自然言語処理　137－12
音声言語情報処理　31－7
（2000．6．2）

マルチドメイン音声対話システムの構築手法

長森 誠† 河口 信夫† 松原 茂樹‡ 外山 勝彦†§ 稲垣 康善†

名古屋大学大学院工学研究科† 名古屋大学言語文化部‡
名古屋大学統合音響情報研究拠点§

概要

これまでに，様々な音声対話システムが開発されているが，その多くは特定のタスクドメインを対象としている．近い将来，音声インタフェースの普及により複数の対話システムを同時に使用する状況が予想される．そこで本稿では，複数の音声対話システムの結合に基づく，マルチドメイン音声対話システムのアーキテクチャを提案し，その実現方法について述べる．本手法では，フラグメントの分配と統合という概念を用いてシステムを構築する．また，このアーキテクチャに基づいて実現した簡単なマルチドメイン音声対話システムについて述べる．

A Framework for Multi-Domain Conversational Systems

Makoto Nagamori† Nobuo Kawaguchi† Shigeki Matsubara‡
Katsuhiko Toyama†§ Yasuyoshi Inagaki†

Graduate School of Engineering,Nagoya University.†
Faculy of Language and Culture,Nagoya University.‡
Center for Integrated Acoustic Information Research,Nagoya University.§

Abstract

Many spoken dialogue systems have been developed so far. Most of them hold the architecture for managing a single task domain. However, such the architecture is not suitable for the conversational systems for managing several task domains. This paper proposes the architecture for multi-domain spoken dialogue systems. The key concepts are distribution and integration of data fragments. We have made an experiment on a simple system based on our architecture.

1 はじめに

近年，音声対話システムの研究が活発に進められ，特定のタスクドメインに対しては実用的なシステムが実現されつつある [1,3,4,5,6,7]．近い将来，様々な種類の音声対話システムが登場し，生活の中に浸透されることが予想される．例えば，自動車内においては，カーナビ，エアコン，カーオーディオなどに対する音声対話システムが搭載されると考えられる．

しかし，これらのシステムを独立に実現し，車内に設置した場合，ユーザの音声がどのシステムへの入力であるのかを各システムが正しく判断することは難しい．それらを同時に使用するには，複数の対話ドメインを処理可能な音声対話システムが求められる．我々は，このようなシステムをマルチドメイン音声対話システムと呼ぶ [2]．

本研究の目的は，拡張性を備え，しかも構成

が容易なマルチドメイン音声対話システムを実現することである．本稿では，それぞれ独立に作成された単一ドメインを処理する音声対話システムを組み合わせた，アーキテクチャを提案し，その構成方法について述べる．本手法では，フラグメントと呼ぶ，音声入出力や付加情報を表現したデータを各モジュール間で分配・統合する．すべてのモジュールがフラグメントを送受信できるように設計することにより，それらを結合するだけでシステムを構成できる．

本稿では，2節でマルチドメイン音声対話システムについて述べ，3節でマルチドメイン音声対話システムのアーキテクチャを提案する．4節では，対話ドメインの決定方法について述べ，5節では，類似した複数のドメインを扱うシステムのマネージャについて述べる．

2 マルチドメイン音声対話システム

マルチドメイン音声対話システムは，種類の異なる複数のドメインを扱う．

我々は，理想的なマルチドメイン音声対話システムは以下の3つの性質を持つべきであると考えた．

1. 拡張性
   ドメインを容易に追加することができる．それぞれのドメインの処理を独立に考えればよい．

2. スケーラビリティ
   多くのドメインを扱う場合でも，妥当な速度で処理できる．扱うドメインの数がユーザの応答性に大きな影響を及ぼさない．

3. ユーザビリティ
   あるドメインのみを利用する場合，ユーザはあたかも単一ドメインの音声対話システムを扱うかのように利用できる．

この中で最も重要な性質は拡張性である．目的は複数のドメインを快適に扱うことであり，扱うドメインはユーザの要望によって変化していく．ドメインの追加や削除が容易であれば，ユーザの要望を満たす柔軟なシステムの実現が可能になる．

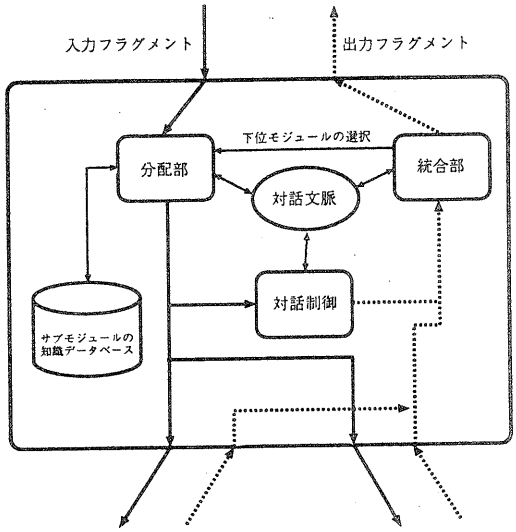

図 1: マネージャの設計

3 マルチドメイン音声対話システムのアーキテクチャ

本節では，提案するマルチドメイン音声対話システムのアーキテクチャについて述べる．我々は拡張性のあるシステムを容易に構成することを目指しており，各ドメインを単一の対話エンジンを用いて統一的に扱うのではなく，ドメイン毎に独立に対話システムを実現し，その結合によってマルチドメインシステムを実現する．

以下では，アーキテクチャを構成するモジュール，システムで扱うデータ，本アーキテクチャに基づいた典型的なシステムの動作例について述べる．

3.1 モジュール

本手法ではアーキテクチャの構成要素をモジュールと呼ぶ．モジュールを以下のように定義する．

- ワークモジュール
  特定のドメインの処理を行なうモジュールである．単独でも音声認識エンジンと音声合成エンジンと接続すれば音声対話システムとして使用することが可能である．

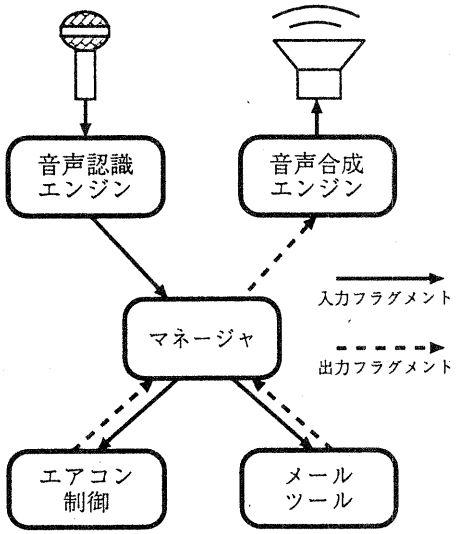

図 2: 典型的なシステム例

- マネージャ
  いくつかのサブモジュールと結合しているモジュールである．マネージャの役割はデータの分配や統合，結合しているサブモジュールの選択，ユーザとの対話の制御などであり，結合するサブモジュールに応じてその役割は変化する．マネージャを構成する要素として，受け取ったデータをサブモジュールに分配する分配部，サブマネージャから送られたデータを集め処理を行ない，上位のモジュールにデータを送る統合部，ユーザとの対話の制御を行なう対話制御部，ユーザとの対話の流れを情報を保持する対話文脈部，サブマネージャに関する知識を持つ知識データベースから構成される (図 1).

図 2 に本手法に基づく典型的なシステム構成例を示す．この例ではエアコン制御とメールツールという 2 つのワークモジュールをマネージャが結合することにより，2 つのドメインを扱うシステムを構築している．

3.2 フラグメント

システム内で受渡しされるデータをフラグメントと呼ぶ．フラグメントは，入力フラグメントと出力フラグメントの二種類あり，前者はユーザの発話に関する情報を含み，出力フラグメントはシステムの応答に関する情報を含む．フラグメントの例を図 3 に示す．フラグメントは属性と値の対の集合であり，フレーム構造で表される．図 3 の例の (1) では ID,phrase,relevance が属性であり，それらに続くものが値である．各モジュールは処理を行なう際，必要なアトリビュートや値だけを参照すれば良いので，どのようなモジュールでも扱うことが可能であり，フラグメントを扱うことのできる仕組みを持つモジュールであるならば容易にシステムを結合させることができ，また拡張することもできる．

```
(1)エアコン制御に対する入力フラグメント
//マネージャからエアコン制御に送信される
input :{ID          : 20147
        phrase      : 長森 から きた メール
        relevance   : 0.20 }

(2)メールツールに対する入力フラグメント
//マネージャからメールツールに送信される
input :{ID          : 20147
        phrase      : 長森 から きた メール
        relevance   : 0.80 }

(3)エアコン制御からの出力フラグメント
//エアコン制御からマネージャに送信される
output:{ID          : 20147
        module ID   : エアコン制御
        utterance   : (null)
        relevance   : 0.0 }

(4)メールツールからの出力フラグメント
//メールツールからマネージャに送信される
output:{ID          : 20147
        module ID   : メールツール
        utterance   : 3通あります
        relevance   : 0.80 }

(5)メールツールに対する入力フラグメント
//マネージャからメールツールに送信される
control:{ID         : 20147
         control    : selected }

(6)エアコン制御に対する入力フラグメント
//マネージャからエアコン制御に送信される
control:{ID         : 20147
         control    : no selected }
```

図 3: フラグメントの例

入力フラグメントがマネージャを経由してすべてのワークモジュールに分配され，ワークモジュールの処理結果である出力フラグメントをマネージャが統合する．マネージャは統合した出力フラグメントから処理を行なわせるワークモジュールを選択し，制御情報を持つ入力フラグメントをワークモジュールに分配することで 1 つの入力に対する処理が完了する．

3.3 動作例

本節では図2を用いてシステムの動作例を示す．() 内の数字は図3に対応するフラグメントを示している．

ユーザが以下の発話をした場合，

U:「長森 から 来た メール」

音声認識エンジンの認識結果がマネージャに渡される．マネージャは接続されているドメインの知識を用いてそれぞれのワークモジュールに対する入力フラグメント (1)(2) の関連度を計算する．関連度の計算は，例えばワークモジュールの語彙を用いる方法が考えられる．マネージャの分配部は2つのワークモジュールに入力フラグメントを分配する．入力フラグメントはID，認識結果，関連度についての情報を持つ．

2つのワークモジュールは入力フラグメント受信後，入力発話の解析を行なう．メールツールでは「長森」「から」「メール」を理解することができるため関連度を変更しない．エアコン制御ではすべての語句を理解することができないので，関連度を0にする．2つのワークモジュールは出力フラグメント (3)(4) をマネージャに送信する．出力フラグメントはID，ワークモジュールID，入力発話に対する応答，関連度を情報として持つ．

マネージャの統合部は2つのワークモジュールからの出力フラグメントを受け取り，関連度の比較からメールツールの出力フラグメントを音声合成エンジンに，入力フラグメントを制御メッセージとしてメールツールは選択したという情報 (5)，エアコン制御には選択しなかったという情報 (6) を送信する．

以上のように，マネージャにより下位のモジュールの応答が統合され，統合部が対話の流れや関連度などを考慮して選択が行なわれる．

4 入力発話に対する関連度の計算

複数のワークモジュールを束ねるマネージャでは，入力に対する各ワークモジュールの関連度の計算が必要となる．そこで関連度計算の一手法を示す．

本節では以下の3つの事柄を手がかりとして関連度計算を行なう．

- ドメインに関連する語句が入力発話中に多く存在する．
- 発話系列中には同じドメインを扱う発話が連続して出現する．
- いままで連続して扱っていたドメインとは異なるドメインを扱う場合，そのドメインを特徴付ける語句が入力発話中に存在する．

語彙知識と特徴語彙知識を生成する．特徴語彙とは，あるドメインを扱う場合，ドメインに対して関連の深い語句のことであり，特徴語彙知識はその集合である．語彙知識はワークモジュールの辞書から特徴単語知識を除いたもので構成される．

関連度は語句の一致数，履歴重み，特徴語彙重みの3つを用いて計算する．履歴重みとは1つ前に選択されたワークモジュールの関連度を大きくする作用があり，特徴語彙重みはドメインの特徴を表す語句が発話文中に存在した場合にそのワークモジュールの関連度を大きく作用を持つ．

A,Bのそれぞれの関連度を R_A, R_B，それぞれの発話文中の語句と語彙知識との一致数を α_A, α_B，特徴語彙との一致数を β_A, β_B，スコアを S_A, S_B，特徴語彙重みを W，履歴重みを h_A, h_B としたとき，以下のように計算する．なお，W は1以上，h_A, h_B は1つ前に選択されているならば1，選択されていないならば1以下の値となる．

$$S_A = (\alpha_A + \beta_A \times W) \times h_A$$

$$S_B = (\alpha_B + \beta_B \times W) \times h_B$$

$$R_A = \frac{S_A}{S_A + S_B}$$

$$R_B = \frac{S_B}{S_A + S_B}$$

関連度が計算された後，その大小をマネージャが比較して発話処理させるワークモジュールを選択する．

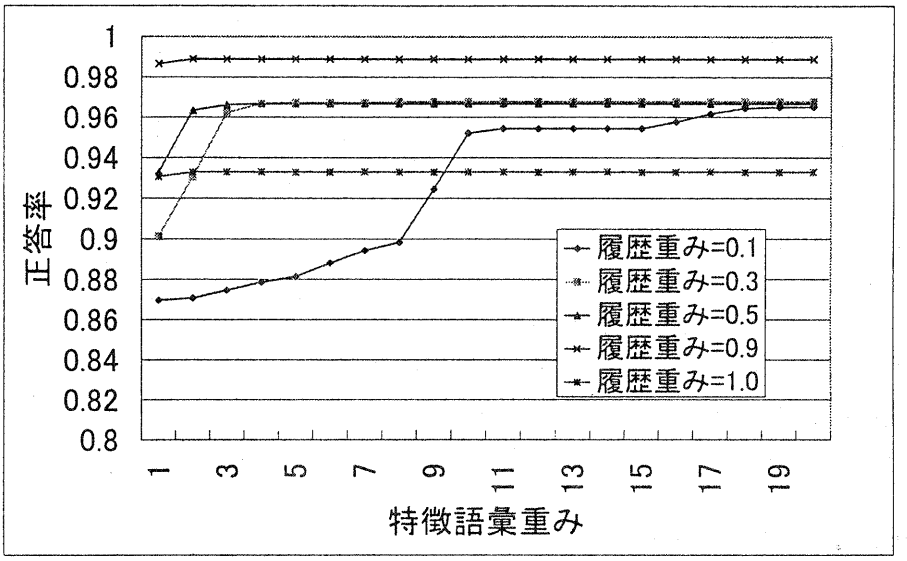

図 4: 関連度計算における正答率実験結果

4.1 ドメイン選択実験

発話入力させドメインを選択する実験を行なった．使用するワークモジュールは MP3 プレーヤとメールツールである．入力文は 1778 文 (7527 単語) を使用し，それらを 18 セットに分割した．入力文は我々の研究室で被験者を用いて収録した MP3 プレーヤを操作する発話と，メールツールを操作する発話の書き起こし文を利用し，それらを混ぜて使用した．

今回は入力文からワークモジュールが正しく選択された割合 (正答率) と履歴重み，特徴語彙重みの関係を調査した．

実験結果を図 4 に示す．それぞれのグラフは履歴重みを表し，横軸が特徴語彙重み，縦軸が正答率である．履歴重み 0.9，特徴語彙重み 2 以上が最も良い結果であり，正解率は 98.9% であった．実験の結果より，入力発話からのワークモジュール選択は，入力発話中に特徴語彙が出現した場合は，特徴語彙を持つワークモジュールを優先的に選択し，特徴語彙一数が同じであるならば，一つ前に選択されたワークモジュールと同じにすれば高い精度で正しい選択を行なうことができるということがわかる．

この方法の問題点は辞書のレベルの違いによる誤りである．ワークモジュールの辞書に登録されている語句の種類や登録数が関連度計算に影響を与えてしまい，正しい選択ができなくなってしまうことが挙げられる．

5 類似ドメインを選択するマネージャの実現

前節の方法は 2 つのドメインの関連がほとんどない場合には有効であると考えられる．しかし，類似したドメインを扱う場合，2 つのワークモジュールの辞書に共に登録されている語句が多くなると予想され，関連度の比較のみでは選択できない場合がある．そこで関連度計算とは異なったアプローチでワークモジュールの選択を行なう．

現在，我々はマネージャが選択することができない場合，ユーザに問い合わせを行ない，そ

の答で選択を行なうシステムを作成している．

本節ではユーザに問い合わせを行なう，MP3プレーヤとラジオを扱うシステムを示す．

5.1 オーディオマネージャ

今回のシステムでは以下の条件を満たすように設計した．扱うワークモジュールはMP3プレーヤとラジオである．

- 一方のワークモジュールのみ動作できる場合は，そのワークモジュールを選択する．
- 入力発話に対し，MP3プレーヤとラジオが共に動作できる場合，ユーザに問い合わせる
- MP3プレーヤとラジオは同時に音を流すことはできない．例えばラジオから音が流れている時にMP3を再生する場合，ラジオを停止しMP3を再生する．

発話が入力されるとマネージャが2つのワークモジュールに入力発話を分配する．それぞれのワークモジュールは入力発話を解析して出力フラグメントをマネージャに返す．出力フラグメントには入力発話に対してどのような動作ができるかという情報が含まれている．マネージャは受け取った2つの結果から条件を満たすように処理を行なう．

以下に動作結果の例を示す．

```
--1-----------------------------
入力発話: 音楽 流して
output{モジュール:MP3 プレーヤ
        処理結果   : 再生}
output{モジュール: ラジオ
        処理結果   : 再生}
応答発話: どちらの音楽を流すのですか
MP3 プレーヤ 状態: 停止中
ラジオ       状態: 停止中
--2-----------------------------
入力発話: ラジオ
control{モジュール:MP3 プレーヤ
         実行       : 不選択}
control{モジュール: ラジオ
         実行       : 選択}
MP3 プレーヤ 状態: 停止中
ラジオ       状態: 動作中
--3-----------------------------
入力発話: 運命 が 聞きたい
output{モジュール:MP3 プレーヤ
        処理結果   : 再生}
output{モジュール: ラジオ
        処理結果   :(null)}
control{モジュール:MP3 プレーヤ
         実行       : 再生}
control{モジュール: ラジオ
         実行       : 停止}
MP3 プレーヤ 状態: 動作中
ラジオ       状態: 停止中
```

1番目の発話はどちらワークモジュールも動作できる発話であるので，2つのワークモジュールは出力フラグメントをマネージャに送信し，マネージャは結果を受けユーザに対して問い合わせを行なう．2番目の発話で問い合わせに対し，ユーザが「ラジオ」と答えたため，マネージャはラジオには実行命令を含む入力フラグメントを，MP3プレーヤには選択しなかったという情報を持つ入力フラグメントを送信する．3番目の発話はMP3プレーヤのみ動作できる発話である．このためラジオの出力フラグメントには処理結果がない．マネージャはMP3プレーヤを選択するが，条件から同時に2つのシステムを動作させることができないので，マネージャはラジオには停止命令を，MP3には再生命令を送信する．その結果，MP3プレーヤは再生し，ラジオは停止する．

6 まとめ

本稿では独立に処理を行なうモジュールを階層的に結合することによって，マルチドメイン音声対話システムを実現する，アーキテクチャを提案した．このアーキテクチャは以下の特徴を持つ．

1. それぞれのワークモジュールは他のワークモジュールやマネージャのことを考慮する必要がない．ワークモジュールは独立に作

成することが可能である．

2. マネージャは下位のモジュールの知識だけを必要とすればよい．また，この知識を用いて適切な下位モジュールを選択することができる．システム全体としてみれば様々なドメインを扱っているが，あたかも単一ドメインの音声対話システムを扱うかのように利用できる．

3. フラグメントを用いてデータのやりとりを行なうので，マネージャやワークモジュールの接続が容易である．このため、システムは柔軟性と拡張性を満たす．

4. すべてのワークモジュールやマネージャは同時に処理することができる．このため提案手法に基づくシステムはスケーラビリティを持つ．

また，入力発話からドメインに関連する語句の一致数から関連度を計算し，その大小からワークモジュールを選択するシステムを作成し実験を行なった．実験の結果，ドメインが類似していないならば，ほとんどの発話から適切なワークモジュールを選択することができることを確認した．

今後の課題として，ドメインが類似している場合のワークモジュールの選択方法が挙げられる．また，関連度計算を行なう方法とユーザに問い合わせる方法の比較をする予定である．

参考文献

[1] Goddey,C.,Brill,E.,Glass,J.,Pao,C.,Phillips,M.,Polifroni,J., Seneff,S. and Zue,V. "GALAXY: A Human Language Interface to Online Travel Information", Proc. ICSLP'94, pp.707-710 (1994).

[2] Nobuo Kawaguchi,Shigeki Matsubara, "An Archtecture for Multi-Domain Spoken Dialog Systems", Proceedings of the 5th Natural Language Processing Paciffic Rim Symposium (NLPRS-99), pp.463-466 (1999).

[3] 松原 茂樹，河口 信夫，外山 勝彦，稲垣 康善，"マルチモーダル図形エディタのための漸進的な話し言葉処理手法"，情報処理学会研究報告，SLP23-2，pp.7-12 (1998).

[4] Matsubara,S.,Yamamoto,H.,Kawaguchi,N., Inagaki,Y.,Toyama,K., "An Interactive Multi-modal Drawing System based on Incremental Interpretation", IJCAI97 Workshop:Intelligent Multimodal Systems, pp.55-62 (1997).

[5] 松永 悟，松原 茂樹，河口 信夫，外山 勝彦，稲垣 康善，"Sync/Mail: 話し言葉の漸進的変換に基づく即時応答インタフェース"，電子情報通信学会技術報告，NLC98-36, pp.33-40 (1998).

[6] Takebayashi,Y.,Tsuboi,H.,Sadamoto,Y., Hashimoto,H. and Shinchi,H., "A Real-time Speech Dialogue System Using Spontaneous Speech Understanding", Proc. ICSLP'92, pp.651-654 (1992).

[7] Zue,V., "Conversational Interface: Advances and Challenges", Proc. EUROSPEECH'97, pp.KN9-KN18 (1997).