3ZB-8

複数人ユーザに対応可能な情報提供会話エージェント

馬場 直哉[1]　黄 宏軒[2]　林 佑樹[3]　中野 有紀子[3]

成蹊大学大学院　理工学研究科[1]　立命館大学　情報理工学部[2]　成蹊大学　理工学部[3]

1.はじめに

ショッピングモールや博物館等の公共施設では，複数人から構成されるグループで情報提供端末を利用することが多い，そこでは，グループ構成員が互いに相談しながら，必要に応じて情報を得るための操作を行うのが一般的である．このようなグループユーザが利用できる情報キオスクとなる会話エージェントを実現するには，多人数会話を遂行する機能を実装することが必要となる．ユーザが一人の場合には，相手の発話は必ずエージェントに向けられていることがわかるが，ユーザが複数人になると，ユーザ同士の会話が増えるため，ターン交代の制御が必須となる．この問題に関して，我々はまず，ユーザ 2 人とエージェントとの 3 人会話において，ユーザの発話が誰に向けられているか，つまり受話者を推定する機構を構築した[1]．本研究では，この受話者推定機構を利用し，複数人ユーザに対応した会話制御が行える情報提供会話エージェントを提案する．

2.複数人対応会話エージェントの必要機能

エージェントが複数人ユーザとの会話を遂行するためには，以下に列挙する機能が不可欠である．

- ターン交代制御機能
- 音声認識・言語理解部
- 会話状態管理部
- 会話制御部

前章で述べたとおり，ユーザが複数人になると，いつエージェントが話すべきかを判断するためのターン交代の制御機構が必須となる．本研究では，発話の話し手(発話者)と発話の受け手(受話者)を推定することにより，適切なターン交代を実現する機構を実現し，複数人のユーザに対応する．次に，ユーザの発話入力に対して音声認識を行い，認識された文字列から言語理解を行う必要がある．さらに，これらのターン情報や言語理解情報を管理するための会話状態管理部，エージェントがユーザに応答しながら会話を進行するための会話制御部が必要である．

3.システムアーキテクチャ

本研究で構築した複数人対応型会話エージェントのシステムアーキテクチャを図 1 に示す．この構成は2章で述べた要件を満たすものである．また，図 2 はシステムのスナップショットである．以下に，本システムの詳細を示す．

Information Kiosk Agents for Multiple Users

1 Naoya Baba: Graduate School of Science and Technology, Seikei University,

2 Hung-Hsuan HUANG: College of Information Science and Engineering, Ritsumeikan University

3 Yuki Hayashi, Yukiko Nakano: Faculty of Science and Technology, Seikei University

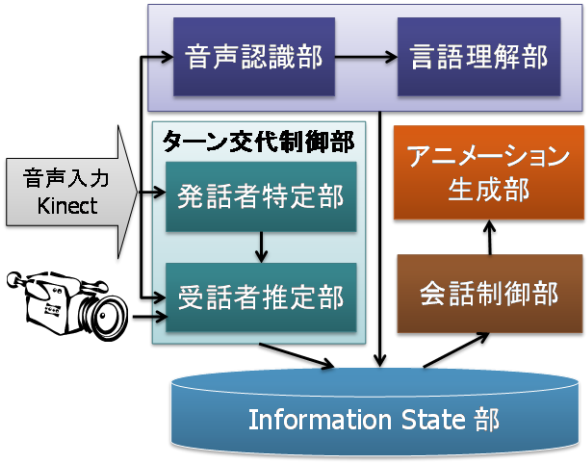


図 1：システムアーキテクチャ

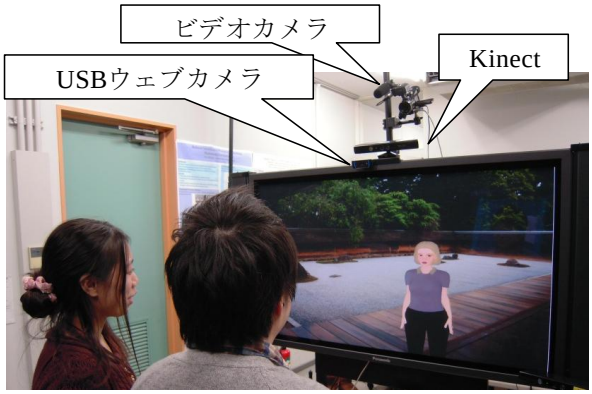


図 2：複数人対応型会話エージェント

3.1. ターン交代制御部

ターン交代制御部では，ユーザがエージェントに対して話しかけている時に，エージェントが確実に応答できるように，発話者と受話者を推定することによって，ターン交代の制御を行う．

3.1.1. 発話者特定部

Microsoft Kinect のマイクロフォンアレイを用い，音源方向を取得し，その方向によって，発話者の特定を行った．図 3 のように，Kinect から垂直方向を 0° とし，音源方向角度がプラスなら右側のユーザ，マイナスなら左側のユーザを発話者とし，受話者推定部に送信する．マ

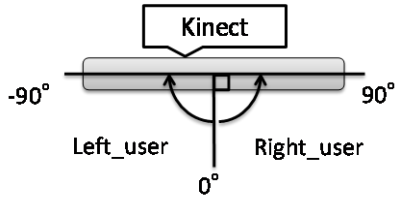


図 3：マイクロフォンアレイによる発話者特定

Copyright ©2013 Information Processing Society of Japan.
All Rights Reserved.

イクロフォンアレイのプログラムは，Kinect for Windows SDK で提供されているものを上記のとおりに改良した．

3.1.2. 受話者推定部

我々の受話者推定方式は，音声と顔向きの情報から受話者を推定しようとするものである[1]．Wizard-of-Oz（WOZ）法により 2 名のユーザと 1 体のエージェントとの会話を収集し，これを分析した結果，エージェントに対しては，高い声で，大きく，ゆっくりと話すということがわかった．また，顔向き情報のみでは受話者を推定するには不十分であることも明らかになった．これらの分析結果より，ピッチとパワーの平均値，話速，発話継続長等の音声情報と，エージェントを見る/相手を見る/その他の方向を見る，の 3 種類の顔向き方向と，これらの遷移パターンの情報を特徴量とする受話者推定モデルを機械学習により構築した．

その結果，80%の分類精度で，受話者がエージェントかもう一人のユーザであるかを推定できるモデルを得ることができた．また，各特徴量をリアルタイムで取得し，受話者推定を行う機構を実装し，評価を行ったところ，隣の人の発話や咳払いなどの不正な音声入力を除くと分類精度が 83%とオフラインでの評価と同等の結果を得ることができた．その後，精度低下の原因となっていた隣の人の発話が本人の発話と認識されてしまう問題については，kinect のマイクロフォンアレイを用いて発話者を特定することにより解決した．

3.2. 音声認識・言語理解部

(1) キーワードの設定

キーワードは WOZ 実験で集めたコーパスに基づき選定した．具体的には，ユーザからエージェントに向けられた発話のみをコーパス中から抽出し，これらの発話を形態素解析し，そこに含まれる名詞・動詞・形容詞・感動詞をキーワードとした．

(2) 入力発話ベクトルの作成

音声認識部では，まず，Google の音声認識システムから 5-best までの発話音声認識結果を得ると，これらを形態素解析し，その中に含まれるキーワードを検出することにより，入力発話をキーワードのベクトルとして表現する．また，ベクトル内の各キーワードの重みは，音声認識結果の尤度により決定した．つまり，1-best の認識結果に含まれる単語がキーワードにマッチしている場合は，これらの単語に 1.0 の重みを与え，順次認識結果の尤度の順位にしたがって重みを 0.1 ずつ小さくする．最後に，最も重みの大きいキーワードの重みが 1.0 になるように正規化する．これにより，認識結果の尤度により重みづけられた入力発話のキーワードベクトルが作成される．

(3) 発話候補ベクトルの作成

上記(1)でキーワード選定に用いられたエージェントに対する発話の集合をユーザ発話候補群とし，本システムではこれらを認識・理解の対象とする．そのために，各発話候補を(1)で選択したキーワードのベクトルとして表現する．また，ベクトル中の各キーワードの重みには全発話候補を文書群として計算した TF*IDF を用いた．

(4) ベクトル空間モデルによる言語理解

言語理解部では，ベクトル空間モデルを利用し，(2)で示した方法で，音声認識部で生成された入力発話のキーワードベクトルと，(3)で作成した各発話候補ベクトルとの類似度を計算し，最も類似度の高い発話候補ベクトルを言語理解結果とする．その結果，図 4 に示すような XML 形式の意味表現が言語理解結果として得られる．この例は，金閣寺について尋ねるという意味表現である．

この言語理解結果は，各変数に代入される値とともに会話状態管理部に送られ，会話の状態が更新される．

```
<UttrUstd uid = “1”>
  <Intention>ask</Intention>
  <Arg>金閣寺</Arg>
<UttrUstd>
```

図 4：言語理解結果の出力例

3.3. 会話制御部

会話制御部では，言語理解部から出力されるユーザの発話意味表現と，受話者推定部から出力される受話者を入力とし，有限状態遷移モデルによって会話の制御を行う．もし，受話者がエージェントと推定されると，認識されたユーザ発話の意味表現に応じて状態が遷移し，遷移した状態に基づきエージェント発話が決定され，出力すべき発話の文字列がアニメーション生成部に送信される．

本研究で作成したシナリオは，京都の観光スポット 14 ヶ所のうち，3 ヶ所を自由に回ることができる旅行クーポンを入手したという設定で，旅行代理店のスタッフとなったエージェントから情報を得ることにより，訪問する場所を 2 人で話し合って決めるというものである．状態遷移はユーザが観光スポットを 3 ヶ所決めることができたら，終了状態になるよう設計した．

3.4. 会話状態管理部

本研究では，Information State の考え方に基づき，会話状態管理部を設計した．具体的には，発話者・受話者・言語理解結果・会話の状態遷移モデル中の状態名などの管理を行い，発話が終了する度に情報の更新を行う．

3.5. アニメーション生成部

アニメーション生成部では，Visage-SDK[2]を利用し，CG キャラクタのアニメーションを生成している．これは，会話制御部によって決定された文字列やキャラクターの非言語動作や場面の切り替えを指示するコマンドメッセージを受け取ると，リアルタイムでそれらを再生する．

4. おわりに

本研究では，複数人対応型会話エージェントを構築するために，受話者推定によるターン制御を行い，2 人のユーザとエージェントとの会話を制御する多人数対応型会話エージェントを実装した．今後は，本システムに対し評価実験を行うことで，各モジュールの性能と会話エージェントシステムの有用性を検証する．

参考文献

[1] 馬場直哉，黄宏軒，中野有紀子：「人対会話エージェントにおける頭部方向と音声情報を用いた受話者推定機構」，人工知能学会論文誌，Vol.28, No.2.

[2] Visage-SDK, Visage Technologies AB, http://www.visagetechnologies.com.

Copyright ©2013 Information Processing Society of Japan.
All Rights Reserved.