

# 音声認識による Web サーフィン統合インタフェース

1 K-5

呂 い 中西泰人 金井秀明 箱崎勝也  
電気通信大学大学院情報システム学研究所

## 1. はじめに

近年、ウェアラブルコンピュータが注目を浴びている。コンピュータを身に付け、行動しながら情報を操作する状況においては、マウスやキーボードのような入力手段よりも、音声入力のほうが有効である。このような状況を想定して、音声認識による WWW ブラウザの操作システムを試作した。

最近、音声入力のインタフェースを持った WWW ブラウザが多く試作されている[1,2]。ここで試作したシステムは、ブラウザを監視するプログラムを起動し、ウェブブラウザから入手した HTML を解析し、アンカー部の抽出とその読みの指定を行って、音声認識機能を起動する。

## 2. インタフェースの設計

本システムは、WWW ページのリンクを音声で読み上げると、音声認識を行い、指定されたリンクのページへジャンプする。ブラウザ自体の操作（例えば、前ページに戻る、進む、ホーム、リロード、終了など）を音声でコントロールすることも可能である。

試作した WWW ブラウザと音声認識の統合インタフェースの画面を図1に示す。一つのリンクに対する音声入力方法には次の三つがある。

1. リンクリストの番号を読み上げる
2. リンクをそのまま読み上げる
3. リンクに対応する表記の第一単語(短縮語と呼

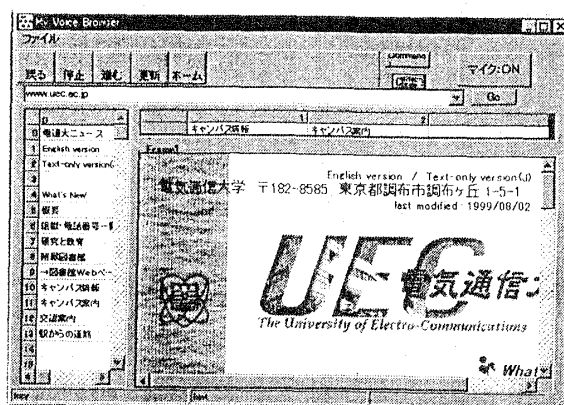


図1 統合インタフェースの画面

ぶ) を読み上げる

1の場合、リンクリストに Web ページ中のすべてのリンクに番号をつけ、表示する。その番号の発声でリンクできる。表記が英語混じりであったり、発声できない記号の場合、読みが困難なため、リンク一覧の番号の音声入力は無効である。

2の場合、リンクに対応する表記の全部を発声してリンクする。ページの中のリンクを見ながら読み上げるので、ユーザにとって自然である。

3の場合、同じ短縮語を持つ複数のリンクが存在するとき、すべてのリンク候補を候補リストに番号をつけて、表示する。ユーザは候補リストの番号を音声入力して、該当のページを指定する。

例えば、「きゃんばす」と発声したとき、今表示しているページに「キャンパス情報」と「キャンパス案内」の二つのリンクがあるとすると、この二つに番号を付けて、候補リストに表示する。

	1	2
キャンパス情報		キャンパス案内

候補リスト

次にユーザが番号を発声して、選択する。

ユーザが希望の短縮語だけを音声入力し、システムがすべてのリンク先候補を表示してくれるので、リンク先の検索が容易であり、また画面に表示しき

れない場合にも有効である。

### 3. 音声操作システムの実装

システム構成と動作の仕組みの概要を図2に示す。

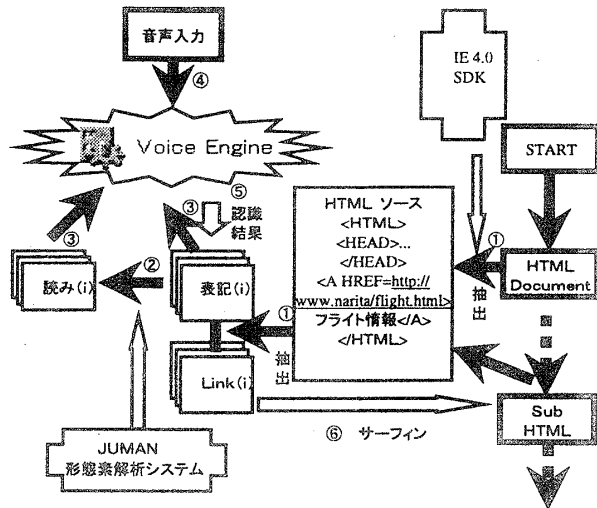


図2 システム構成

システムは次のような順序で動作する。

- ①. ブラウザが表示しているページのHTMLソースファイルからアンカー部の抽出を行う。HTMLの書式からアンカー部<A HREF="..."></A>で挟まれた文字列を抽出し音声認識の対象とする。
- ②. 発声の対象となる表記を抽出後、形態素解析を行って、表記の読み方を指定する。
- ③. 読み方をすべて音声認識エンジンの辞書に登録する。同時にリンクの表記に番号を付けて、リンクリストのフレームに表示する。
- ④. 音声認識エンジンが音声入力の受付開始を要求する。
- ⑤. 音声認識エンジンは入力音声を逐次処理し、認識結果と一致するすべてのリンク表記をブラウザの候補リストフレームに表示して、音声入力待ち状態に戻る。
- ⑥. 音声認識エンジンはユーザが発声した番号を認識し、次のジャンプする先のURLを決定する。
- ⑦. 手順①に戻る。

試作システムは、Internet Explorer SDK[3]、

Juman形態素解析システム[4]、NEC VOICE音声認識エンジンのコンポーネントを組み込んで、Visual Basicで実現した。

### 4. 考察

パーソナルコンピュータ上で3つの音声入力方法でWWWブラウジングの評価実験を行っている。その結果、リンク番号による入力は、発声対象は数字なので、相対的に短く、ユーザにとって発音が楽であるが、ページの中のリンクをそのまま発音できず、リンクリストを探して、その番号を読み上げる必要があるため、不自然さがある。リンク表記が発声できない記号混じりの場合、リンク番号による方法は一番有効である。

リンクそのまま読み上げによる入力は、一番自然だが、画面に表示しきれない場合、あるいは発声できない記号混じりのリンクの場合には入力が困難である。

短縮語による入力は、希望の短縮語だけ入力し、システムはすべてのリンク先候補を表示してくれるので、ホームページのリンク数が多い場合、また画面に表示しきれない場合、希望のリンク先の検索が容易である。しかし、システムが短縮語の認識と候補リストの番号の認識を二段階で処理するので、リンクをそのまま読み上げる方法と比べると、少し時間がかかる。

### 5. おわりに

この3つの音声入力方法の使いやすさと認識率の関係、周囲の雑音の影響、話者発音の学習による効果などの評価は今後の課題である。

### 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-を用いたWWWブラウザの音声操作システム”、情報処理研究会資料、SLP-20-14(1998.2)
- [2]野波 成：“ブラウザと音声認識システムの統合インタフェース”、システムソフトウェアとオペレーティング・システム、78-7 (1998.5)
- [3]<http://www.microsoft.com/msdn/sdk/inetsdk/>
- [4]<http://www.nagao.kuee.kyoto-u.ac.jp/nl-resource/>