

音声操作可能な Web ブラウザの開発を通じた 音声操作インタフェースの模索

伊東 一樹[†] 平川 昌宏[†] 笥 捷彦[‡]

早稲田大学大学院基幹理工学研究科[†] 早稲田大学理工学術院[‡]

1. 背景

現在，音声認識技術はロボット，カーナビ，医療システム，携帯電話など，様々な分野で利用されている。

その音声認識技術を PC 操作に利用し，有効に扱うためにはどのような形で用いれば良いかを探ることを目的として研究を行った。今回は，PC 操作の中でも高頻度で行われる Web 閲覧という分野に焦点を絞って研究を行った。

具体的には，Web ブラウザに音声操作を行うための機能を追加していき，それらの機能を評価・改良を行うことで，PC 操作に適した音声操作インタフェースの模索を行っていく。

2. 開発方針

本研究では，音声操作機能を既存の Web ブラウザ (IE や Firefox など) に組み込むのではなく，初めから Web ブラウザの開発を行い，それに音声操作機能を組み込む形で研究を行った。これは，音声操作機能を追加しやすくするためである。

また，音声認識機能はフリーで提供されている大語彙連続音声認識システム Julius [1] を用いて実装を行った。Julius は音声認識結果を文字列として出力するので，本システムではその文字列によって Web ブラウザに動作を行わせる。

音声操作機能を実装するにあたり，音声操作の短所として，次の 3 点を挙げる。

- ・位置のピンポイントな指定が行えない。
- ・100%の精度での認識はまず不可能
- ・1回の入力に1秒程度はかかってしまう。

よって，本システムではこれらの短所に対処するという方針で開発を行った。

The Evaluation of the Interface of Speech Recognition through Development of the Web Browser for Speech Recognition

[†] Kazuki ITO, Masahiro HIRAKAWA, Graduate School of Fundamental Science and Engineering, Waseda University

[‡] Katsuhiko KAKEHI, Faculty of Science and Engineering, Waseda University.

3. システム動作概要

本システムにおいて音声操作は，次に示す手順で行う。

- ① 現在表示している Web ページを解析する
- ② 入力された音声に対して音声認識を実行し，結果を文字列として取得する
- ③ 認識結果に従って動作を実行

まず，リンクの URL やテキストボックスの ID などを取得する。同時に常に Julius をスレッドとして実行させておき，音声入力が行われる度，音声認識結果を文字列として取得する。この文字列を参照してブラウザに動作を行わせることで，音声入力機能を実現している。

4. 音声操作機能

本研究では以下の 4 つの機能の実装・評価によって音声操作インタフェースの模索を行った。

- ① ブラウザの基本動作
⇒特定の動作を行う方法の模索
- ② ページ中のリンク選択
⇒画面上に表示されている要素の選択方法の模索
- ③ 文字入力
⇒テキストの入力方法の模索
- ④ 音声入力に適した検索
⇒音声入力で選択しやすい表示方法の模索

4.1. ブラウザの基本動作

ブラウザの基本動作とは，戻る，スクロール，ブックマークなどの動作を示す。あらかじめ動作と音声を登録しておくことで，その音声を入力することで動作を実行することができる。これによって，1回の音声入力だけで複雑な動作を行うことを現した。

また，音声の登録を行う際，登録する単語の認識のしやすさを数値で設定することができる。これによって認識の精度を向上

させることができる。動作の登録は、図1に示すウィンドウを使用して行う。



図1 動作の登録を行うウィンドウ

4.2. ページ中のリンク選択

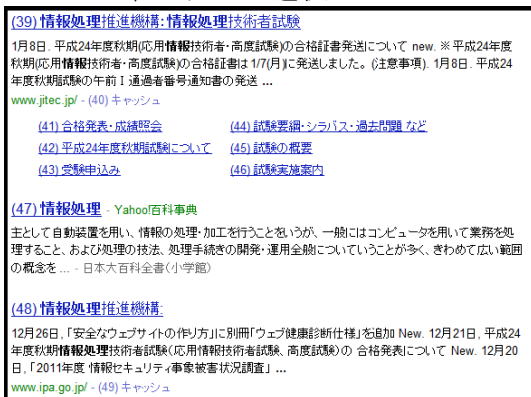


図2 ページ中のリンクへの番号の割り振り

Web ページ中のリンクやテキストボックスに図2のように番号を割り振る。

リンクを選択し、そのリンクに移動したい時には、割り振られている数字を読み上げ、その後「OK」と発声することで移動を行う。

4.3. 文字入力

ページ中のテキストボックスへの文字の入力は、必要最小限の入力で済むように、次の手順で行う。

- I. Web ページ中のテキストボックスを選択
- II. 入力したい文章を直接読み上げて入力
- III. サブミットなどを実行

4.3.1. テキストボックスの選択方法

リンクと同じように、ページ中に存在するテキストボックスにアルファベットを割り振る。

割り振られているアルファベットを読み

上げることで、テキストボックスの選択を行い、テキスト入力モードに遷移する。

4.3.2. 文章の入力

認識した音声はかな漢字変換を行い、すぐに入力する。この時、変換は Julius の辞書の中で一番コストの高い単語を入力する。

また、パスワードなどの音声入力を行いづらい文章は、動作登録と同様にあらかじめ音声と合わせて登録しておくことで、1単語の入力によって文章の入力を行えるようにした。

4.3.3. 文章の修正

最後の1文字を消す、最後の単語を消すなど、様々な範囲の修正方法を用意し、それらを1単語の発声によって実行できるようにすることで、認識ミス時に素早く対応をできるようにした。

また、入力した単語がうまく変換されなかった場合のため、様々な修正範囲の選択方法を用意した。ユーザは、それを発声することで、変換ミスがあった単語をすぐに選択する。システムは、その変換候補を番号と共に提示し、ユーザはその番号を発声することで、正しい変換結果を選択する。

4.4. 音声入力に適した検索

Google や Yahoo といったポータルサイトを使わずとも Web 検索を行えるように、Web 検索の結果を表示する機能を実装した。ただ検索結果を表示するだけでなく、音声入力でもリンクを選択しやすい様に検索結果を表示するように実装した。

5. まとめ

Web ブラウザに音声操作機能を実装、評価を繰り返し行うことで、PC 操作に適した音声操作インタフェースの模索を行った。この研究の成果が、PC 上で音声操作システムを作成する助けになると考えられる。

参考文献

- [1] Julius, 名古屋工業大学 Julius 開発チーム
2012年1月11日訪問, <http://julius.sourceforge.jp/>