2003－HI－103 (1)
2003／5／16

PDA で動作する旅行会話向け音声翻訳システムのインタフェース評価

水谷 研治　小沼 知浩　遠藤 充　南部 太郎　脇田 由実

あらまし

キーワードに基づく用例検索型の音声翻訳システムにおいて、検索モードを複数化したユーザインタフェースの評価実験を行った。評価尺度としては用例の検索時間と検索精度を使用した。検索時間が 30 秒以内の場合で、複数のモードを自由に使用できる場合の検索精度はクローズドテストで 86.8%、オープンテストで 76.8%であった。また、クローズドテストにおいて 1 回の操作で検索が成功するときの平均検索時間は 10.3 秒であった。同時間内での検索精度は 1 つの検索モードだけで検索する場合の最大検索精度よりも 12.0%高く、検索モードの複数化によるユーザビリティの向上が確認された。

User interface evaluation of a speech translation system for travel conversation installed in PDA

Kenji Mizutani　Tomohiro Konuma　Mitsuru Endoh　Taro Nambu　Yumi Wakita

Abstract

We evaluated user interface of our keyword-driven speech translation system whose input modes to retrieve an example sentence were experimentally enhanced. Our evaluation criteria consisted of the retrieval time and precision. When testees could use all of the modes at will, the precision within 30 sec was 86.8% for a closed test set and was 76.8% for an open test set. When they succeeded to retrieve with one operation, the average time was 10.3 sec for a closed set. The precision within the time was 12.0% higher than the maximum one when only one of the modes was available. Multi-modality of the interface was effective to raise the system's usability.

1 はじめに

人と人のコミュニケーションにおいて、相手の発声を待つ時間は双方に心理的な負担を与える場合が多い。特に音声翻訳システムを用いて異言語の２人が対話をする場合、ユーザが所望の翻訳結果を獲得するまでに費やす時間はコミュニケーションの本質にまったく関係が無い。システムの操作時間は音声翻訳システムのユーザビリティに直接的に大きな影響を与える要因の１つであり、可能な限り短い方が望ましい。

われわれはこのような操作時間を短縮するためにキーワード主導型の多言語音声翻訳システムを提案してきた。[1] 従来、用例検索のための入力インタフェースのモードとしては、音声認識による文入力だけであったが、PDA への移植に際して、実験的にソフトキーボードによるキーワード入力と、対話場面による用例の絞込みという 2 つのモードを追加した。

本稿では、モードの複数化がシステムのユーザビリティに与える影響を調べるために行ったユーザインタフェースの評価実験とその結果について報告する。入力インタフェースの評価尺度には用例の検索時間と検索精度を使用した。まず、入力インタフェースとして１つのモードしか使用できない場合と、すべてのモードを自由に使用できる場合とで各モードの比較実験を行った。また、システムが保持する用例を被験者が既知の場合と未知の場合についても比較実験を行い、実験室環境と実際の使用状況に近い環境との差分を確認した。

松下電器産業株式会社 先端技術研究所 ヒューマンウェア研究所
Humanware Technology Research Laboratory, Advanced Technology Research Laboratories
Matsushita Electric Industrial Co., Ltd.

2 音声翻訳システムの概要

2.1 ソフトウェアの構成

PDA で動作する音声翻訳システムの実現には様々な手法が提案されている。[2][3] しかし機械翻訳による誤訳をユーザが検出できないとコミュニケーションがスムーズに進行しない可能性がある。

われわれが提案するキーワード主導型の音声翻訳では、正しい対訳が付けられた用例データベースを利用し、その対訳の範囲内で翻訳結果を出力する。したがって理論的に誤訳は発生しない。ユーザはまず原言語の用例を検索してから、それを目的言語に変換する。[4][5]

原言語の用例を検索するために、ユーザは文発声による音声検索、ソフトキーボードによるキーワード検索、対話場面の絞込みによるアイコン検索のいずれかを行う。図１にシステムの構成を示す。

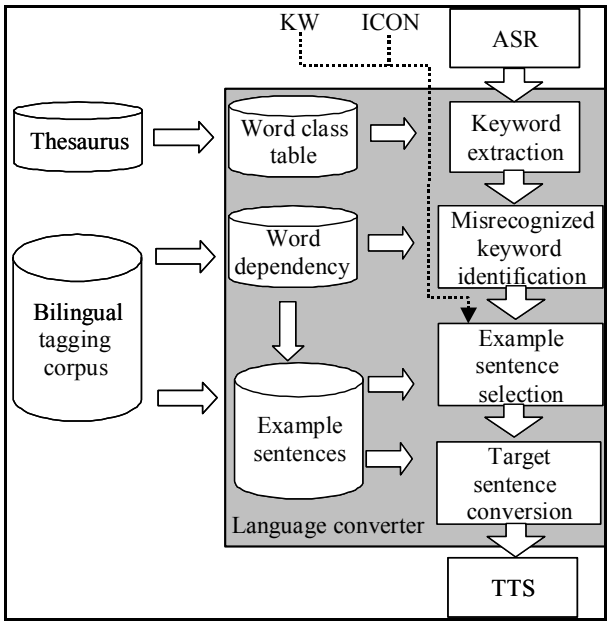

図１　音声翻訳システムの構成

2.2 音声認識部

日本語について連続音声認識を行う。音響モデルは不特定話者に対応する。声道長正規化係数は可変で話者に適応する。言語モデルは 1493 単語の trigram で、平均単語正解率は 78.1%である。

2.3 言語変換部

日本語から目的言語への言語変換を行う。音声認識結果に含まれる誤認識単語を訂正する機能を備える。用例は 1392 文、クラス単語は 737 語である。

用例には 1371 のキーワードが関連付けられ、キーワードから直接検索することもできる。また、用例は 85 の対話場面に分類されている。対話場面は木構造を構成し、最大３階層である。

音声認識の結果を中国語へ変換する一例を図２に示す。

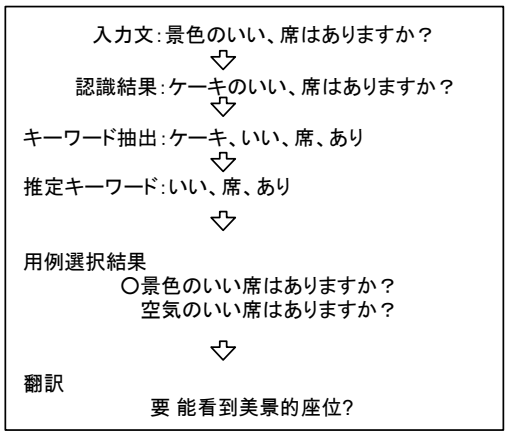

図２　言語変換の一例

2.4 ハードウェアの仕様

音声翻訳システムを実装した PDA の仕様を以下に示す。

- CPU: Intel StrongARM 206MHz
- RAM: 64MB(SDRAM 100MHz)
- 外部メモリ: 使用しない
- OS: Microsoft PocketPC（日本語版）

ソフトウェアの開発には Microsoft eMbedded Visual C++ 3.0 を使用した。

2.5 ユーザインタフェースの仕様

システムは基本的にスタイラスで操作し、ボタンやカーソルキーなどは使用しない。画面下部のメニューをタップするとそれぞれの検索モードが起動する。

音声検索モードを図３に示す。ユーザは音声検索開始ボタンを押してから文を発声する。音声認識終了後、検索された用例がリスト表示される。一度に閲覧できる用例は７文であり、それ以上の用例はスクロールバーを操作して閲覧する。所望の用例が見つかったらそれをスタイラスでタップすると目的言語に変換されて合成音声が出力される。

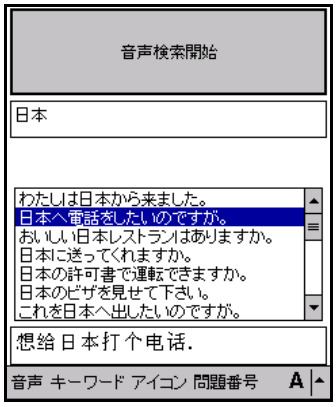

図３　音声検索のインタフェース

キーワード検索モードを図４に示す。ユーザは画面上のソフトキーボードを用いてキーワードの平仮名読みを入力する。第１キーワードは読みに従って前方一致検索され、リスト表示される。一度に閲覧できる第１キーワードは７単語であり、それ以上のキーワードはスクロールバーを操作して閲覧する。所望の第１キーワードが見つかり、それをスタイラスでタップすると第１キーワードを含む用例がリスト表示される。用例の表示と操作は音声検索の場合と同じである。

ただし、表示される用例の数が 15 を超える場合は、その用例に含まれる第１キーワード以外のキーワード（第２キーワード）が表示される。一度に閲覧できる第２キーワードは 5 単語である。第２キーワードを選択すると用例が表示される。

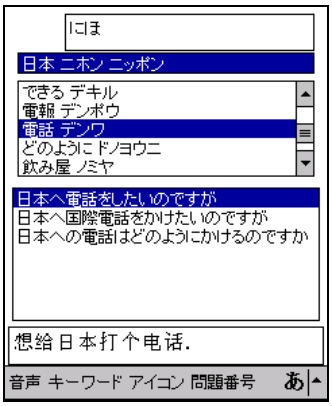

図４　キーワード検索のインタフェース

アイコン検索モードを図５に示す。ユーザは対話の場面をアイコンで選択しながら用例を絞り込む。第１階層のアイコンを選択するときは用例を参照することができないが、第２階層以下では図６に示すように、音声検索と同条件でその階層以下に含まれるすべての用例のリストを閲覧することができる。

図５　アイコン検索のインタフェース
（第１階層）

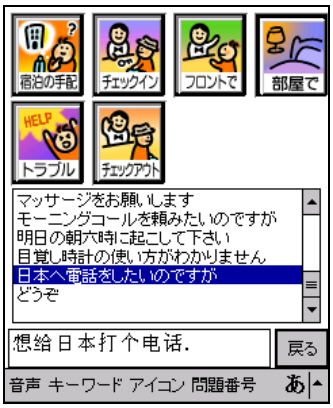

図６　アイコン検索のインタフェース
（第２階層以下）

3 ユーザインタフェースの評価方法

3.1 前提条件

PC や携帯電話などの情報機器を日常的に使用していて、GUI の操作に精通していることを被験者に要請する。被験者には実験前に本システムの使い方を十分に説明する。特に、音声認識とソフトキーボードについては実際に何度も使用してもらい、発声タイミングや声の大きさ、平仮名の入力方法について習熟しておく。また、本システムで検索することが可能な用例とその分類構造についても開示する。

なお、本システムの実用的な観点から、検索時間の上限は 60 秒に設定し、60 秒を超えた場合は検索に成功しても検索失敗として扱う。

3.2 クローズドテストによるモード比較

被験者に用例データベースの中から１つの用例を提示して、異なる３つのモードで検索してもらい、それぞれの検索時間を測定して比較する。

ところで、１つの用例の検索が T 秒以内で成功する確率 P(t≦T)は、ユーザがモードとして音声検索、キーワード検索、アイコン検索を使用する確率をそれぞれ P(V), P(K), P(I)とすると、

$$P(t \leqq T) = P(t \leqq T \mid V)P(V) + P(t \leqq T \mid K)\,P(K) + P(t \leqq T \mid I)\,P(I)$$

ただし $P(V)+P(K)+P(I) = 1$

となる。P(V), P(K), P(I)を適切に推定するためには長期的な実験を行う必要があるので、短期的な実験では P(V) = P(K) = P(I)の場合の P(t≦T)を平均的な検索精度として採用することがある。しかし、システムの入力インタフェースに対して習熟が進むと、ユーザは検索時間を最短にする検索モードをある程度の確度で選択するようになると予想される。そこで、１つの用例の検索が T 秒以内で成功する限界確率 $P_{sup}(t \leqq T)$を、

$$P_{sup}(t \leqq T) = \max_{m \in \{V,K,I\}} P(t \leqq T \mid m)$$

と定義し、インタフェース性能の上限値として次節以下の実験で利用する。

3.3 クローズドテストによる I/F 性能評価

被験者に用例データベースの中から１つの用例を提示して検索してもらい、検索時間を測定する。ただし、モードの使用制限は与えない。どのモードから検索を開始してもよく、また、あるモードでの検索を断念して別のモードで検索を開始してもよい。

3.4 オープンテストによる I/F 性能評価

被験者には直接用例を提示せず、海外旅行中のある場面を提示して、その場面の登場人物の１人になったつもりで自発的に用例を考えてもらう。そして用例が確定してからそれを検索してもらい、検索時間を測定する。

なお、被験者が考えた用例がそのままシステムに存在しない場合がある。本システムの実際の使用場面を想定すると、相手との対話においては、ユーザは用例の正確さよりも相手に与える待ち時間の方を優先する傾向がある。したがって、主観的ではあるがこれでよいと思った用例を検索結果として確定するように指示する。

3.4 操作回数から見た I/F 性能評価

以上の評価方法は実時間から見た検索時間である。しかし、音声翻訳システムを使用するユーザが心理的に感じる時間の経過は、検索成功までの操作回数によるところが大きい。

ここで操作回数を以下のように定義する。

- 音声検索モード
  発声開始ボタンを押して発声→１回
- キーワード検索モード
  第１キーワードの選択→１回
  第２キーワードの選択→１回
- アイコン検索モード
  最上位の階層の切り替え→１回

そして、n 回の操作で検索が成功するときの平均検索時間 T(n)を、実時間軸を定性的に解釈するために利用する。また、N 回以内の操作で検索が成功する確率 P(n≦N)を計算して、その増加率も確認する。

4 実験結果と考察

19 人の被験者（男性 13 名、女性６名）について実験を行った。

各被験者についてクローズドテストによるモード比較、クローズドテストによる性能評価、オープンテストによる性能評価を連続的に実施した。

4.1 クローズドテストによるモード比較

10 文の用例について行った実験結果を図７に示す。限界性能 $P_{sup}(t \leqq T)$は 30 秒で 97.9%である。

９ 秒以内の検索ではアイコンによる検索精度 P(t≦T | I)が最大であるが、９秒を超えると音声による検索精度 P(t≦T | V)が最大になる。アイコンの検索精度が高いのはドメイン構造の記憶によるところが大きいと考えられる。

16 秒を超えるとアイコンよりもキーワードによる検索精度 P(t≦T | K)が高くなる。これは、被験者の多くがキーワードの前方一致検索機能を活

用しなかったために、ソフトキーボードによる第１キーワードの入力に平均 10 秒の時間を要しているからである。なお、実験の範囲外ではあるが、検索時間 60 秒以上ではキーワード検索の精度は音声検索の精度を上回っている。

アイコン検索では、他のモードほど時間の経過と共に検索精度が上昇しない。これは、用例が含まれると思う最初のアイコンで検索に失敗すると、それ以外のアイコンを場当たり的に検索する傾向があるからである。

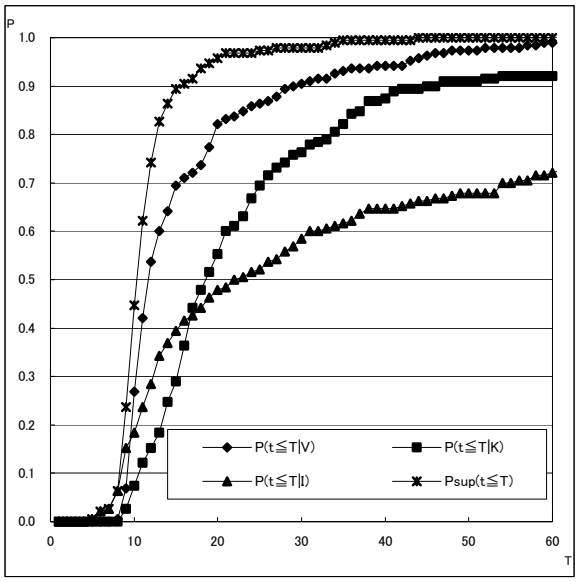

図７　モード比較評価

4.2 クローズドテストによる I/F 性能評価

10 文の用例についてモードの使用に制限を設定せず、検索時間を測定した。この実験を開始する時点で、被験者は各モードの特性をある程度習熟済みである。T 秒以内で検索が成功する確率 P_{cl}(t≦T)を図８に示す。検索時間 30 秒の検索精度は 86.8%である。

4.1 の実験結果と比較すると、13 秒以内の検索ではインタフェースの限界性能に近づいている。これは被験者が、検索時間が最短になるように各モードを使い分けたためである。しかし 13 秒以上では、音声検索だけを使用する場合よりも最大 5.8%低い値となっている。これは被験者がモードを切り替えてすぐに検索を開始することができない場合が多く、検索開始前の思考時間によると考えられる。

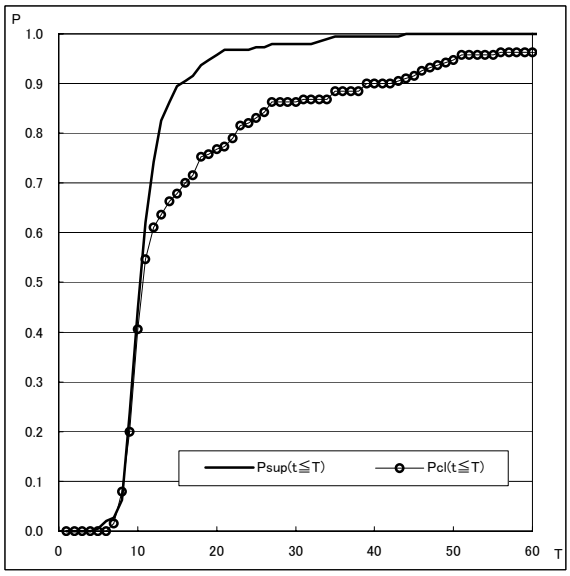

図８　クローズドテストによる I/F 性能評価

4.3 オープンテストによる I/F 性能評価

10 の対話場面についてモードの使用に制限を設定せず、検索時間を測定した。T 秒以内で検索が成功する確率 P_{op}(t≦T)を図９に示す。検索時間 30 秒の検索精度は 76.8%である。

期待した通りの用例が検索できない場合や、それに近い用例が出現した場合の判断時間が含まれているので、4.2 の実験結果よりもインタフェースの限界性能から離れる結果となった。

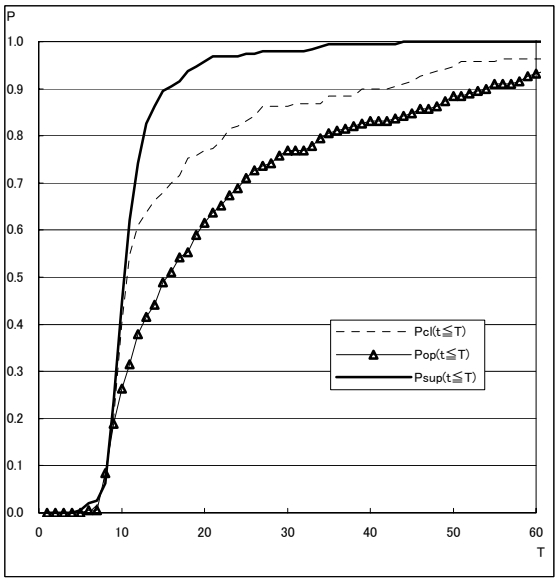

図９　オープンテストによる I/F 性能評価

4.4 操作回数から見た I/F 性能評価

操作回数が n 回のときの平均検索時間 T(n)を、4.2 および 4.3 の実験結果について計算した T_{cl}(n)と T_{op} (n)を図１０に示す。1 回の操作で検索が成功するときの T_{cl}(1)は 10.3 秒である。

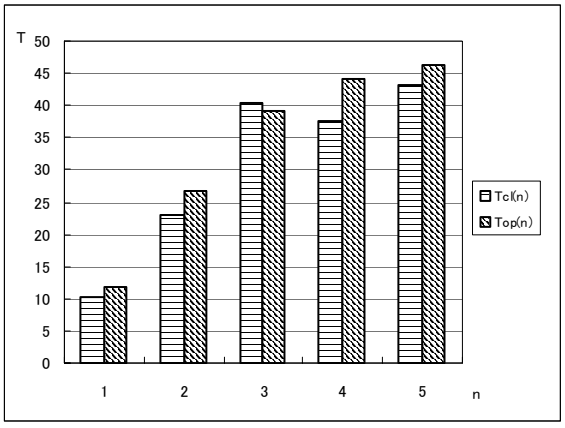

図１０　操作回数と平均操作時間の関係

4.1 の実験結果において、同時間内での検索精度は 1 つの検索モードで検索する場合の最大検索精度（音声検索モードの 42.1%）よりも 12.0%高いことから、モードの複数化は特に 1 回の操作で検索が成功する場合に有効であることがわかる。

また、N 回以内の操作で検索が成功する確率 P(n≦N)を、4.2 および 4.3 の実験結果についてそれぞれ計算した P_{cl} (n≦N)と P_{op} (n≦N)を図１１に示す。ユーザが受けるシステムの印象は、1 回だけ失敗が許されるときの値と大きく関係する。本システムでは、クローズドテストで 89.4%、オープンテストで 81.6%である。

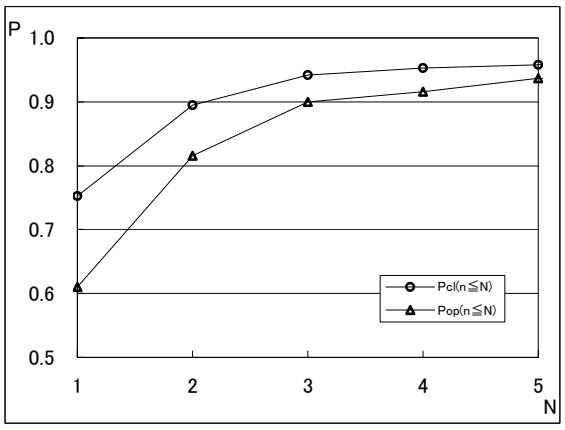

図１１　操作回数と検索精度の関係

5 おわりに

用例の検索時間と検索精度について音声翻訳システムの入力インタフェースの評価を行った。複数のモードを自由に使用できる場合は、30 秒以内の検索精度はクローズドテストで 86.8%、オープンテストで 76.8%であった。また、クローズドテストにおいて 1 回の操作で検索が成功するときの平均検索時間は 10.3 秒であった。同時間内での検索精度は 1 つのモードのみ使用する場合の最大検索精度よりも 12.0%高く、モードの複数化による検索精度の向上が確認された。

実際の使用状況を考慮すると所望の用例が検索できるまで 1 つのモードを連続的に使用することは少ない。4.1 の P(t≦T|V)、P(t≦T|K)、P(t≦T|I)はいくらか過大な評価値であり、4.2 の P_{cl}(t≦T)の方が実際的な評価値である。したがって 2 回以上の操作で検索が成功する場合についても同等の検索精度が得られていると考えられる。

今後は中国語以外の他の言語への応用を考えながら[6]、音声翻訳システムのユーザビリティ向上について研究を進める予定である。

参考文献

[1]T.Konuma, K.Matsui, Y.Wakita, K.Mizutani, M.Endoh, M.Murata: An experimental multilingual bi-directional speech translation system, 9th International Conference on Theoretical and Methodological Issues in Machine Translation, 2002

[2]山端潔, 磯谷亮輔, 安藤真一, 花沢健, 石川晋也, 江守正, 磯健一, 服部浩明, 奥村明俊, 渡辺隆夫: PDA で動作する旅行会話向け日英双方向音声翻訳システム, IPSJ SIG-NL150, 2002

[3]高電社:翻訳ウォーカーj・北京, http://www.kodensha.jp/

[4]Y.Wakita, K.Matsui, Y.Sagisaka: Fine keyword clustering using a thesaurus and example sentences for speech translation, ICSLP2000

[5]Y.Wakita, K.Matsui, Y.Sagisaka: Robust speech translation using fine keyword clustering, Workshop on Multi-Lingual Speech Communication 2000

[6]堀一成, 青野繁治, 藤家洋昭, 石島悌, 脇田由実, 高階美行: 『多言語同時処理』研究の射程と言語間バリアフリー, 情報処理学会第 65 会全国大会, No.5, pp347-350, 2003