ネットサーフィン支援のための音声対話システム

Spoken Dialog System for Net-Surfing

西本 卓也　　新美 康永
Takuya NISHIMOTO　　Yasuhisa NIIMI
京都工芸繊維大学 工芸学部 電子情報工学科
Department of Electronics and Information Science, Kyoto Institute of Technology

Abstract: To successfully retrieve information via net-surfing, the users must perceive a sense of control over their interaction. At present, however, there are many difficulties in the net-surfing environment, because some operations are time-consuming and not transparent to the users. To achieve a balance between the users' skills and the challenges of the interaction, we designed a net-surfing support system, which works in the three modes; direct manipulation mode, user directed dialog mode, and system directed dialog mode. We also investigated on the speech input for the system, and created an experimental voice operated web-browser.

1　はじめに

ウェブ環境は操作が容易であるため、仕事や趣味において多くのユーザに利用されるようになっている。しかし、ウェブ環境において、有用な情報を快適かつ効率よく手に入れるためには、さらに多くの改善が必要である。

我々はネットサーフィンをタスク達成のための手段と考え、ウェブ環境が持つ問題点をユーザの視点から整理した。また、ユーザの操作を対話的に支援することでこれらの問題を解決するための、システム動作モードについて整理した。さらに、これらのモードに応じた音声対話を行うウェブブラウザ VOXplorer を作成した。

2　ネットサーフィン支援の必要性

Hoffman らが提案するネットサーフィンの認知モデル [1] によれば、ユーザのスキルとコンテンツのチャレンジが均衡したときに flow とよばれる状態が生じ、このときユーザはウェブ環境に対する能動的な参加感覚を得ることができる。

また、flow には目標指向と体験指向の 2 種類の均衡があり、両者は大きく異なるとされている。目標指向 flow においては常に選択的な操作が行なわれる。また、体験指向 flow は時間潰し・娯楽的な要素が強い。

一般にユーザはネットサーフィンによって何らかのタスクを達成しようと試みる。しかし、ブラウザのユーザインタフェースやウェブページの構成などによっては、操作に困難を感じてしまい、タスクを達成できない。また、対象となるウェブ情報の量や質がユーザの要求を満たさない場合や、ユーザが対象への興味を失ってしまった場合なども同様である。

ネットサーフィンにおいて、情報検索タスクの達成を妨げる要因には、次のようなものが挙げられる。

操作労力に関する困難　例えば、目的のページへ移動するための操作量が多かったり、操作が単純でない場合に、ユーザは困難を感じる。また、ブラウザの動作モードの変更をしたい場合などにも、ユーザは複雑な操作を強いられる。

システムの透過性に関する困難　例えば、ネットワーク性能の制約により処理を待たされる場合などにおいて、ブラウザの内部処理の状態はユーザに理解されにくく、これがユーザの不安を引き起こす。

西本卓也, 〒606-8216 京都市左京区松ヶ崎御所海道町, 京都工芸繊維大学 電子情報工学科, Tel:075-724-7492, Fax:075-724-7400, nishi@dj.kit.ac.jp, http://www-vox.dj.kit.ac.jp

3　システムの設計

ブラウザ操作に不慣れなユーザや、検索対象の話題を熟知していないユーザに対して、ユーザのスキルを補助し、コンテンツのチャレンジを下げるような支援システムを提供することにより、ユーザの操作効率や満足度を高めることができる。

このようなシステムがどの程度のユーザ支援を行うべきかは、ユーザやタスクによって異なると考えられる。そこで、既存のブラウザに単語音声認識およびテキスト音声合成を併用する場合のシステム動作モードを以下のように分類した。

1. 直接操作モード:　従来のブラウザ操作に加えて、マウス操作に対応する音声コマンド入力のみを付加する。
2. ユーザ主導対話モード:　モード 1 の機能に加えて、ユーザが要求した場合にのみ、以前のページへの直接移動や、リンク先ページの先読みなど、操作労力を軽減する機能を提供する。
3. システム主導対話モード:　モード 1, 2 の機能に加えて、ユーザの操作労力を軽減したり、ユーザの行うべき操作をシステムが積極的に提示する。

これらのモードに応じて、システムはユーザにメッセージを提示すると同時に、各モードで許されたコマンド入力を受理できるものとする。

また、特に音声コマンドの設計においては、マウスによる既存のブラウザ操作体系はそのまま併用できるようにする。また、言語処理や音声処理の誤りに対する頑堅性を高めるために以下のような配慮を行う。

- 音声入力により他のページに移動する場合は、ユーザの選択内容を音声で復唱し、「はい」「そうです」などの確認を要求する。
- ページ内の文字列が複数の読みを持つ場合には、これらをともに音声コマンド候補とする。また、文字列の一部分（例えば「芸術と人文」における「人文」など）も音声コマンドとする。
- 音声コマンド候補の最大語彙数を制限し、音声認識部の応答速度を維持する。

4　システムの実装

作成したシステム VOXplorer (Voice Operated Explorer) はクライアントとサーバから構成される。クライアントは Microsoft Windows 95 および Internet Explorer 4.0 上で動作する。音声認識・合成には Catena ドキュメントトーカ Ver.3 を使用している。OCX を用いることにより音声入出力とブラウザとの連携を行う。

VOXplorer サーバは UNIX WS 上の Apache httpd とその CGI として実装されている。クライアント側のブラウザから要求を受けると、VOXplorer サーバは表示する HTML ファイルを実際の WWW サーバからダウンロードし、下記のような変換を行ったうえで、結果をブラウザ側に送る。

1. HTML ファイルが含む画像ファイルや他ページへのリンク等の URL を絶対パスに変換する。
2. ページが複数のフレームから構成される場合は、個々のフレームを table を用いて結合し、擬似フレーム画面を生成する。これは OCX と script の埋め込み処理を容易にするためである。
3. 形態素解析システム JUMAN を用いてページ内の文字列から単語境界と読み情報を得る。
4. HTML タグ情報と対応する単語列から音声入力語彙リストを生成し、適切な語彙数を選択する。
5. HTML ファイルに対し、音声入力が行われた場合の処理（確認やページ移動など）を行う script を埋め込む。マウスクリックによるリンク操作は再度 CGI を介するようにする。

5　まとめ・今後の課題

ネットサーフィンにおけるユーザの操作を対話的に支援するために、複数のシステム動作モードを選択的に使用することを提案し、音声ウェブブラウザ VOXplorer を作成した。

今後は、システム動作モードやタスクの特性（目的指向／体験指向）が異なる場合など、各種の条件において音声利用の効果を評価する予定である。さらに、システム動作モードのあり方や、モードの適切な選択方法についても検討し、システムの改良を行っていく予定である。

参考文献

[1] Donna L. Hoffman, Thomas P. Novak, “Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations”, http: //www2000. ogsm. vanderbilt. edu, 1995.

[2] 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, 日本音響学会講演論文集, 2-Q-23, pp165-166, 1998-03.