

論文 / 著書情報
Article / Book Information

論題(和文)	情報検索のための音声対話システムの構成法と評価
Title(English)	Construction and Evaluation of Spoken Dialogue Systems for Information Retrieval
著者(和文)	斉藤朗, 古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会 2001年春季講演論文集, Vol. , No. 1-3-22, pp. 47-48
Citation(English)	, Vol. , No. 1-3-22, pp. 47-48
発行日 / Pub. date	2001, 3

情報検索のための音声対話システムの構成法と評価*

◎齊藤 朗 古井 貞熙 (東工大)

1. はじめに

現在我々は、店舗情報検索を目的とした対話システムの開発を行っているが、ユーザにとって使いやすいシステムを構築するためには、自由発話による入力が望まれる。そこで、検索キーワードを自由発話により受理するシステムを開発し、ユーザから高い評価を得ることを目指しているが、自由発話の認識性能やシステムの構成法によっては、孤立単語発声の入力の方がユーザにとって使いやすいという場合がありうる。実際、本稿でも示すが、我々が初期に構築したシステムでは、孤立単語発声システムよりも低い評価結果となってしまう。

一方、システムの改良を進めるにあたって、ユーザ評価を強く反映する客観的な数値（コスト）を定義することが出来れば、被験者を要する主観評価実験を不必要に行わずに済み、改良の方針も立ちやすい。そこで、この客観的な値を初期のシステムによる被験者実験を通して定義し、その数値を有効に減らす方針でシステムの改良を行う。

2. 初期のシステム構成

タスクとしては東京近郊の店舗情報案内を扱っている[1]。具体的には、ユーザに場所と業種を発声してもらい、その2つのキーワードから検索される情報をユーザに提示するものである。

NTT-AT社の「高精度音声認識ソフトウェア」の開発キットを用い、PC上で動作する対話システムを作成した。主な仕様は、マイク入力、ディスプレイ表示（認識結果は最大10個まで表示）、マウス操作で、ユーザへは音声ガイダンスで指示を与える（システム主導型）。そこでまず、以下に示す3種類のシステムを作成した（初期システム）。

ver1. 場所のみ入力→確認→業種のみ入力→確認→受理

ver2. 場所+業種入力→場所確認→業種確認→受理

ver3. 場所+業種入力→場所業種確認→受理

各システムの確認段階で、候補がない場合は再入力するものとする。入力に関して、ver1は単語発声、ver2,3は自由発話による入力を許している。認識部には前述の開発キット付属の音響モデル

と、言語モデルとして単語ネットワークを用いた（キーワード数:約550, ガーベージ:約50）。

3. 初期システムによる被験者実験

3.1 実験条件

3つのシステムを36名（男性28名, 女性8名）の被験者に使用してもらい、操作性などを5段階（1:非常に良い~5:非常に悪い）で主観的に評価し、さらにシステムの順位付けを行ってもらった。ユーザの慣れによるデータの偏りをなくすため、使用するシステムの順番はランダムで行った。

3.2 実験結果

被験者による主観評価結果を表1に示す。次に、この主観評価を反映すると考えられる数値として、「ユーザが発声し結果が返ってくるまでの時間(U)」、「入力音声の中のキーワードの認識率(P)」、「返ってきた結果をユーザが見て、次の行動を起こすまでの反応時間(C)」、「候補を選択する時間(S)」の4つに着目し、これらを a)場所+業種、b)場所のみ、c)業種のみ、のそれぞれの入力場面での値を測定した。各平均値を表2に示す。

表1. 主観評価結果

	認識性能	認識時間	操作性	順位
ver1	2.47	2.75	2.42	1.33
ver2	2.92	3.39	2.58	2.17
ver3	3.03	3.31	2.89	2.47

表2. 客観的データ

	U(秒)	C(秒)	S(秒)	P(%)
a	14.72	5.45	4.18	48.9
b	6.30	3.09	4.02	77.0
c	7.62	3.05	4.01	71.0

表1を見ると、今回の実験ではユーザの主観評価結果は、全ての項目でver1が最も良い値となっており、3つの初期システムの中で、孤立単語発声のシステムがユーザにとって最も使いやすいという結果となった。

4. システムの改善

初期システムでの結果を踏まえ、自由発話入力のver2, ver3の改善を考える。

4.1 システム構成の改善

場所+業種入力で、場所、業種個別の認識率を調

*Construction and evaluation of spoken dialogue systems for information retrieval.

べた所、「場所が誤りで業種が正解」となる確率は約10%程と他の組み合わせに比べ小さいものであった。この性質を利用し、ver2を「場所が誤っていた時は、業種も誤っていると考え、確認せずに無条件で業種の発声をユーザに求める」というシステム構成に変更した(ver2')。

4.2 コストの定義

表2を見ると、場所と業種を入力する時(a)の、U,C,Pの値が他の2つ(b,c)よりも大きく劣っており、これが表1の主観評価につながっていると考えられる。そこでこれらを用いたコストを定義し、このコストを減らす方針でシステムの改良を行う。以下にコストの定義を示す。

発声コスト(U_t) = 1タスクを完了するまでに
要する総発声時間の期待値

反応コスト(C_t) = 1タスクを完了するまでに
要する総反応時間の期待値

4.3 コストの削減

U_t を削減するためにはU,Pの値を、 C_t を削減するためにはC,Pの値を改善する必要がある。そこでまず、音声認識ソフトウェアのバージョンをあげ認識性能(P)の向上をはかった。これにより、aにおけるPの値は68%となり、20%近く向上した。同時に認識時間も改善され、 $U=8.37$ と6秒近く短縮された。また、Cの改善のために、ボタン表記の簡素化、ボタン数削減などの、インタフェースの改良を行った。その結果、 $C=3.15$ と2秒以上短縮することが出来た。

5. 改良システムの被験者実験

3節の時と同じ条件のver1、認識性能を改善したver2',ver3(それぞれver2'R,ver3Rとする)、さらにこれら2システムのインタフェース改良版(ver2'RI,ver3RI)の計5システムについて、11名の話者(男性10名,女性1名)による被験者実験を行った。実験で得られたユーザの主観評価結果、及び発声コストと反応コストを表3に示す。

表3. 改良システムの主観評価結果と各コスト

	認識 性能	認識 時間	操作 性	順位	U_t	C_t
ver1	1.82	2.00	2.45	3.00	18.9	8.31
ver2' R	1.82	2.18	3.00	3.82	13.0	8.42
ver3R	2.27	2.18	3.55	4.73	12.0	9.05
ver2' RI	2.00	2.64	1.73	1.55	13.0	6.42
ver3RI	1.82	2.09	1.73	1.91	12.0	6.69

表3をみると、 U_t とユーザの主観評価にはあまり相関が見られない。一方、 C_t は操作性や順位とよ

く合致しており、ユーザの主観評価を反映する数値であると考えられる。ここで、aでの認識率を変化させた時の、各システムにおける C_t の理論値の変化を図1に示す。この図から傾向を見てみると、認識性能を変える前(認識率49%)の時、ver2'R,ver3R(初期システムのver2,ver3に相当)はともにver1よりも C_t が高く、表1の順位とも一致している。認識性能を改善し、認識率68%にした時も、両システムはver1より C_t の値が高く、評価順位もver1より劣っている。しかし両システムのインタフェースを改良したもの(ver2'RI,ver3RI)については、 C_t の値がver1よりも下がり、順位もver1より良いものになった。このような関係をみてもコスト C_t が主観評価値とよく合致していることが見受けられる。

また、この図からは、認識性能の改善だけでは認識率を75%以上にしなければver1を上回らないが、インタフェースの改良により、比較的低い認識率であってもver1より良い結果を得ることが可能であると予想もされ、インタフェースの改良が有効であったことも伺える。

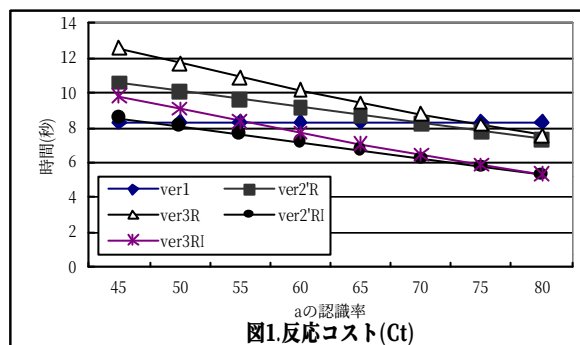


図1.反応コスト(C_t)

6. まとめ

インタフェースの改善を伴う反応コスト C_t を削減するような改良を行うことで、自由発話を入力とするシステムの評価を向上させることができた。これは、本システムにおいては、認識に要する時間よりも、判断決定する手間(時間)の方が、ユーザの評価に影響を及ぼしていることを示唆している。

今後は、ユーザへの使用法の説明やヘルプ機能など、今回は扱わなかった要因の、システム評価への影響について検証を重ねていきたい。

謝辞

ご助言を頂いたNTTサイバースペース研究所の、大附克年氏に感謝します。被験者実験に御協力頂いた東工大学生、古井研究室の方々に感謝します。

参考文献

- [1]山口他, “音声対話を用いた情報検索システムの研究”, 春季音講論集, 3-6-19, 1998.