

複数話者による対話システム

松坂要佐 東條剛史 久保田千太郎 田宮大介 古川賢司 早田啓介 中野裕一郎 小林哲則
早稲田大学 理工学部

1 はじめに

従来の人・機械の対話システムにおいては、人と機械が1対1で対面して対話することが前提とされてきた。しかし、通常我々は同一空間内で複数人を相手に多対多の対話を行なっている。

将来、人を補助するロボットが我々の生活空間に入ってくるのが期待されているが、それら人の生活空間において活動するロボットのインターフェースとしては、複数人のユーザを同時に相手にできるものが望ましい。

本稿においては、複数人との対話を行なうにあたって、対話における対話参加者の役割とその関係パターンを分析するとともに、分析を元に構築した画像入力機能を有する人型ロボットを利用することにより複数人を相手に対話を行なうことができる対話システムを紹介する。

2 対話における対話参加者の役割とその関係パターン

まず、複数人での対話の定義であるが、一つの話題を複数人で共有している状態を指すものとする。たとえば複数人が距離的に近くに寄って言葉を交わしていても、あるグループが別のグループが共有している話題とは違う話題を取り上げていたら、そこには2つの複数人での対話が存在することになる。

複数人での対話においては、複数人が無秩序に発言したのでは対話が進行しないため、対話をリードする役割をする人物が常に一人存在している。対話をリードしている人物は固定されているわけではなく、局面や話題の変化に応じて入れ替わっていく。

対話をリードする役割を負っている人物は、その視線や顔の向いている方向によって、その他の対話参加者に対して発言権を与える。その他の対話参加者はリードしている人物に対して声をかけ、振り向かせることにより、自ら発言権を取得する場合もある。(それでも無視され、発言権の取得に失敗した話者が、その対話

グループから分化して、自らがリーダーとなって新たな対話グループを形成する場合もある。)

3 対話システム概要

3.1 システム構成

複数人を相手とした対話システムを構築するにあたり、対話をリードしている人物の発見の必要性から画像情報を取得する機能が不可欠であった。また、自らが対話のリーダーとなった場合は、複数の対話参加者に対して発言権を与える／与えないをアピールする能力が必要となるため、人に模した眼球・頭を持つことにより人に対して視線を投げかけることが可能なロボットを使うことによって、人と同等の発言権の付与能力をシステムに与えた。

3.2 対話におけるリーダーの発見

対話をリードしている人物は、その他の対話参加者の視線の集中している方向をたどっていくことで見つけることができる。本システムでは、対話グループの中でのリーダーの発見のために、対話参加者の顔向き認識をおこなう。ロボット頭部に設置したカメラからの取り込み画像より、統計的手法によって肌色尤度を求めることで肌色領域(顔に対応する)を抽出し[2]、得られた肌色領域画像を顔向きデータベースを使って作っておいた固有空間に投射することにより、領域がどの顔向きカテゴリに属するか判別する。

対話参加者からの顔向き以外の意志表現として、指さし動作・挙手の動作も認識できるようにした。指さし動作・挙手の動作の認識としては、取り込み画像からエッジを抽出し、肌色領域(顔および手に対応する)の座標と、顔・手位置と肘位置の関係を収集した姿勢データベースをもとに、画像中の人物の肩位置、肘位置を探索することにより指し示した方向を得る。

システムは、対話をリードしている人物を発見したら、ロボットにその人物を正視させることにより、自らが対話に参加していることをアピールする。

3.3 自らがリーダーとなった場合

対話参加者の視線(顔の向き)がロボット自身に集中していた場合、ロボット自身が対話のリーダーとなる。システムはロボットに視線を投げかけさせることにより、対話参加者に対して発言権の付与を行なう。対話

〒169 東京都新宿区大久保 3-4-1
Tel:03-3209-3211(内線 3174) Fax:03-3205-9381
E-mail: yosuke@tk.elec.waseda.ac.jp

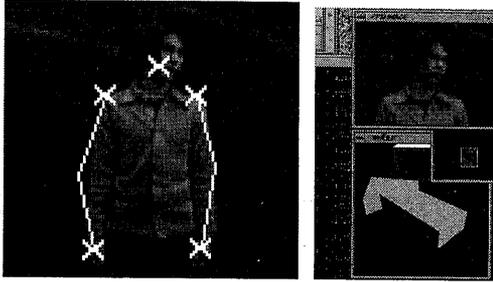


図 1: 対話参加者の顔の向きと姿勢の解析

参加者はロボットに正視されることで、自分に発言権が与えられたことを実感できる。

また、対話者参加者からの呼びかけによる発言権の取得にも対応できるように、音源の定位も行なう。顔の両脇に設置した2つのマイクから入力された2系統の音声の高域の減衰特性の比較によって、音源の方向を推定し、その方向にいる人を画像処理によって探す。

システムは、現在対話中の人物との対話が途切れるまで発言権を与え続けるが、その間にその他の話者が呼びかけによって発言権を取得しようとした場合には、目の自由度と首の自由度を使い分けた視線の投げかけをおこなうことにより、呼びかけには応じないことを意志表示する。たとえば、顔は他の人を向いたまま目だけを対話者に向け「待ってください」などと発話することで、その対話者を無視しているわけではないが、現在の意識は他の対話者に向いていることを表現できる。目と首の自由度を使い分けた対話進行の例を図2に示す。

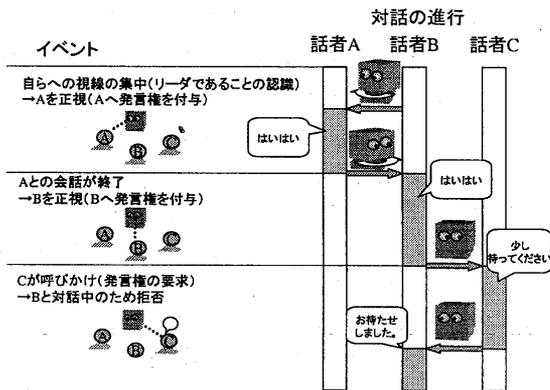


図 2: 顔向きにより発言権の付与を行なった対話進行

音声認識には語彙数約七百の連続音声認識を行う [3]。入力された文が、システムに対する質問文であった時は言語処理により、あらかじめ用意しておいた質問意図 (約 50 種類) に分類して返答文を出力する。システムに対する質問文以外の文には「はいはい」という発話を行ない次の発言を促す。

3.4 実験

上記のシステムを用いて、対話者参加者数を2名 (ロボットを含めて3名) に限定した簡単な対話実験を行なった (図3)。

タスクとしては、ロボットの持つ機能についての一問一答形式の自己紹介タスクを設定した。各対話者はロボットの前にならんで座り、任意の対話参加者を相手に発言を行なう。実験は比較的静かな研究室で行ない、暗幕を背景とした。

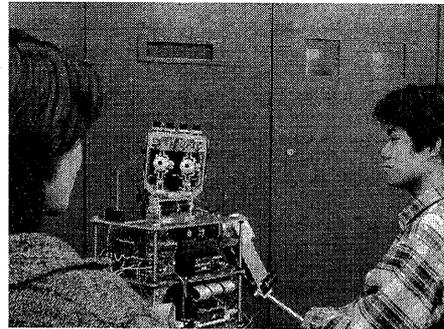


図 3: 対話タスク実行風景

4 むすび

複数対話者のなかで、視線を投げかけ合いながら一問一答の対話を行なう対話システムを構築した。

対話参加者は複数対話者の中でも違和感なく発話を行なうことができ、ロボットの身体表現により対話の進行をスムーズに制御することができることを確認した。

参考文献

- [1] 肥田木康明, 小林哲則 他 “アイコンタクト機能を有する複数ユーザとの対話ロボット”, 音声言語情報処理 17-1, pp.1-6, (1997)
- [2] 春山智, 小林哲則 “動画像処理による手振り動作認識”, 信学総大'97, D-12-55, (1997)
- [3] 小林紀彦, 小林哲則 他 “統計的言語モデルにおける高頻度形態素連鎖語の辞書登録による効果”, 日本音響学会 平成 10 年度春季研究発表会, pp.29-30, (1997)