

F-9 大語彙音声認識を用いた PDA 向け音声 UI の試作

林美奈子 長友健太郎 江森正 磯健一 畑崎香一郎

NEC マルチメディア研究所

1. はじめに

PDA の需要が年々拡大している。PDA は、小さく持ち運びが簡単である反面、画面が小さく、文字入力がかしこいという弱点がある。従って、歩きながらや片手がふさがった状態での入力は難しい。こうした弱点を補うためにいくつかのユーザインタフェース(UI)の研究が行われている[1]。

今回、音声入力と通常の画面入力とを組み合わせることで PDA を使いやすくするため、PDA 向けの大語彙音声認識エンジン[2]で動作する音声 UI を開発した。試作では、経路探索ソフトの出発地および到着地を入力するための UI として実装した。

2. 大語彙音声認識を用いた音声 UI

2-1 音声 UI を用いた駅名検索

音声 UI は、大量の選択肢から1つを選び出す場合有効である。例えば、経路探索ソフトの駅名入力では、多数の駅の中から選択するために階層メニューを使用したり、文字入力のためにソフトキーボードを使ったりするため、たとえ両手が使うことができても非常に煩雑な操作になる。一方、音声 UI では、目的の選択肢を発声するだけでよいので、前述の階層メニューなどに比べ有効である。また、キーボードやマウスなどに比べ導入時の訓練がほとんど必要ないことも利点の一つである。

こうした背景に基づいて、市販のアプリケーションをそのまま利用して、自然に音声を入力できることが音声 UI を開発した。具体的には、音声 UI を市販の経路探索ソフトであるJRトラベルナビゲータと組み合わせて試作した。ユーザは、従来のソフトキーボードでの入力の代わりに、PDA 側面の録音ボタンを押しながら駅名を発声する。図1に示すように、認識結果は、出発地または目的地のテキストボックスに入力される。また、画面も従来空白であった部分に認識エンジンの状態を表示していることを除けば、従来のアプリケーションと同様である。

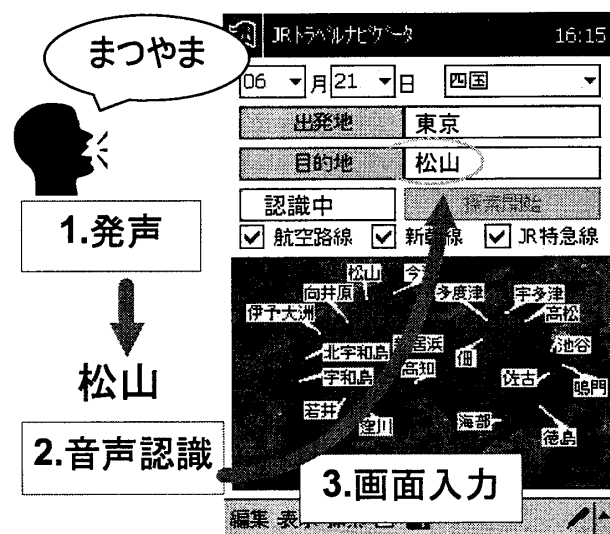


図1 音声 UI を用いた入力操作

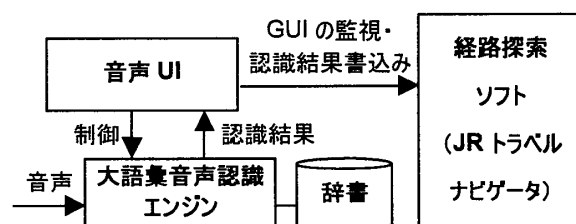


図2 システム構成

2-2 システム構成

システム構成を図2に示す。音声 UI は、PDA 上で動作する大語彙音声認識エンジン上に実装されている。音声 UI は、認識エンジンに対して動作の制御と認識辞書の切替えを行う。また経路探索ソフトの GUI を監視し、また、到着地や目的地のテキストボックスへ認識結果の書き込みを行う。

本システムで用いた音声認識エンジンは、ネットワーク文法を受け付けることができる連続音声認識エンジンで、駅名などの大語彙固有有名詞を圧縮した木構造辞書として扱うことにより少ないメモリ量で動作することができる[2]。

2-3 音声入力の制御

(1) 認識辞書の切替え

経路探索で入力可能な日本全国の駅は2万6千以

上であるため、関東地方など地域ごとに辞書を切替えることで各地域用辞書のサイズを低減し、認識性能とレスポンス速度を向上させた。

また、経路探索ソフトの地域切替え機能と連動して辞書を切り替える仕様とした。ユーザが地域を変更すると、自動的に変更後の地域の辞書がロードされるのでユーザが辞書切替えについて意識することはない。

(2) マイクスイッチ

ユーザは、録音ボタンを押しながら音声を入力する。入力の UI をメモの録音と統一することでユーザにとって直感的な入力が可能になる。また、録音ボタン制御は、録音ボタンを押している間だけ音声認識がオンになるプッシュ・トゥーク方式を採用した。このように明示的に音声入力の開始/終了を指定するので周辺雑音の混入を抑えられる。

3. ユーザビリティ評価

アプリケーションの使い勝手を評価するために、年齢や性別の異なるユーザで経路探索を行うユーザビリティ評価を行った。表1に実験条件を示す。

表 1 実験条件

被験者	7名(男性3名、女性4名)
辞書	離散単語辞書 (関東地域の全駅名 1877 駅) (関西地域の全駅名 1363 駅)
話者モデル	男女不特定モデル
実験環境	会議室

表 2 ユーザに与えられた課題

1)	田町から御徒町までの経路
2)	赤坂(東京)から桜木町(横浜)までの経路
3)	分倍河原(東京)から難波(大阪)までの経路
4)	自宅から NEC 本社(三田/田町)までの経路

実験では、表2に示すように各ユーザに 1)発声した駅名が選択地域内に一つしか存在しない経路、2)発声した駅名が選択地域内に複数存在するため候補から選択を要する経路、3)操作の途中で地域の切替えすなわち辞書の切替えを要する経路、4)地域の切替えを含み、かつユーザ自身が設定する経路の 4 つの課題で経路探索を行ってもらった。また、実験では、被験者に対し操作方法を紙面で説明

し、口との距離や姿勢など音声 UI の利用に関する説明は行わなかった。

表 3 経路探索課題における正解率

	1回目	2回目(言い直し)	3回目(言い直し)
1)	86%	100%	100%
2)	57%	76%	93%
3)	36%	64%	93%
4)	71%	86%	100%

表3に示すように、課題 1)では、1 回目の正解率が 86%、言い直しを含め 3 回までの発声まで範囲を広げると 100%と、高い。課題 2)および 3)では、1 回目の正解率が 57%、36%と低い。これは、課題 2)ではタップと発声とを交互に行う過程での録音ボタンの押し忘れたこと、課題 3)では地域の切替え忘れたことが主な原因である。しかし、言い直しを含めた場合、課題 2)および 3)の正解率は、ともに 93%と高くなった。これは、操作の過程で使い方を習得したと考えることができる。もっとも複雑な入力操作を要する課題 4)でも、1 回目の正解率が 71%、言い直しを含めると 100%と高い。これは、1)~3)の課題評価の過程で使い方を習得したことによると考えられる。以上から、試作した音声 UI は複雑な説明や訓練なしで使いこなすことができることがわかった。

4. おわりに

PDA 向け大語彙音声 UI を試作した。これを経路探索ソフト JR トラベルナビゲータに組み込み、評価を行った。評価の結果、実用的な認識性能と使い勝手が得られることがわかった。

謝辞

(株)ジェイアール東日本企画殿、(株)ビーマップ殿に、JR トラベルナビゲータの利用のご承諾とご協力を感謝いたします。

参考文献

- [1] Y.Y. Wang, "Robust Language Understanding in MIPAD", Eurospeech2001, 2001
- [2] 友枝他,「木構造辞書とネットワーク文法を用いたコンパクト大語彙連続認識エンジン」, 音学論 2001 年 3 月