

## 音声キーワードによる WWW のブラウジング

桂 浦 誠† 中 村 哲† 鹿 野 清 宏†

WWW ブラウザに音声認識を組み込んだシステムとして、マウスをクリックする代わりにリンク項目を音声認識を用いて入力するシステムが考えられる。しかし、これだけではマウスによる操作を音声入力で行うにすぎず、入力手段としての音声を有効に利用しているとはいえない。本研究では、現在表示されているページからある程度ページの先読みを行い、音声キーワードにより効率的な情報検索を実現するシステムを提案する。また、システムの有効性を検証するために被験者を用いて実験を行い、朝日新聞のホームページを用い、指定した記事を検索するタスクを達成するまでの所要時間、マウスをクリックした回数を計測した。その結果、ページの構造が複雑な場合にはユーザが要した所要時間も短く、マウスをクリックする回数も少なく、音声キーワードを用いたほうが効率的な情報検索が行われていることが分かった。さらに、利便性に関する主観評価を行った結果、5段階評価で4.8という値が得られ、音声キーワードの有効性を示すことができた。

### The WWW Browser System with Spoken Keyword Recognition

MAKOTO KATSUURA,<sup>†</sup> SATOSHI NAKAMURA<sup>†</sup>  
and KIYOHIRO SHIKANO<sup>†</sup>

The use of speech recognition for WWW browsers has been investigated. The conventional systems realized the speech input for the hyperlink selection in a current page instead of the mouse operation. However, it is nothing but the replacement of the mouse and keyboard operation with the speech input. In order to make full use of speech, we propose a speech interface to a WWW browser which enables us to retrieve information efficiently by recognizing speech keywords. The proposed system looks ahead Web contents linked to the current page, and acquires the vocabulary for speech recognition dynamically. We measured completion time and the number of mouse clicks required to complete tasks of searching for the target Web page using the specified input modality in the Asahi News Paper site. As a result, when hyperlink structures are complicated, the proposed system reduced both completion time and the number of mouse clicks compared with the system only using the mouse operation. It is confirmed that recognition of speech keywords are very useful in that condition. We also carried out the subjective opinion test. User's reports say that the WWW browser system with a complementary use of the speech keywords input and the mouse is more convenient than that only with the mouse.

#### 1. はじめに

近年、インターネットの普及にともない、ネットワーク上には膨大な情報が存在している。そこで、ユーザが自分にとって必要な情報を効率良く、しかも快適に検索する方法が求められている。現在、インターネット上の情報を検索する手段として、Yahoo (<http://www.yahoo.com/>) などのディレクトリサービスや Altavista (<http://altavista.digital.com/>) などのキーワード検索システムが実現されており、世界中の情報にアクセスすることが可能となってきている。

一方、音声認識の分野においては、長期間の研究の結果、認識性能はめざましく向上し、一般のパーソナルコンピュータ上で高速に連続音声認識が可能となったり、カーナビゲーションなどに応用されるまでになっている。

現在、通常の WWW (World Wide Web) ブラウザに音声認識システムを組み込み、マウスやキーボードだけでなく音声によっても操作できる音声ブラウザがいくつか実現されている<sup>1)~7)</sup>。入力手段としての音声を考えると、コンピュータの初心者、熟練者に関係なく自然に利用でき、人とコンピュータのインタフェースとして非常に適していると考えられる。しかし、これらのシステムでは、音声認識により操作を行う対象のページ (以後、対象ドメインと呼ぶ) が固定であっ

† 奈良先端科学技術大学院大学情報科学研究科  
Graduate School of Information Science, Nara Institute  
of Science and Technology

たり、現在画面に表示されている情報しか音声による入力ができないといった問題点を抱えている。マウスやキーボードの代替手段としてのみでは、音声の利用方法として十分ではなく、効率的になっているとはいえない。これらをふまえて、過去に見たページの情報も直接音声認識を行うことができるようなシステムも実現されている<sup>4)</sup>。

本研究では、音声を利用して、ネットサーフィンを行いながらインターネット上の情報を効率的に検索できるシステムの実現を目的としている。単純に現存する検索エンジンのようにインターネット上からできる限りの情報を収集し、そのまま認識辞書を構成することも考えられるが、語彙数が膨大になり現実的ではない。また、検索エンジンを使用した際に目的とは関係のないページが数多くヒットするのと同様に、ユーザにとって不必要な語彙もかなり含まれている。

そこで本論文では、現在表示されているページからある程度ページの先読みを行い、動的に認識語彙を獲得しながら、ユーザによるキーワードの音声入力により目的のページに直接ジャンプすることを可能とする方法を提案する<sup>8)~10)</sup>。この方法により、ユーザは冗長なマウス操作を避けることができ、効率的な情報検索を行うことが可能となる。

本論文では、これらのうち、ページの先読みとキーワードの音声入力について述べる。さらに、音声入力の原理的な有効性を確認するための被験者による評価実験について述べる。以下、2章で提案手法について詳細に説明し、3章でシステムの構成について述べる。さらに、4章で評価実験を行い、最後に5章でまとめと今後の課題について述べる。

## 2. 音声キーワードを用いたネットサーフィン

### 2.1 WWW ブラウザと音声認識

インターネットの情報にアクセスするのにWWWブラウザは非常に便利なアプリケーションであるが、その設計方針は既存の入力方法(マウス、キーボード)に基づいている。つまり、快適に操作を行うためにはある程度の慣れが必要である。また、マウスやキーボードの操作は、その性質上、現在表示されている画面に制限される。

一方、入力手段としての音声を考えて、コンピュータの初心者、熟練者に関係なく自然に利用でき、人とコンピュータのインタフェースとして非常に適していると考えられる。また、音声入力は現在表示されている画面にとらわれることなく入力を行うことが可能である。よって、ユーザインタフェースの観点から見

て、音声入力をWWWブラウザに組み込み、種々の入力手段を有効に組み合わせるマルチモーダル化が非常に有益だと考えられる。さらに、現在表示されているページからある程度のページの先読みをしながら、動的にページ情報を獲得し、局所的な検索エンジンを実現できれば便利である。

本研究では、マウスやキーボードなどの既存のインタフェースに音声入力を加え、ユーザがネットサーフィンを行いながらインターネット上の情報を効率的に検索できるシステムを提案し、その中で、音声入力の原理的な有効性についての検討を行っている。

### 2.2 従来の音声ネットサーフィン

現在、WWWブラウザに音声認識を組み込んだシステムがいくつか提案されている<sup>1)~3)</sup>。これらの音声ブラウザの主な機能としては以下のようなものがある。

- (1) 特定ドメイン内の情報検索
- (2) 現在表示されているページのリンク項目の認識
- (3) ブラウザ機能の操作

(1)は、特定のページ範囲内の情報検索を音声で行うシステムである。単語単位で正規文法に基づき認識を行う音声認識技術を用いる場合、あらかじめ検索の対象となるページについて、認識語彙、文法を規定する必要があり、対象ドメインが特定される。しかし、実際のインターネットの情報は、時々刻々変化し、また、世界中のページにリンクされることが特長である。したがって、特定ドメインの検索では、インターネットのWWWブラウザの機能を生かしているとはいえない。

(2)は、現在表示されているページに書かれているリンク項目をマウスでクリックする代わりに、音声を用いるという方法である。確かにマウスを使用しなくてもよいので初心者などには便利だと思われるが、リンク項目が文章のように長いときには、読み上げに時間がかかる、読み上げの誤りが生じる確率も高くなることなどの問題があり、必ずしも効率的でない。その点、マウスクリックでは、一定の短い速度で操作でき、選択の確認などが即座にできる点で優れている。

(3)は、ブラウザの機能の操作を音声で行うものである。つまり、キーボードショートカットの機能を音声で実現したものである。ブラウザにボタンとして実現されている操作(バック、フォワード、リロードなど)については、(2)と同様にマウス操作のほうが効率的である。逆に、ブックマークの選択などでは、マウスでは何段階かの操作を強いられるので音声を用いたほうが便利であると思われる。

さきほど述べたように、ブラウザの操作における音

声入力のメリットは、キーボードやマウスの複雑な操作を音声入力により効率化することができる点にある。特に、ブックマークや現在表示されていないページの情報検索については、キーボードやマウスより、効率化できる可能性がある。また、現在の音声ブラウザの機能で不足している点は、不特定ドメインの情報検索である。インターネットの WWW ブラウザを実現するためには、ときどき刻々変化し、いろいろなところにリンクが張られている不特定で動的なドメインの検索を行う必要がある。

### 2.3 提案手法

本論文では上記の音声入力の特徴をふまえたうえで、音声認識を有効に利用しユーザインタフェースを改善する WWW ブラウザを実現する手法として、

- リンクの文章単位でなくリンク中のキーワードの抽出
- 不特定話者音素モデルの利用による任意キーワードの認識
- 動的なページ先読みによる不特定ドメインの検索を備えた音声 WWW ブラウザを提案する。

#### 2.3.1 ページの先読み

先ほど述べたように、音声ブラウザには、不特定ドメインを検索する機能が不可欠である。しかし、現状では、不特定ドメインの任意単語の大語彙音声認識は、非常に困難である。一方、実際にページ検索をする状況では、現在のページにリンクされている先のページの中から所望の情報を取り出す利用形態が非常に多い。そこで、本論文では、このような利用形態に注目し、動的にドメインの焦点を移して、認識対象語彙を変更しながら、認識語彙数を制約しながらも不特定ドメインに対処するページ先読み方式を導入する。

次に、ページ先読みについて説明する。現在表示されているページのリンク項目に対して音声認識を行うことは、1回のマウスのクリックを1度の音声発話に置き換えることである。通常、目的のページに到達するためには複数回のマウスクリックを必要とする。そこで、これを1度の音声発話で置き換えることを考える。そのためには、あらかじめある程度の深さまでのページを先読みし、リンク項目に対して音声認識を行えるようにすればよい。図1において、現在表示されているページからページ先読みを行い、網掛けの部分のリンク項目を認識辞書とする。ユーザの欲しい情報が図中の②にあるとすると、①のページ中のリンク項目は認識辞書に入っているため、3回のクリックが1度の音声発話で置き換えられることが分かる。

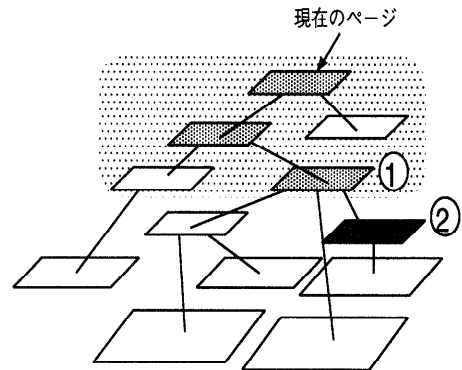


図1 ページの先読み

Fig. 1 WWW page lookahead.

#### 2.3.2 音声キーワードによるリンク項目の認識

認識辞書の作成において、リンク項目の文字列をそのまま読み直して認識辞書としてもよいが、先読みしたページのリンク項目をユーザがそのまま発話するとは考えにくい。また、現在表示されているページのリンク項目を見て音声で入力し、認識させるとしても、リンク項目が長くなったり、読みにくい字が出てきたりした場合には不便である。そこで、リンク項目の中からキーワードとして名詞を抽出し、キーワード集合を発声することによりリンク項目を認識できるようにする。

たとえば、「温暖化防止京都会議始まる」というリンク項目があったとする。認識辞書には、「おんだんか」、「ほうし」、「きょうと」、「かいぎ」の4つの名詞が加えられる。ユーザは「きょうと、おんだんか」のようにキーワード集合を発声すれば、このリンク項目が認識される。

以上により、ユーザは現在表示されているページからキーワードを連想し、音声入力することによって、目的のページに直接たどりつくことができる。

以下、キーワードという言葉は、リンク項目の文字列から抽出した名詞を指すこととする。

## 3. システム構成

### 3.1 システムの概要

本システムは、UNIX ワークステーション上で実現されており、図2に示すように大きく分けて次の3つのモジュールから構成されている。図中の矢印はデータの流れを表す。

- キーワード生成モジュール
- 音声認識モジュール
- インタフェースモジュール

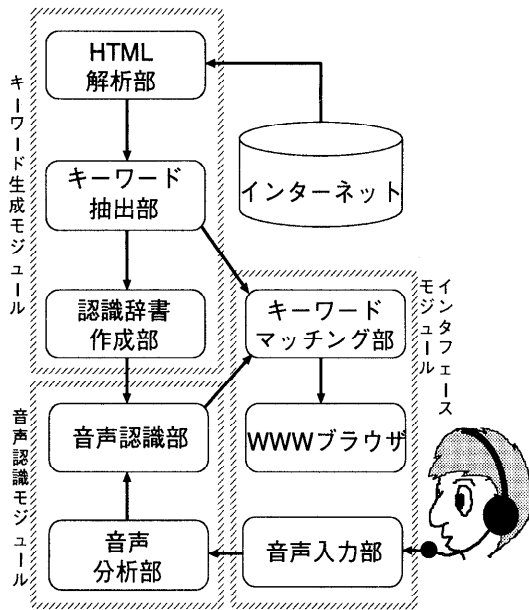


図2 システム構成  
Fig. 2 System architecture.

ユーザが発話を行う前に、システムはページを先読みして認識語彙を獲得しなければならない。キーワード生成モジュールでは、現在表示されているページからページの先読みを行いながらリンク項目のテキストを抽出し、認識辞書を抽出する。音声認識モジュールでは、入力された音声を分析し、音声認識を行う。インタフェースモジュールでは、キーワード生成モジュールで抽出されたキーワードと音声認識結果を照合し、次のリンク先の候補を提示する。以下に各モジュールの詳細を述べる。

3.2 キーワード生成モジュール

3.2.1 HTML 解析部

HTML 解析部では、URL と階層数を指定して、その URL から指定した階層数だけインターネットからファイルをダウンロードする。あるファイルをダウンロードすると、そのファイルからタグを解析してリンク項目の文字列、リンク先の URL を抽出する。図3に例を示しながら説明する。ダウンロードしたページに、

```
<A HREF="http://www.zzz.jp/weather">
  明日の天気</A>
```

という記述があったとする。「<A HREF」で始まるタグは次のリンク先を示すので、ここから URL およびリンク項目の文字列を抽出すると、

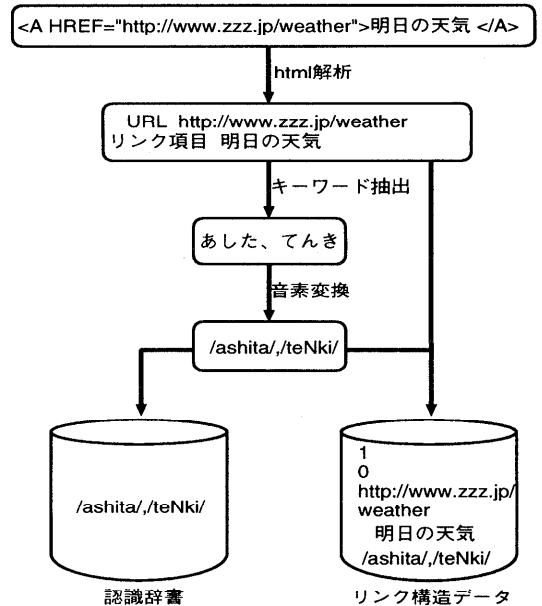


図3 キーワード抽出の例  
Fig. 3 An example of keyword extraction.

```
URL      : http://www.zzz.jp/weather
リンク項目 : 明日の天気
```

となる。

3.2.2 キーワード抽出部

キーワード抽出部では、形態素解析システム「茶筌」<sup>11)</sup>を用い、リンク項目の文字列からキーワードとして名詞のみを抽出し読みをふる。例では、リンク項目である「明日の天気」から名詞を抽出し、

```
あした、てんき
```

という2つのキーワードが抽出される。

3.2.3 認識辞書作成部

認識辞書作成部では、抽出されたキーワードを音素列に変換し、音声認識で用いる認識辞書を作成する。例では、キーワード「あした、てんき」を音素単位に変換し、

```
/ashita/,/teNki/
```

が認識辞書に登録される。

3.2.4 リンク構造データ

キーワード生成モジュールでは、認識辞書とは別にキーワードマッチング部で用いるリンク構造データを

0 0 <http://www.asahi.com/main.html> TOP PAGE  
 9 0 <http://www.asahi.com/cinfo/index.html> 当社概要と採用  
 案内 [tousha- gaiyou saiyou aNnai](#)  
 3 0 <http://www.asahi.com/flash/flash.html> ニュース速報 [nyu-  
usu soku+hou](#)  
 1 0 <http://www.asahi.com/home.html> トップ・ニュース [toQpu  
nyuusu](#)  
 23 0 <http://www.asahi.com/information/copyright.html> 著  
 作権 [chosaku+keN](#)  
 15 0 <http://www.asahi.com/information/howto.html> asahi.  
 com の使い方 [tsu+kaikata](#)  
 8 0 <http://www.asahi.com/information/information.html>  
 朝日新聞のお知らせ [asahi+shiNbuN oshirase](#)  
 81 3 <http://www.asahi.com/flash/fnational.html> #fnational  
 \_140 「キネ旬」ベストテン、邦画 1 位は「うなぎ」 [besu+toteN  
honga unagi](#)

図 4 リンク構造データ  
 Fig. 4 Link structure table.

作成する。これは、各リンクのインデックス番号、親ページのインデックス番号、URL、リンク項目文字列、キーワードの音素表記をセットにしたもので、どのキーワードがどのリンクに対応しているかを示すものである。図 3 においては、

1 (リンクのインデックス番号)
0 (リンクの親のインデックス番号)
<a href="http://www.zzz.jp/weather">http://www.zzz.jp/weather</a> (アドレス)
明日の天気 (リンク項目)
/ashita/,/teNki/ (キーワードの音素表記)

がリンク構造データとなる。実際のリンク構造データの例を図 4 に示す。

### 3.3 音声認識モジュール

音声認識モジュールでは、ユーザからの入力音声およびキーワード生成モジュールで作成された認識辞書を用いて音声認識処理を行う。分析条件を表 1 に示す。音声認識部では、音声分析部で生成した特徴パラメータを基に HMM (Hidden Markov Model) を使用して認識を行っている。本システムにおいては、HMM は音素単位のモデルを使用している。認識部の仕様を表 2 に示す。HMM の出力確率分布には、すべての状態で共通の確率密度分布を持ち、重みを変えることで頑健かつ精密に音素を表現できる Tied-Mixture 型 HMM を用いている。分布の数は、メルケプストラムが 256、 $\Delta$  メルケプストラムが 256、 $\Delta$  対数パワーが 64 である。また、音声認識の文法には、正規文法を用

表 1 分析条件

Table 1 Analysis conditions.

サンプリング周波数	12 kHz
分析窓長	32 msec
シフト幅	8 msec
特徴パラメータ	メルケプストラム 16 次元 $\Delta$ メルケプストラム 16 次元 $\Delta$ 対数パワー 1 次元

表 2 音声認識部

Table 2 Specification of the speech recognizer.

モデル	Tied-mixture 型 HMM
モデル数	音素単位 54
学習データ	日本音響学会文章データ (64 人分)
216 単語認識率	94.9 %

いており、抽出されたキーワードが任意回数あらゆる組合せで接続可能な文法としている。これは、複合語の場合のように、キーワードがユーザの発話の中に埋め込まれて発話された場合にもある程度の対処をするためである。さらに、実際にシステムを使う際、周りには環境雑音が存在し、またマイクの伝達特性なども認識に影響を与える。本システムでは、これらの環境に比較的簡単に対処できる E-CMN/CSS 法<sup>12)~14)</sup>を用いている。

### 3.4 インタフェースモジュール

インタフェースモジュールでは、ユーザとシステムの間で音声データやページのアドレスなどの受渡しを行う。図 5 は本システムで用いているインタフェースモジュールの検索結果表示画面である。このインタフェースは Tcl/Tk を用いて作成されている。

#### 3.4.1 キーワードマッチング部

キーワードマッチング部では、音声認識結果とキーワード抽出部で作成したリンク構造データを照合し、ユーザに次のリンク先の候補を提示する。リンク先の候補はリンク項目をラベルとするボタンで実現され、マッチしたキーワードの多い順に表示される (図 5 下部)。また “Score” は、マッチしたキーワードの数を表す。これにより、複合語が分割されて、認識辞書に複数のキーワードとしても登録された場合でも、所望のページが上位に提示されるようになる。ユーザはボタンを押すことで希望のページに直接ジャンプできる。この際、ボタンには直接ジャンプできるページへのリンク項目だけではなく、途中経路のリンク項目も同時に表示されている。よって、ユーザはどのような経路をたどってジャンプするかをあらかじめ知ることができ、現在自分がどこにいるのかつねに把握しておくことができる。図 5 は、ユーザが発声した「スポーツ」

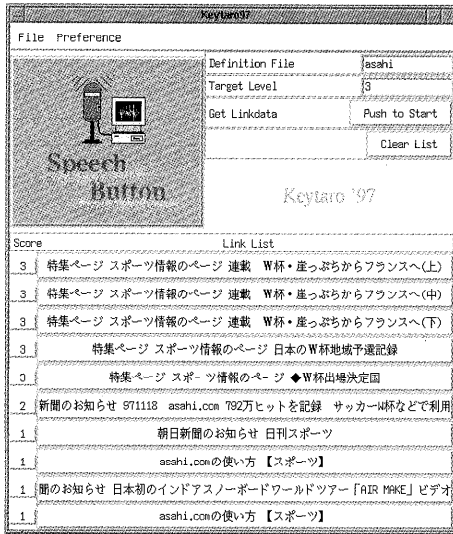


図5 インタフェースモジュールの検索結果表示画面

Fig. 5 An example of the user interface window.

「ワールド」「カップ」の3キーワードが認識された例である。音声認識モジュールで認識が行われた後、キーワードマッチング部において、上記3キーワードが含まれたリンク項目をサーチする。その結果、図5のLink Listの1番目の候補では、Scoreの表示から3キーワードとも含まれるリンク項目が存在することが分かり、現在のページから、「特集ページ」「スポーツ情報のページ」「連載 W杯・崖っぷちからフランスへ(上)」と、3階層下のページに到達できることになる。この表示部分がボタンになっているので、実際にはこのボタンを押すことにより、ページのジャンプが行われる。なお、図中の「W杯」の読みは、手作業で「ワールド」と「カップ」に修正されている。

### 3.4.2 WWW ブラウザ

最終的に選ばれたリンク先はWWWブラウザに表示される。本システムはブラウザとしてNetscape Navigatorを用いている。ブラウザ本体には何も手を加えていないので、ユーザは通常のマウスやキーボードを使ってブラウザを操作することも可能である。キーワードマッチング部において、ユーザのボタン操作に対応してブラウザの表示を切り替えなければならないが、この動作はNetscape Navigatorの-remoteオプションによって実現されている。

## 4. 評価実験

### 4.1 実験内容

本システムにおけるページ先読みのキーワード音声認識の有効性を検証するために、被験者に実際にシス

1. 1月27日の天声人語を読みたい
2. 新聞の関西面を読みたい
3. 長野五輪のジャンプの選手を知りたい
4. 全豪テニスの結果を知りたい
5. スキー場(びわ湖バレー)の情報を知りたい

図6 タスク内容

Fig. 6 Target tasks.

テムを使ってもらい評価実験を行う。この実験では、従来のマウス、キーボードなどの入力手段に音声を加え、さらにページの先読み、キーワードの利用をすることにより、WWWのブラウジングがどう変化するかを調べることが目的である。

朝日新聞社 (<http://www.asahi.com/>) のページをある程度ダウンロードして、ファイルに保存しておく。これは、ネットワークの負荷を避けるためと、すべての被験者に平等な条件で実験を行ってもらうためである。今回の実験では、ページを移動するたびに認識辞書を作成するのではなく、キーワード抽出および認識辞書の作成はあらかじめ行っている。この状態で、こちらから図6に示すような5つのタスクを与え、

- (a) 通常のマウスクリックでの操作
- (b) キーワード発声とマウスを併用した操作
- (c) キーボードを用いた操作

によりタスクに書かれたページを探すよう指示する。そして、被験者の操作の様子およびコンピュータの画面をビデオカメラで撮影し、タスク達成までの実行時間と実行手数を計測する。ここで、実行手数とはマウスのクリック数である。(b)においては、形態素解析の結果がタスクの実行にどのような影響を及ぼすかを調べるために、

- (b-1) 手作業により解析, 作成された辞書
- (b-2) 茶筌により解析, 作成された辞書

の2種類の認識辞書を使って実験を行う。茶筌により形態素解析を行い、読みを与え、音素列に変換して音声認識辞書を作成した場合、形態素解析の誤りにより、読み誤り、自立語の区分されすぎなどの状況が生じ、そのまま音声認識辞書に登録され、発声が正しくても認識できない場合が生じる。これらに起因する性能劣化を評価するために、手作業により作成した誤りのない辞書との比較を行う。(b-1)と(b-2)の実験については、始めは必ずキーワードの発声で行い、あとは音声を使ってもマウスを使っても自由という指示をした。なお、(b-2)の辞書については、音声認識の際の無駄な湧き出し誤りを避けるために、茶筌の解析の後に、Perlによるフィルタを通し、音素数が2以下の単語、数詞を辞書から省く処理、名詞の後に続く接尾

表3 実験条件  
Table 3 Experiment conditions.

被験者	本学学生 10 名
対象ページ	朝日新聞
ページ数	346
リンク数	546
認識辞書	手作業で辞書を作成 (語彙数 773)
	茶釜の結果に自動で後処理 (語彙数 702)

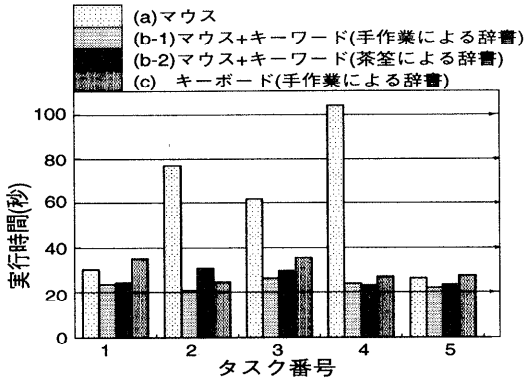


図7 タスク達成までの実行時間  
Fig. 7 Elapsed time for task completion.

辞はマージして複合名詞とする処理を行っている。また (c) においては、音声入力のみ有効性を評価するために、キーワードをキーボードから入力するモードを作成した。キーワード入力には kinput2 を用い、ローマ字で平仮名入力された1つあるいは複数のキーワードを手作業で作成した辞書と照合することにより検索を行う。実験条件を表3に示す。

4.2 実験結果

4.2.1 タスク達成までの実行時間

タスク達成までの実行時間を図7に示す。各値は、タスクごとに被験者数で平均をとった値である。今回の実験では、ユーザの操作時間に注目して実験を行っているため、図7中の (b-1) と (b-2) の時間には音声認識時間 (発声終了後から認識結果出力までの時間) は含まれていない。

図7において、(a) マウスのみと (b-1) マウス+キーワード (手作業により解析した辞書) を比べてみると、タスク1, 5においてはマウスのみの場合と音声キーワードを併用した場合とはあまり差がみられない。タスク2, 3, 4においては音声キーワードを併用した方が実行時間が短くなっている。これは、タスク1, 5ではトップページから目的のページまでが比較的分かりやすい構造になっていたからだと思う。

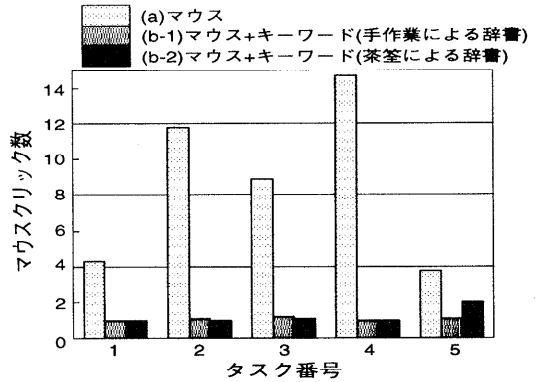


図8 タスク達成までのマウスクリック数  
Fig. 8 Total clicks for task completion.

たとえば、タスク5では、トップのページに「スキー場情報」といったリンク項目があり、マウスでたどっていきやすい。逆にタスク2, 3, 4では目的のページまでの構造が複雑であったり、見えにくい場所にリンクが存在したためにマウスの操作に時間がかかり、音声キーワードの有利さが現れた結果となった。タスク4などでは、一度スポーツ面を探してから、さらにテニスの記事を探す必要があるといった場合である。

参考までに、タスク4において、タスク達成までの実行時間の標準偏差を示すと、(a) 51.9秒、(b-1) 11.4秒、(b-2) 5.4秒、(c) 6.9秒であった。この理由を考えると、まず被験者の絶対数の不足が考えられるが、もう1つの大きな理由として、システムに対する個人差によるばらつきが考えられる。この差は、被験者個人のマウス、キーボード、音声入力に対する習熟度に起因している。ある人にとって便利なインターフェースが他の人にとっても便利とは限らないので、個人差に適応できるようなシステムを検討していく必要がある。

4.2.2 タスク実行までの実行手数

タスク達成までのマウスのクリック数を図8に示す。実行時間と同様、各値はタスクごとに被験者数で平均をとった値である。この図により、(b-1) と (b-2) では、各タスクともマウスのクリック数がほぼ1回 (リンク先を選択するボタンを押す操作) におさえられ、冗長なマウスクリックが省かれていることが分かる。

4.2.3 形態素解析の性能による影響

タスク達成時間

手作業で作成された辞書と茶釜により作成された辞書では、当然前者の辞書のほうが正確であるのでタスク達成までの時間に差が出てくるものと思われる。しかし、図7において (b-1) と (b-2) の時間を比べると、ほとんど差がないという結果が得られた。

表4 発声キーワードが認識辞書に含まれる割合  
Table 4 Percentage of spoken keywords in a keyword dictionary.

手作業で解析を行った辞書	22語中17語 (77%)
茶釜で解析を行った辞書	23語中14語 (63%)

表5 キーワード認識率および1発話あたりの平均発話キーワード数

Table 5 Keyword recognition rates and average number of keywords per an utterance.

	認識率	1発話あたりの キーワード数
手作業で解析を行った辞書 (語彙数773)	73.2%	2.8語
茶釜で解析を行った辞書 (語彙数702)	66.2%	2.6語

#### 発声キーワードと認識辞書の関係

被験者が発声したキーワードがどの程度認識辞書に含まれているか調べた結果、表4のようになった。これより、手作業で解析を行った辞書の方がユーザの連想するキーワードをより含んでいることが分かる。

ここで問題となるのは、ユーザが発声したキーワードが辞書に入っているかどうかをユーザが知ることができないということである。ユーザは自分が発声したキーワードが未知語と分かれば入力キーワードを変えることができ、システムの利便性が増すものと思われる。

#### キーワード認識率

2種類の辞書を使ったときのそれぞれのキーワード認識率およびユーザの1発話あたりの平均発話キーワード数を表5に示す。キーワード認識率の計算は以下の式によって行っている。

$$\text{認識率} = \frac{\text{正解キーワード数}}{\text{のべ発声キーワード数} - \text{未知語数}}$$

この式によって分かるように、認識率の計算の際には、辞書に入っていない単語、すなわち未知語はカウントしていない。表より、茶釜で解析を行った辞書を用いたほうが認識率が悪い結果となった。これは、茶釜の解析した辞書のほうが単語の切り方が短いため、ユーザの発話したキーワードをより短い単語の連続としてまちがうケースが見られたためであると考えられる。また、キーワード認識率と1発話あたりの平均発話キーワード数の両方の結果から、ユーザの発話したキーワードのうち約2語は正解となるのでキーワードマッチで上位の結果となり、目的のリンク先にたどりつけることが分かる。

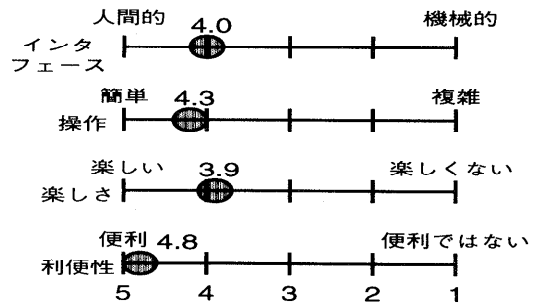


図9 アンケート  
Fig. 9 Questionnaires.

#### 4.2.4 主観評価

実験終了後、音声キーワードを使うことによって、マウスの場合に比べてシステムがどのように変わったと感じたか、被験者に5段階評価のアンケートに答えてもらった。質問項目は、ユーザインタフェースの親和性を調べるために「インタフェースが人間的か機械的か」、利便性の改善度を調べるために「操作が簡単になったか」、「便利になったか」、また、利用者が持続的にシステムを利用するかどうかにおいて重要と思われる「楽しかったか」の4種類である。結果を図9に示す。最も良い値が5、最も悪い値が1である。まず、インタフェースが人間的か機械的に関しては、マウスだけの単調な操作よりも、音声入力という入力手段が加わることによってよりユーザフレンドリに感じられたと考えられる。操作の複雑度では、マウスを何度もクリックしてページを行き来するよりも、音声キーワードを使ったほうが操作が簡単だという結果が得られ、音声キーワードの有効性が示されている。楽しいか楽しくないかについては、マウスだけよりも音声を併用したほうが楽しいという結果になった。最後に利便性については4.8という結果が得られ、音声キーワードを併用したほうがより便利という感想を得ることができ、本システムの有効性を確認できた。

#### 4.2.5 キーボード入力との比較

図7に同じ被験者についてキーワードの入力をキーボードから行った場合の結果を示す。タスクは先と同じ5つである。キーボード入力の場合はローマ字入力のためもあり、マウスと音声キーワードを併用した場合とほぼ同等かやや時間が多めにかかるという結果が得られた。今回の被験者がキーボード入力に日頃慣れ親しんでいる学生であることを考えると、キーボードに不慣れな初心者の場合にはさらに大きな差が生じることが予想される。



## 5. ま と め

本論文では、音声キーワードを用いて、効率的なネットサーフィンを実現するシステムを提案し、評価実験を行うことによりシステムの有効性を検証した。

今回の実験では、試作したシステムにおいて、ネットワークの負荷、音声認識時間を考えないといった条件で、純粋にユーザの操作だけに注目した評価を行った。この評価では、ページの構造が複雑な場合に、キーワード入力による操作のほうがマウスのみ比べて効率が良いという結果が得られ、マルチモーダルインタフェースにおける音声の有効性を示すことができた。

しかし、本システムには以下のような課題が残されている。

- インタフェースの改良
- 未知語への対処
- 適切なキーワードの選択
- 実働システムの構築および評価

実験後の被験者の感想として、リンク候補の表示において、認識されたキーワードが表示されたリンク候補のどの部分に相当しているのかわかりにくいのでボタンの選択に苦労したといった意見が出たので、ユーザがより直観的に操作できるようにインタフェースを改良していく必要がある。

次に、ユーザは自分が発声したキーワードが辞書に入っているかどうかを知ることができないので、ガーベジモデルなどの未知語対処の手法やシソーラス辞書を用いてユーザの発話を柔軟に吸収していく必要がある。先読みの深さについても、実際の利用形態の調査を通して最適化していく必要がある。

また、本研究では、音声キーワード候補としてリンク項目の名詞を使用したのが、これがページ内の情報を的確に表しているとは限らない。実験において、被験者の発話したキーワードでリンク項目内にはなかったが本文中には存在するといった例も見られた。また、本システムでは画像リンクには対応していないので、イメージタグ中のテキストについても同様の処理を行い、キーワードを抽出することが必要である。これらをふまえて、ユーザのイメージした音声キーワードに近い語彙をシステムが動的に獲得できるようにしていくようにしなければならない。

最後に、今回の実験では音声入力の原理的な有効性を検討するために、実際の環境とは違う環境で実験を行っている。実働可能なシステムを構築するためには、ネットワークの負荷を減らすためにプロキシサーバを利用する、音声認識率の低下を防ぐために獲得する認

識対象語彙をカウントしながら先読みをコントロールする、ブラウジングを行いながら辞書を動的に切り替えるといったテクニックが必要となる。今後、これらを実装し、音声認識時間、ネットワークの負荷をふまえた実際の環境で、さらに広い層の多くの被験者に対して評価を行っていく予定である。

## 参 考 文 献

- 1) 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, 第 52 回情報処理学会全国大会論文集, 5J-1, pp.407-408 (1996).
- 2) Kondo, K. and Hemphill, C.: Surfin' the World Wide Web with Japanese, *ICASSP97*, pp.1151-1154 (1997).
- 3) 甲斐充彦, 中野宗広, 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, 情報処理学会音声言語情報処理研究会, 98-SLP-20-14, pp.81-86 (1998).
- 4) 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, 日本音響学会講演論文集, 2-Q-21, pp.179-180 (1997).
- 5) 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, 電子情報通信学会音声研究会, SP97-59, pp.13-18 (1997).
- 6) 西本卓也, 新美康永: ネットサーフィンにおける音声入力語彙とその役割, 情報処理学会音声言語処理研究会, 98-SLP-20-13, pp.75-80 (1998).
- 7) 近藤和弘, チャールズヘンツヒル: 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 D-II, Vol.81, No.2, pp.257-267 (1998).
- 8) 桂浦 誠, 中村 哲, 鹿野清宏: キーワードを用いた音声によるネットサーフィン, 日本音響学会講演論文集, 2-Q-34, pp.183-184 (1997).
- 9) 桂浦 誠, 中村 哲, 鹿野清宏: 音声キーワードによるネットサーフィンの実現, 情報処理学会音声言語情報処理研究会, 98-SLP-20-12, pp.69-74 (1998).
- 10) 桂浦 誠, 中村 哲, 鹿野清宏: 音声ネットサーフィンにおけるキーワード辞書自動抽出の評価, 日本音響学会講演論文集, 2-Q-22, pp.163-164 (1998).
- 11) 松本裕治, 北内 啓, 山下達雄, 今一 修, 今村友明: 日本語形態素解析システム『茶釜』version 1.0 使用説明書, <http://cactus.aist-nara.ac.jp/lab/nlt/chasen.html>, NAIST Technical Report, NAIST-IS-TR97007 (1997).
- 12) 庄境 誠: 加法的雑音および情報性歪みが存在する自動車環境のロバストな音声認識に関する研究, Doctoral Thesis, NAIST-IS-DT9561207 (1998).
- 13) 庄境 誠, 中村 哲, 鹿野清宏: 音声強調手法

E-CMN/CSSの自動車環境内での音声認識における評価, 電子情報通信学会 (D-II), Vol.81, No.1, pp.1-9 (1998).

- 14) 庄境 誠, 中村 哲, 鹿野清宏: 音声認識における音声強調手法及びモデル適応化手法の検討, 電子情報通信学会音声研究会, SP96-19, pp.53-60 (1996).

(平成 10 年 6 月 1 日受付)

(平成 10 年 12 月 7 日採録)



桂浦 誠

昭和 47 年生. 平成 8 年大阪大学工学部電子工学科卒業. 平成 10 年奈良先端科学技術大学院大学情報科学研究科情報処理学専攻博士前期課程修了. 現在, ヤマハ (株) 勤務. 在学中, 音声認識, 音声対話システムの研究に従事. 音響学会会員.



中村 哲 (正会員)

昭和 33 年生. 昭和 56 年京都工芸繊維大学工芸学部電子工学科卒業. 昭和 56~平成 6 年シャープ (株) 中央研究所および情報技術研究所に勤務. 昭和 61~平成元年 ATR 自動翻訳電話研究所に出向. 平成 6 年 4 月より奈良先端科学技術大学院大学情報科学研究科助教授. 平成 8 年 3~8 月 Rutgers University · CAIP Center 客員教授. 音声情報処理, 主として音声認識の研究に従事. 京都大学博士 (工学). 平成 4 年日本音響学会粟屋学術奨励賞受賞. IEEE, 電子情報通信学会, 日本音響学会, 人工知能学会各会員.



鹿野 清宏 (正会員)

昭和 22 年生. 昭和 45 年名古屋大学工学部電気学科卒業. 昭和 47 年同大学院工学研究科修士課程修了. 同年電電公社武蔵野電気通信研究所入所. 昭和 59~61 年カーネギーメロン大学客員研究員. 昭和 61~平成 2 年 ATR 自動翻訳電話研究所音声情報処理研究室長. 平成 4 年 NTT ヒューマンインタフェース研究所主席研究員. 平成 6 年 4 月より奈良先端科学技術大学院大学情報科学研究科教授. 音情報処理学講座を担当, 主として音声・音情報処理の研究および研究指導に従事. 工学博士. 昭和 50 年電子情報通信学会米沢賞, 平成 3 年 IEEE SP 1990 Senior Award, 平成 6 年日本音響学会技術開発賞受賞. IEEE, 音響学会各会員.