

音声検索システムにおけるユーザインターフェースの考察

杉崎 茜[†] 井ノ上 直己[‡] 橋本 和夫[‡] 米山 正秀[†]

[†]東洋大学
〒350-8585 埼玉県川越市鯨井2100
E-mail: akane@yo.cs.toyo.ac.jp
yoneyama@cs.toyo.ac.jp

[‡]KDD 研究所
〒356-8502 埼玉県上福岡市大原2-1-15
E-mail: {inoue,kh}@kbip.kddlabs.co.jp

あらまし 筆者らは、高齢者や障害者のようなコンピュータ操作が不慣れな人だけでなく、一般の人にも容易に扱える優しいユーザインターフェースの開発を目的に、音声によりコンピュータと対話しながら検索を行えるシステムの開発を行っている。このシステムにおける音声認識誤りの訂正方式として、筆者らは既に認識誤り時にだけ同じ音声を繰り返し入力する方式を提案した。本稿では、コンピュータ操作が不慣れな高齢者に、実際に提案手法を実装した情報検索システムを使用してもらい、我々の提案手法による入力方式を評価してもらった結果を報告する。

キーワード 音声入力 情報検索 ユーザインターフェース 繰り返し音声

Consideration of User Interface in Voice Input Information Retrieval System

Akane Sugizaki[†] Naomi Inoue[‡] Kazuo Hashimoto[‡] Masahide Yoneyama[†]

[†]Toyo University
2100, Kujirai, Kawagoe, Saitama 350-8585
E-mail: akane@yo.cs.toyo.ac.jp
yoneyama@cs.toyo.ac.jp

[‡]KDD R&D Laboratories Inc
2-1-15, Ohara, Kamihukuoka, Saitama 356-8502
E-mail: {inoue,kh}@kbip.kddlabs.co.jp

Abstract We are developing an information retrieval system with voice input man machine interface, in order to design a system which can be easily handled by not only the elderly but also ordinary person. For this point of view, we have already proposed a man machine interface where user can re-input the same speech as the previous speech only when the system misrecognized it. In this paper, we present evaluation results that the elderly use the information retrieval system in which our proposed man machine interface is implemented.

Key words Speech Input, Information Retrieval, User Interface, Repeated speech

1. まえがき

近年、高度情報化社会が進む中では、コンピュータと人間との情報交換が重要な役割を担うと考

えられる。また、インターネットの普及によって情報検索システムが一般化し、種々の情報が利用者のもとに容易に手に入る環境が整う時代となっ

てきた。しかしながら、高齢者や障害者はキーボード操作に不慣れであったり、障害のためにキーボード入力ができなかったりするため、情報検索システムを利用できない人が多い。このような状況を改善するためには、利用者にとって自然な方法でコンピュータと対話しながら情報検索が行える方式を開発する必要がある。特に、入力手段では、人間の持つ最も基本的で便利なメディアである音声の利用が期待されている。

そこで筆者らは、高齢者や障害者のようなコンピュータ操作が不慣れな人だけでなく、一般の人にも容易に情報検索システムが扱える優しいユーザインタフェースの開発を目的に、音声入力による情報検索システムの研究開発を行っている^[1]。

利用者とシステムとの対話を円滑に進めるためには、入力された音声に対して認識を誤ったという事実を把握しながらタスクを遂行する必要がある。例えば、システムの認識語彙に登録されていない未知語が入力された場合、システムが正しく未知語を検出できずに誤った認識結果に基づいて情報検索を行うと、利用者が求めている情報が得られなくなったりする。そのため、通常は利用者による認識結果の確認・訂正が必要となる。

誤り訂正方式には種々存在するが、筆者らは、利用者の手数が最も少なく済む方式が最も優しい方式だと考え、認識誤りにだけ利用者が同じ音声を繰り返すことができる方式を提案した^[2]。この方式では、システムが、同じ音声が繰り返し入力されたことを検出すると、直前の認識結果が誤っていたと判断し、直前の認識結果を訂正することで、利用者の手数を減らしている。しかし、システムが何度も認識誤りを繰り返した場合、かえって利用者がわずらわしく感じると思われることから、この方式にも工夫が必要だと考えられる。そこで、コンピュータ操作が不慣れな高齢者に、実際に提案手法を実装した情報検索システムを使用してもらい、我々の提案手法による入力方式を評価してもらった。本稿では、この評価実験結果について報告する。

以下、第2章では筆者らが開発している音声入力による情報検索システムの概要を示し、第3章では本稿で比較対象とした認識誤り訂正方式を示す。第4章では評価実験結果について示し、最後に、第5章で今回の実験に対するまとめ及び今後の予定について述べる。

2. 音声入力による情報検索システム

筆者らは、高齢者や障害者だけでなく一般の人にも容易に情報検索システムを扱える優しいインタフェースの開発を目的に、既存の情報検索エンジンを音声入力により制御できるシステム（以下、音声検索システム）の開発を進めている。ここでは、この音声検索システムの概要を示す。

2.1 検索エンジン

音声検索システムでは、1997年3月5日までの約10年間分の朝日新聞記事約180万件を検索対象としており、これらの検索には市販の全文検索エンジンを用いている^[3]。

また、図1に示すように、利用者は「セナ」「事故」「死亡」のように検索したい新聞記事の内容を単語ごとに区切って発声し、システムは1単語認識する度に認識結果を画面表示するとともに音声合成装置を利用して認識結果を復唱する。

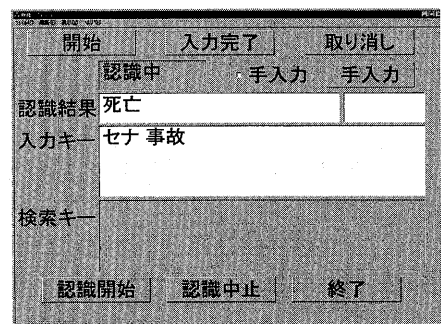


図1 音声検索システムの画面

2.2 認識誤り訂正方式

筆者らは、正しく認識できた場合は、続けて次の検索キーワードを入力でき、音声認識を誤った場合にだけ、同じ音声を繰り返し入力することで認識誤りを訂正できる誤り訂正方式を提案している^[2]。例えば、図2に示すように、入力が「事故」に対し認識結果が「実行」と誤ると利用者は「事故」と再び入力する。システムは今入力された音声が入力された音声と同じであるか否かを自動判定し、同じと判断すると直前の出力結果「実行」を「事故」と書き換える。ここで、同じ「事故」という入力でも、正しく認識された後に入力された「事故」を異語音声、認識を誤った後に入力された「事故」を同語音声と呼び区別する。

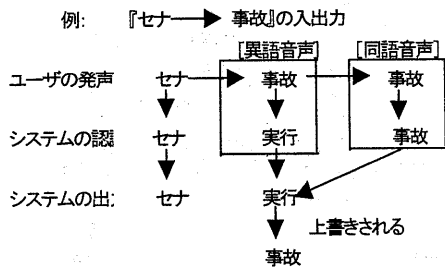


図2 誤り訂正処理の流れ

3. 認識誤り訂正方式の改良

図2には既に提案した認識誤り訂正処理を示した。しかし、この方式では、システムの認識語彙に登録されていない未知語が入力された場合など、システムが正しく未知語を検出できずに認識を誤ると利用者は何度も同じ単語を繰り返して入力することになる。これは、利用者には返ってわずらわしと感じられると思われるため、音声を繰り返し入力して認識誤りを訂正する方式にも工夫を行う必要がある。そこで、以下に示す2通りのバリエーションを提案し比較することとした。

(1) 2回認識を誤った際、候補一覧表を表示する方式

この方式は、図2に示した認識誤り訂正方式で、何度も同じ単語の入力を避けるため、同語音声が入力された時点で、認識候補一覧を表示する方式である。例えば、図3に示す例では、「事故」に対し認識結果が「実行」と誤ると、利用者は「事故」と再び入力し、それでも認識を「時効」と誤った場合には、再度「事故」と発声する。この際、画面には認識候補一覧が表示され、正解を選択することにより、直前の認識結果「時効」を「事故」と訂正できる。ここで、認識候補一覧から正解を選択する方法には、キーボードやマウスを使うことも考えられるが、特にコンピュータ操作に不慣れた人にも容易に利用できると思われるタッチパネルを利用することとした。また、画面に認識候補一覧を表示したとしても、その中に発声単語がない場合は、これ以上同じ音声の入力はしなくて済むように、システムは「入力単語は認識できませんでした。別のキーワードを入力して下さい。」とアナウンスを行い、利用者に別のキーワードの入力を促すこととした。図4に、「事故」を「時効」と誤つ

た後、再び「事故」と繰り返し入力した場合の表示画面を示す。

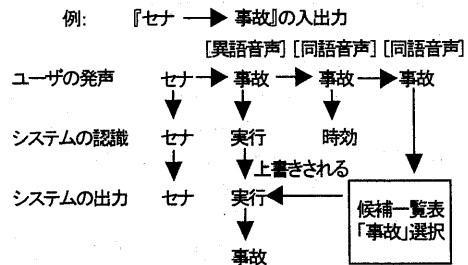


図3 連続して誤った場合に候補一覧を表示する方式

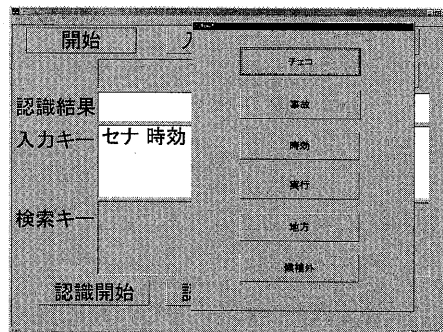


図4 認識候補一覧表示画面

この方式では、2回目の同語音声まで認識候補は1候補しか表示されないため、同語音声を入力する割合が増えることになる。しかし、繰り返し音声だけで認識誤りを訂正できるため、認識率にもよるが検索キーワード全てが入力し終わるに要する時間は短くてすむ可能性がある。

(2) 1回認識を誤った際、候補一覧を表示する方式

この方式は図5に示すように、最初の同語音声入力に認識候補の一覧を表示し、利用者に候補中から正解を選択させる方式である。この方式では、最初の同語音声が入力された時点で認識候補一覧を表示するため、先の(1)の方式と比較し、正解が画面に表示される割合は高くなる。一方、認識誤りを訂正するためには、候補一覧から選択するという操作が必要であるため、認識率にもよるが検索キーワード全てを入力するまでの時間は(1)の方式と比較し長くなるが予想される。(1)の方式と同様、最初の同語音声入力

表示した認識候補一覧中に正解が含まれていない場合は、再々度同じ発声を繰り返し認識候補一覧を表示するが、それでも正解が表示されない場合は、別のキーワード入力を促す。

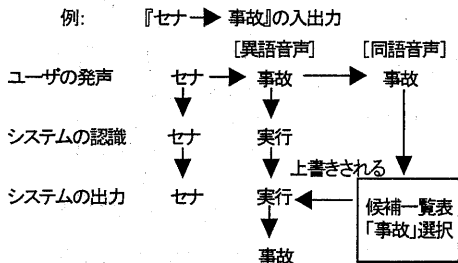


図5 最初の認識誤りで候補一覧を表示する方式

4. 評価実験

高齢者などのコンピュータ操作に不慣れな利用者にとって、前章で示した方式のどちらが有効であるか、また認識を誤った場合、繰り返し同じ音声が発声して誤りを訂正することが受け入れられるか否かを確認するため、評価実験を行った。その際、ベースラインとして、以下の従来方式も用意して比較実験を行った。

4. 1 従来方式

従来から音声認識システムにおける認識結果の確認・訂正方式である以下に示す方式も試した。

・毎回認識候補一覧を表示する方式

この方式では、利用者がキーワードを発声する度にシステムがそれを認識し、認識候補の一覧を画面表示する。例えば、「セナ」という発声に対して表示された候補一覧表の中から「セナ」を選択する。このように、候補一覧中に正解があるとそれを選択し、利用者は次の「事故」と発声する。しかし、画面表示された候補一覧表の中に正解が無かった場合には、一覧表の「候補外」を選択してから再度発声し直す。この方式では、入力度に認識候補一覧を表示して正解を選択させるため、正しく認識した場合（正解が認識候補一覧の第1位の候補になった場合）でも選択処理を必要とする。しかし、入力した単語が画面に表示されない割合（誤認識率）が低くなるため、同語音声の入力回数が減る。そのため、検索キーワード全てを入力し終わるに要する手数は認識率にも依存するが前章に示した方式よりも少なくなる可能性がある。

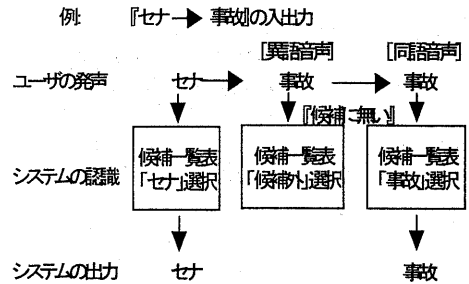


図. 6 毎回認識候補を表示する方式

4. 2 評価用データ

評価実験では、男性10名女性10名の60歳～79歳までの合計20名の話者に、実際に検索したい新聞記事を単語毎に区切り、一人10項目の発声をしてもらった。実験での被験者の分布は表1に示す通りであった。

表.1 被験者の年齢構成

| 人数 | 60～65歳 | 66～70歳 | 71～75歳 | 76～80歳 |
|----|--------|--------|--------|--------|
| 男性 | 4 | 3 | 2 | 1 |
| 女性 | 9 | 1 | 0 | 0 |

4. 3 評価方法

被験者には、予め自分の発声とシステムの表示が正しければ（認識候補一覧を画面表示した場合は正解があれば選択し）次のキーワードを発声してもらい、利用者の発声とシステムの認識結果表示が誤っていたら繰り返し同じ音声を発声してもらい（認識候補一覧を画面表示した場合は正解があれば「候補外」を選択してもらい）ように指示を行った。このようにし、先に示した3通りの方式を試してもらった。その際、いずれの方式でも、同じ検索キーワードの入力を行っている。実験後、3つの方式に対して利用者の感じたユーザインタフェースとしての便利さを5段階で主観評価してもらった。

今回の実験はユーザインタフェースの有効性の実験であるため、3度続けて認識を誤った場合のシステムからの別のキーワード入力を促すアナウンスに対しては無視し、次のキーワードを入力してもらった。

4. 4 実験結果

4. 4. 1 主観評価結果

3つの方式に対するそれぞれの主観評価結果を図7に示す。ここで、第3章で示した2つの方式はそれぞれ「2音声+候補一覧表」および「1音声+候補一覧表」に対応しており、「毎回候補一覧表表示」は第4.1節で示した方式に対応している。図7より、筆者らが第3章で提案した方式のうち、最初の同語音声で認識候補一覧を表示する「1音声+候補一覧表」に対しては、1や2という低い評価をする人がいなかったという、非常にまとまりのある結果が得られた。また、主観評価の平均値は表2に示す通りであり、「1音声+候補一覧表」が最も高い評価を得た。

4.4.2 評価結果の分析

比較評価した3通りの方式に対し、1回の検索試行で最終的に検索キーワードの全てを入力するのに要した手数時間の結果を、図8に示す。図8より、毎回認識候補を表示する方式では、他の2つの方式に比べ手数時間が約10秒長くかかることが分かる。また、「1音声+候補一覧表」、「2音声+候補一覧表」の2つの方式における手数時間には、ほとんど差が見られない。これには1回目の発声による認識率が高いことで、2回目以降に2つの方法で誤り訂正する時間には大きな差が見られないのだと思われる。また、比較評価した3通りの方式における、認識率の結果を図9に示す。ここで、図9に示した認識率とは、1つの単語に対して何回か繰り返し入力したことにより、正解が画面上に表示された割合を示している。すなわち、図3の例では、「事故」という入力が3回繰り返されて、3回目に表示された認識候補一覧中に正解が含まれた場合は「事故」という入力に対する認識は正解したとカウントしている。

図9より、毎回認識候補を表示する方法は当然認識率が高く、ついで、「1音声+候補一覧表」、「2音声+候補一覧表」となっているが、これらの間にはそれぞれ2~3%程度の差しかないことが分かる。

表2では、毎回認識候補を出力する方式が「1音声+候補一覧表」方式よりも評価が0.15ポイント低かったのは、手数時間が大幅にかかったことが原因だと推察される。一方、「1音声+候補一覧表」方式と「2音声+候補一覧表」方式とで手数時間および認識性能に目立った差が見られないにもかかわらず、主観評価の結果0.45ポイントも差があるのは、手数時間や認識性能の他に評価を左右する原因があるとことを

示している。

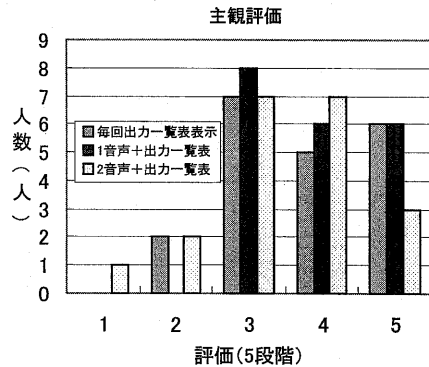


図7 主観評価結果の分布

表2 主観評価の平均値

| | 毎回出力一覧表 | 1音声+出力一覧表 | 2音声+出力一覧表 |
|------|---------|-----------|-----------|
| 評価平均 | 3.75 | 3.9 | 3.45 |

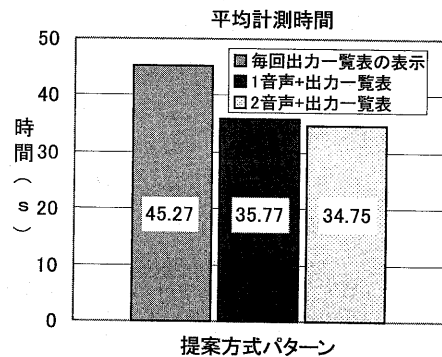


図8 手数時間

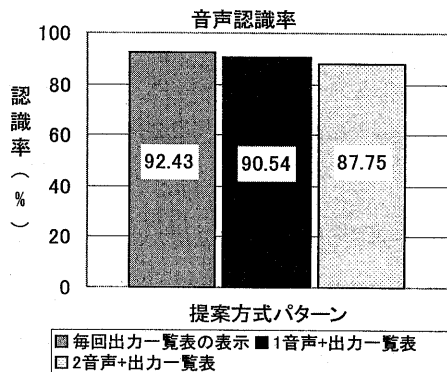


図9 認識率

この原因として、以下のことが考えられる。我々のシステムでは認識候補一覧表示時、正解候補の選択のためタッチパネル操作だけを受け付け、音声の入力は受け付けずに設計していたことから、これによって余計な音声を発してもその音をシステムが拾う心配が無いという安心感を利用者に与えたことが評価に影響していると思われる。逆に言えば、音声だけで認識誤りを訂正する方式では、システムが如何なる全ての音も拾ってしまうことから、余計な音声を発してはいけないというプレッシャーが、利用者に大きな負担をかけたと考えられ、このことが原因で、評価にかなりのばらつきが出たものと思われる。しかし、入力毎に認識候補を表示する方式よりも、最初の1回は同じ音声の繰り返しを受け付ける方式の方が評価が高かったことから、利用者の負担よりも手数時間の大きな差が評価に優先したと思われる。

5. まとめ

筆者らは、音声入力情報検索システムにおけるユーザインタフェースとして、検索キーワードの認識を誤った場合にだけ同じ検索キーワードを繰り返し発声することで、認識誤りを訂正する方式を既に提案した。この方式の改良として(1)2回認識を誤った際、候補一覧表を表示する方式、(2)1回認識を誤った際、候補一覧表を表示する方式を提案した。また、従来方式の(3)毎回出力一覧表にて認識候補を表示する方式も併せ、実際にこれらの方式を実装した音声検索システムを利用した主観評価実験を行った。その結果、(2)の方式が最も主観評価が高く、さらに5段階評価のうち3.9ポイントとかなり高い結果が得られた。このことから、筆者らの提案する認識誤り時に同じ音声を繰り返し入力する方式は、コンピュータ操作に不慣れた利用者にとって、有効なインタフェースであることが分かる。また、(3)の方式では、認識率は一番良い結果が得られたものの、計測時間がかなり過ぎることで、(2)の方式の主観評価に比べ低い結果となったと考えられる。一方、(1)の2回認識を誤った際に候補一覧表を表示する方式が他の2つの方式よりも認識率で大差がない上、手数時間が最も短いにもかかわらず主観評価の結果が低かったことは、主観評価に影響する要因として認識率・手数時間以外の要因があることを示唆する結果となった。現在のところ、この要因として利用者が検索キーワード以外の表現を

入力してはいけないという精神的な負担を想定しているが、今後は利用者が音声を入力する時点でボタンを押下するなど、この仮説を裏付ける実験を行い、さらに優しいインタフェースを開発していきたい。

また、本論文では、従来からあるUNDOコマンドを利用する手法とは比較していないが、我々の手法では、同じ音声の繰り返し入力されたかを判定することにより、UNDOコマンドを利用する場合と比べて手数が少なくできるため、UNDOコマンドを利用するよりも高い評価を得ると考えている。しかし、この比較評価も今後行いたいと考えている。さらに、認識率においてはまだ改善すべき点が多いと思われ、認識率の向上も進めていく予定である。

謝辞

本研究は通信・放送機構(TAO)からの受託研究「高齢者・障害者のための機能代行・支援システム技術の研究開発」の一環として実施した。

参考文献

- [1] 井ノ上直己,谷戸文廣,橋本和夫:”情報検索タスクにおける音声対話の分析”,音講論集 3-1-12, pp109-110,Sept. 1997.
- [2] 今井裕志,井ノ上直己,橋本和夫,米山正秀:”音声入力による情報検索システムのための音声認識誤り訂正方式”,情報処理学会第58回全国大会,pp.81-82,1999
- [3] 松下電気産業:”インターネット高速全文検索登録ソフトウェア PanaSearch / SSW マニュアル”,1996.