと音響情報との併用による認識率の向上を図る。このような機能をシステムに付加することにより、テキスト入力ボックスへの固有名詞入力が可能になる。

Speech input interface of arbitrary character string for form-based text input field on WWW

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

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

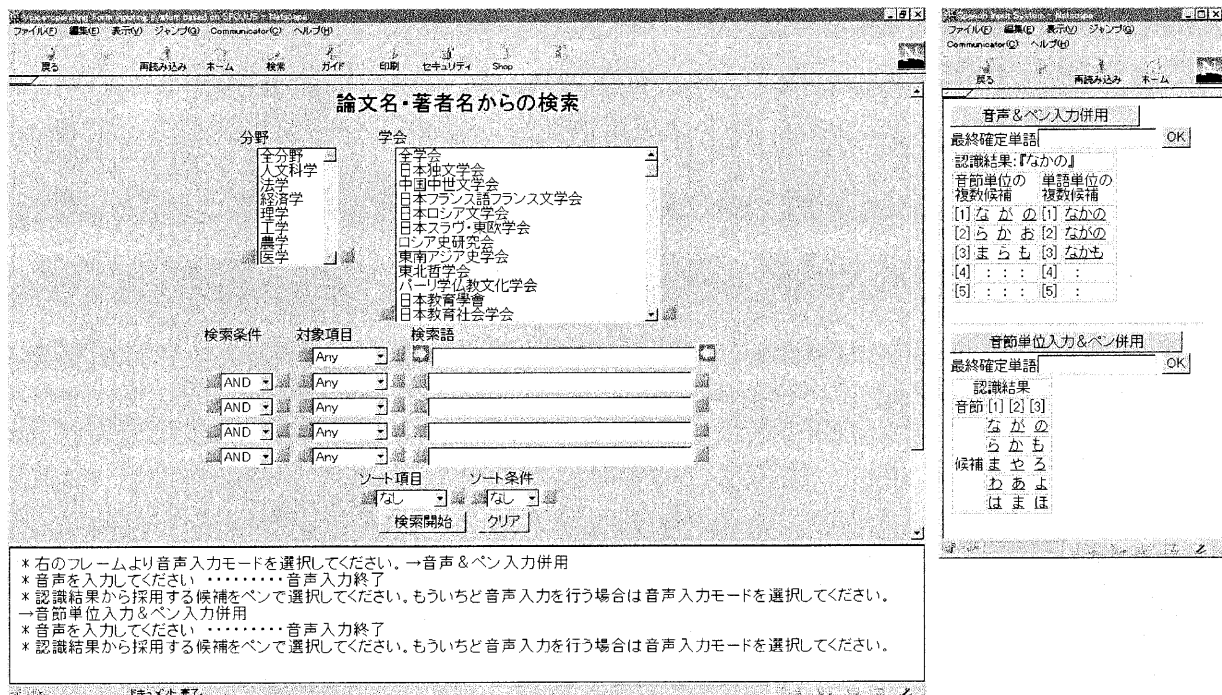


図 1: システムの動作例

4 システム構成

基本となるシステムは、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 [2]。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 Script および 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[4]、形態素解析には茶筌 [5] を用いた。

システムの動作例を図 1 に示す。

左側のウィンドウのうち、上部は WWW ブラウザの画面であり、ユーザに入力を要求する箇所を矢印 (→) で指示している。下部はシステムからユーザに対するプロンプトが表示される。GUI としてユーザに提供するのは主に右側の小ウィンドウであり、この中において、音声認識結果や認識候補の選択などを行う。

右側のウィンドウ上部の、「音声&ペン入力併用」ボタンを押すことにより、音声入力が可能になる。ユーザが音声を入力すると、音声認識結果として、音節単位の複数候補および文字列単位の複数候補 (n-best) が出力される。出力された複数候補をペンでタッチすることにより、最終確定単語フィールドに文字が入力されていく。任意の時点で OK ボタンを押すことにより、左側の当該個所に文字列が転送される。

この両方法においても、目的の文字列が入力できない場合は、右側のウィンドウ下部の、「音節単位入力&ペン入力併用」ボタンの動作に移る。これは、

ユーザが音声を入力する際に、「な」か「の」などのように音節ごとに区切って入力する必要がある。出力としては音節ごとに複数候補が表示されるので、出力された候補をペンでタッチすることにより、文字の入力が可能になる。

5 むすび

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 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 上のフォーム入力タスクに対して音声対話を実現することが可能である。

参考文献

- [1] 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).
- [2] 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).
- [3] 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).
- [4] <http://www.slp.tutics.tut.ac.jp/SPOJUS/>
- [5] <http://cactus.aist-nara.ac.jp/lab/nlt/chasen.html>