音声ウェブブラウザ VOXplorer の評価

◎西本 卓也　藤沢 正樹　新美 康永　（京工繊大・工）

1.　はじめに

音声によるネットサーフィン支援機能は、既存のウェブブラウザの操作体系を維持しつつ、情報検索支援のための対話的インタフェースを容易に組み込むことができる。また、今後はウェブブラウザ機能を備えた車載機器や家電・携帯情報端末などが普及することが予想されており、ポインティングデバイスによらないブラウザ操作は、このような環境における有用な選択肢の一つとなる。

我々は音声を用いたネットサーフィン支援システム VOXplorer (Voice Operated Explorer) を構築してきた。本システムにおけるネットサーフィン支援は、ユーザがシステムに操作を命令できる機能（直接操作）および、ユーザが求める情報に関してシステムが対話的に助言を与える機能（対話的支援）に分類できる。しかし、対話的支援がユーザに支持されるためには、直接操作において、音声の利用にユーザが違和感を持たないことが前提となる。

本稿では特に音声ウェブブラウザの直接操作に注目し、実環境におけるホームページ検索タスクを用いて評価を行った。また、ユーザの音声入力を確認する方法について、確実さを重視する構成と効率を重視する構成の2通りに関する評価を行った。

2.　VOXplorer の概要

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 マシン上の Apache httpd と CGI として実装されている。クライアント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、ページ内容に応じた音声入力語彙を決定し、必要なスクリプトを HTML に埋め込む処理を行ったうえで、結果をブラウザ側に送る。

3.　タスク

本実験では評価用タスクとして、Yahoo Japan (http://www.yahoo.co.jp) のサイト解説文を読んでもらいながら、特定の情報についてのページを検索させた。ただし、文字列による検索は使用せずリンク項目のみを利用させた。

被験者は課題を提示されたのち、Yahoo Japan トップページからのリンク選択によって該当するページを検索した。

課題は「投資信託の選び方がわかるページ」「ウルフルズの公式ページ」「改造車の車検」といった表現で提示した。これらは Yahoo による解説文が含んでいる表記を多少言い替えて引用したものである。

4.　確認応答モデル

VOXplorer を運用するにあたって、下記の2つのモードを設定した。

モード1（確認応答あり）は、操作の確実さを重視する構成である。ユーザがリンク項目を読み上げるたとき、システムがその項目を復唱し、ユーザによる確認を待つ。ユーザが「はい」または「そうです」と発話することによって、システムは「了解」と発話するとともにページ移動を開始する。

モード2（確認応答なし）は、操作効率を重視する構成である。ユーザがリンク項目を読み上げたとき、その項目を復唱した後、ユーザの確認を求めないでページ移動を開始する。誤認識等で間違ったページ移動が行われた場合は「戻る」という操作でページ移動を取り消す（図1）。

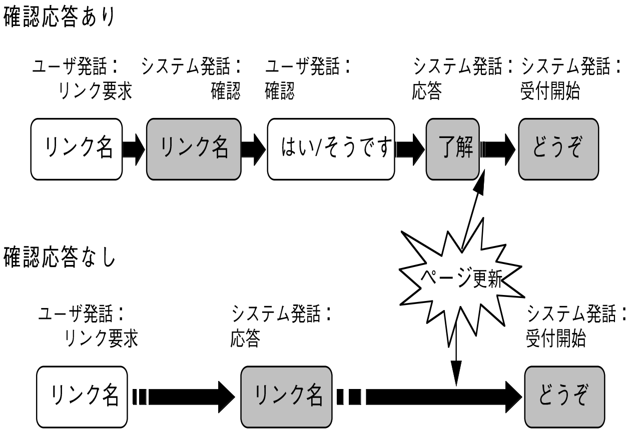

図１．２種類の確認応答モデル

∗ Evaluation of VOXplorer Speech Input Web Browser

By Takuya NISHIMOTO, Masaki FUJISAWA and Yasuhisa NIIMI
(Department of Electronics and Information Science, Kyoto Institute of Technology)

5. 実験

前述の 2 つのモードにおいて評価実験を行った。被験者は日常的にウェブブラウザを使用している大学院生 5 名である。行ったタスク数を表 1 に示す。

表１．実験で行ったタスク数

被験者	1	2	3	4	5	合計
モード 1	4	3	3	3	3	16
モード 2	4	3	3	3	3	16

1 回のタスクに要した時間は約 5 ~ 10 分であった。また、VOXplorer システムが画面更新に要した時間は約 8 秒である。これは、音声入力を受理してから、画面を更新し、全てのインライン画像を表示し、システムが音声プロンプト（「どうぞ」）を発声するまでの時間である。

認識率，ページ移動効率およびタスク達成効率を次式によって求めた。ただし発話数は選択操作に関する発話のみを含め、スクロールなどのページ内操作に関するものを除く。

$$\text{認識率} = \frac{\text{認識発話数}}{\text{認識発話数} + \text{誤認識発話数} + \text{棄却発話数}}$$

$$\text{ページ移動効率} = \frac{\text{発話数}}{\text{ページ移動数}}$$

$$\text{タスク達成効率} = \frac{\text{発話数}}{\text{ページ移動数} - \text{誤認識による移動数}}$$

各モードにおける結果を図 2 に示す。

また、被験者 5 人のアンケート結果の平均点を表 2 に示す。ただし、評価は 1 ~ 5(5 段階) で最高点は 5 である。

表２．被験者のアンケート結果（平均）

	モード 1	モード 2
集中できるか	3.2	2.6
効率がよかったか	1.8	1.2
楽しかったか	3.2	2.2
わかりやすいか	2.8	2.8
簡単か	2.4	2.8

6. 考察

どちらのモードにおいても認識率が高いほどタスク達成効率は高い（図 2）。またモード 2 はモード 1 と比較すると少ない発話数でタスクが達成できている。しかし、ユーザの主観的な評価においては「集中できるか」「効率がよかったか」「楽しかったか」の 3 項目について、モード 1（確認応答あり）の方が高い評価を得た（表 2）。誤認識によりページ移動が行われた場合、誤ったページ移動および元のページへの復帰のための時間が必要であるが、この間ユーザの割り込み操作が不可能であることが、ユーザの指示を得られなかった理由と考えられる。

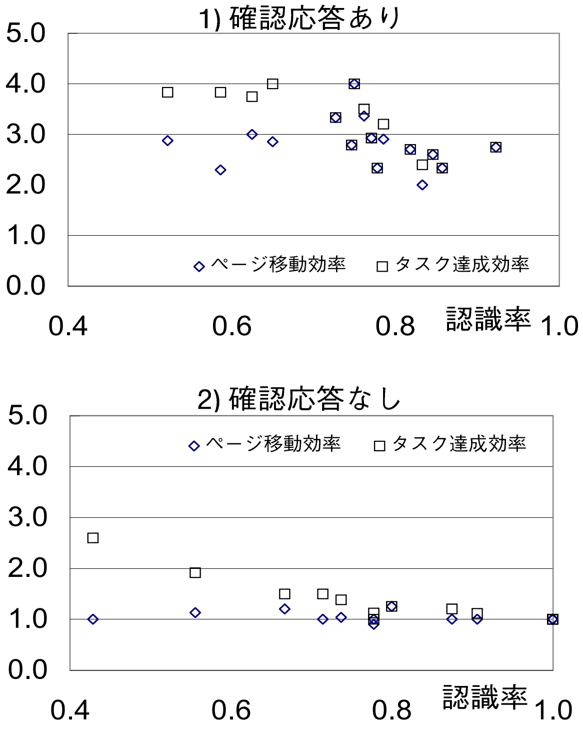

図２．各実験における認識率，ページ移動効率およびタスク達成効率

既存のブラウザ操作との比較においては、アンケート項目「集中できるか」「楽しかったか」の 2 項目については同等または良好と答えた被験者が多かった。このことから、音声ウェブブラウザは没入感や娯楽性といった点において幅広いユーザに支持されうるものと考えられる。

また、誤認識がほとんどなかった被験者は「わかりやすいか」「簡単か」の項目も既存のブラウザ操作と同等であると評価した。

7. まとめ

音声ウェブブラウザ VOXplorer を用いた評価実験の結果、ユーザは実際の操作効率とは逆に、確認応答を要求するシステムに対して効率の良さを感じることが分かった。また、音声ウェブブラウザがマウス操作と同等のわかりやすさや簡単さなどを得るためには誤認識の防止が重要であること、音声操作が没入感や娯楽性などの点で魅力的であることなどが明らかになった。

今後はこれらの結果を踏まえてシステムを改良するとともに、さまざまな熟練度の被験者を対象とした評価を行う予定である。

参考文献

[1] 西本卓也, 新美康永, ネットサーフィン支援のための音声対話システム, 第 12 回人工知能学会全国大会, S7-03, pp.141-142, 1998-06.