

MOSAIC ブラウザーを用いた音声対話システム

5 J-1

木村貞弘 中村 哲 鹿野清宏*

奈良先端科学技術大学院大学 情報科学研究科

1 はじめに

近年インターネットやパソコンの普及に伴い、さまざまな情報がネットワークを通して提供されている。この様な状況の中で、これらネットワーク上の情報を誰もが簡単にアクセスできるためのインターフェースの必要性が生じている。現在 GUI (Graphical User Interface) を用いてこれらの問題に対処しているツールとして、MOSAIC ブラウザーなどが挙げられるが、膨大なデータの中から目的のデータを選択する事は、非常に困難である。そこで今回我々は、MOSAIC のインターフェースに音声入出力を加えてマルチモーダル化したシステム Keytarō を開発した。本稿ではこのシステムの構成と評価実験について示す。

2 音声対話システム Keytarō

2.1 システムの特徴

今回開発したシステムは以下の様な特徴を持っている。

- 基本的タスクドメインは学内案内。
- インターネット上の情報を音声で検索する事ができる。(検索は MOSAIC が行なう)
- MOSAIC の標準的インターフェースを音声で操作できる。

インターネット上の任意のドメインの音声対話を実現する事が、本システムの最終目的であるが、任意のドメインに対処する対話システムの実現は、極めて困難なため、今回は学内案内のドメインに限定し、それ以外のインターネット上の情報に関しては、MOSAIC の機能である BACK、FORWARD、HOTLIST などサポートする事にした。本システムは MOSAIC だけでなく、多くのアプリケーションのマルチモーダル化を図っているため、基本的条件として、MOSAIC 自体の変更及びソースコードの書き換えは行なわず、外部から操作を行なう事とした。

2.2 Keytarō の構成

Keytarō は図 1 の様に、ユーザーと相互作用を行ないながら一つの目的を達成する。音声入力

部で解析され、認識結果文字列に変換される。対話管理部はその変換された文字列を基に、次発話文の予測など対話の流れに関する制御を行なう。意図抽出部はユーザーの意図を実行コマンドに変換する。その変換されたコマンドは、実行部 (MOSAIC) で処理される。

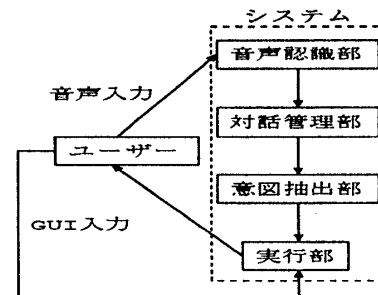


図 1: 音声対話システム Keytarō

2.3 各モジュールの説明

2.3.1 音声認識部

音声認識部は、システムと独立した音声認識サーバーとした。音声データは、この音声認識サーバーに転送され、特徴パラメータ化された後、HMM (隠れマルコフモデル) により認識される。この認識結果は文字列として出力される。

2.3.2 対話管理部

音声認識部より出力された結果文字列を基に、対話管理部は現在の対話の状態を分析し、次発話文の予測を行なう。予測された結果により、次発話における認識対象語彙を変化させる。またデータ構造が既知の場合 (学内案内タスクの場合)、曖昧な語彙に関しては問返しを行なう。ユーザーが「駅に行きたい」と発話した時、この状態では IS-A 関係となる「どこの駅」と言う情報が欠落している。この時、システムはユーザーに「どこの駅ですか」と問返しを行ない、結果文字列の補正を行なう。

2.3.3 意図抽出部

音声認識部、対話管理部を経たデータは、ここで実行コマンドに変換される。このコマンドは、2つに分類される。

- http アドレスを検索するシグナルを転送

*Speech Dialogue System with Mosaic Browser,
Sadahiro Kimura, Satoshi Nakamura and Kiyohiro Shikano,
Graduate School of Information Science, Nara Institute of Science and Technology;

● MOSAIC の標準的インターフェースを操作

まず、MOSAIC に http アドレスを検索するシグナルを転送する処理であるが、MOSAIC に USR1 というシグナルを送る事で可能である。これは MOSAIC に標準装備されている機能である。一方、標準的インターフェース (MOSAIC に装備されている BACK、FORWARD、HOTLIST などの機能) は外部から操作する事ができない。そこで、疑似的なショートカットキーのイベント信号を、MOSAIC に送る方法を採用した。ショートカットキーの情報を図 2 の様なウィンドウ TREE を再帰的に走査して、目的のウィンドウに送るものである。以上 2 点により MOSAIC を外部操作する。

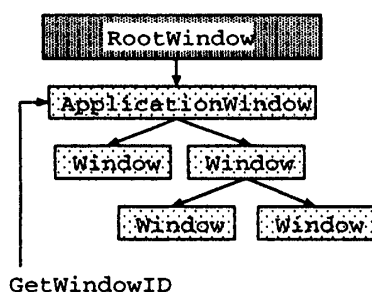


図 2: ウィンドウ TREE

最後にコマンドに変換する方法であるが、ここではキーワード抽出による意味解釈 [1] という方法を用いて行なう。これはタスク (学内案内) が単純であるため構文解析などを行なわなくても、意味解釈ができるためである。

3 評価実験

3.1 実験内容

被験者に本システムを実際使用してもらい、タスクが完了するまでの時間を計測した。ここで、(A)MOSAIC のみを使用した場合、(B)Keytarō を使用した場合のそれぞれにおいて計測した。データの検索には、データ構造の理解と入力との 2 段階がある。実験 (A) の MOSAIC のみを使用した場合は、データの構造は未知で、データ構造の理解は被験者が行なわなければならないが、Keytarō を使用した場合は、システムが行なう。従って、本実験はモダリティーの比較ではなく、システム自体の評価に関するものとなっている。つまり MOSAIC を単独で使用する時よりも、どの程度本システムが検索 (データ構造の理解) の困難さを解消でき、実行時間を短縮できるかの評価となっている。実験条件を表 1 に示す。

表 1: 実験条件

被験者数	4 人 (男性 3 名、女性 1 名)
音声録音時間	4 秒
対話数	20 対話 / 人 (全 80 対話)

ここで、この 4 人の被験者のうち、2 人は MOSAIC にはあまり慣れていない初心者であり、残りの 2 人は MOSAIC に精通している熟練者である。本実験では、このユーザーレベル別で比較した。

3.2 結果

実験結果を表 2 に示す。熟練者において、MOSAIC を単独で使用する時と、Keytarō を使用する時のタスク実行時間の優位な差は現れなかった。一方、初心者においては約 1.6 秒のタスク実行時間を短縮する事ができた。これは、熟練者にとっては Keytarō が有効であるとは言えないものの、初心者にとっては非常に有効であると言える。ちなみに今回の実験では、80 対話中 12 対話の認識ミスが生じ、正解率は 85% であった。

表 2: 結果

評価実験	初心者	熟練者
A	58.0 秒	41.5 秒
B	41.4 秒	39.3 秒

4 まとめ

MOSAIC のみで使用するより、マルチモーダル化した Keytarō は、初心者のタスク実行時間を短縮する事ができる。今回の評価実験で用いた音声認識サーバーの認識に所要する時間は、平均 1.6 秒であった。現在、より高速に認識するサーバーに変更したので、タスク達成時間をもう少し短縮する事ができる。今後の課題として、被験者数、被験者のレベルなどをもっと詳細に考慮した評価実験と、評価尺度を時間以外にも考える必要がある。また、任意のインターネット上の情報の知識構造を自動的に分析し、任意のドメインでの対話を可能にする手法も考えなければならない。

参考文献

- [1] 肥田、伊藤、中川:「音声対話システムにおける自然発話の頑健な一理解法」、情報処理学会第 50 回全国大会講演論文集、2-467、(1995)