

F-030

複数の話者と対話可能なロボットの試作

A trial production of the robot which can have a dialog with some speakers

田村 雅人[†] 金丸 浩司[†] 鈴木 良弥[†] 西崎 博光[†] 関口 芳廣[†]

Masato Tamura Koji Kanemaru Yoshimi Suzuki Hiromitsu Nishizaki Yoshihiro Sekiguchi

1. まえがき

本稿では、音声認識と自然言語処理技術を利用して、複数話者と対話できるロボットの作成を試みている。このロボットは簡単なアルゴリズムで人間二人と対話でき、話者方向に向く動作もできるようにしている。試作したロボットの各モジュールについて説明し、各モジュールの性能評価についても述べる。

2. ロボットの概要

ロボットは、人間二人が「間違い探しのゲーム」をしている傍らにいて、二人の対話を理解する。対話理解のために二人の対話に割り込むことができ、話者の方向に向くこともできる。また、間違い箇所を認識し人間に確認することができる。

このようなロボットを作成するためには、1m程度離れた話者の音声認識ができること、話者の識別ができること、話者方向の識別ができること、重なった音声の識別ができること、対話内容が理解できること、適切な応答音声を返せること等、様々な問題を解決する必要がある。これらの問題に対処する基本的な技術は既に存在するが、本稿ではできるだけ簡単なアルゴリズムでロボットを実現することを試みた。

3. 各部の説明

ロボットの試作にあたっては、部分的な修正が必要な場合が多いので、基本的にはブラックボードモデルを採用する。この方法の欠点は、システム全体の同期をとることが難しいことである。速度は若干遅くなるが、各処理の動作状況を示す指標を用意して、この問題を解決している。各処理を以下のように分けて説明する。

3.1 音声の入力

音声入力は、15cm間隔に無指向性マイクロホン3個配置した、ラインマイクロホンアレーで行なう。音声の標準化は16kHz、量子化は12ビットである。

3.2 話者方向の識別

話者方向の識別は、まず左右の端にある2つのマイクロホンに入った音声の相関をとり、音声の到達時間差を求め、その時間差から話者方向を識別する。ここでは、ロボットの前方に話者がいると仮定して処理をする。

1名の話者に対して行った話者方向識別実験では、1.5mの距離で約90%の識別率を得ている。但しここで得られる角度は、ロボットの頭部を話者方向に向ける等の動作の情報である。よってその話者方向がわかればよいので、 ± 10 度以内に入ると正解としている。

3.3 話者の識別

ロボットと対話をする人の情報は、予めロボットに登録しておくことにする。話者識別の方法は基本的には文

献[1]の方法を応用する。パラメータとして、メルケプストラム、メルケプストラム、ピッチ、ピッチを使う。このシステムでは、

1. 学習が容易なこと
2. 非常に少ない音声データ量で識別が可能なこと

が要求される。そこで学習には挨拶で使用される音声、「おはよう」、「こんにちは」など6単語を使用する。話者識別には1発話が使われるが、できるだけ多くのデータが含まれる「発話全体」と、パラメータの抽出等が安定な「母音部」のみを使った話者識別を行い、その順位を利用して最終的な話者識別を行っている。

7名の話者を登録して行った予備実験では、約94%の話者識別率であった。

3.4 音声の認識

音声認識にはJulius[2]を使用している。話者とロボットの距離が1m程度離れるとSN比の低下などで、接話型マイクロホン利用の場合より音声認識率が低下する傾向が強い。そこで、前述のマイクロホンアレーで得られるデータの相関を取り、音声の到達時間差を求め、遅延時間差を考慮して各マイクロホンの音声の和をとる。これで話者方向に指向性があるマイクロホンになる。この音声を認識対象にすることにより、認識率の向上を計っている。1名の話者が、ロボットから1m離れたところで発音した音声に対して、マイクロホンアレーを使用しない場合と、使用した場合の音声認識率の比較を以下に示す。

- マイクロホンアレーを使用しない場合 30%
- マイクロホンアレーを使用した場合 56%

言語モデル、音響モデル共十分な対処をしていないので、全体の認識率は低いが、マイクロホンアレーの使用により音声認識率の向上が期待できる。

音声重複した場合の処理は、現在話者方向の識別結果を利用して音声を分け認識を行っているが、満足がいく結果を得るためにはさらに工夫が必要だと思われる。

3.5 対話の理解

ユーザの発話を理解して、必要に応じて応答を返すようにする。理解内容の蓄積や対話を円滑にすすめるための割り込み発話なども行う。

3.5.1 意味理解

話者の発話内容を理解するために、音声認識結果を形態素解析する。形態素解析にはJUMAN[3]を用いる。形態素解析の結果より、格文法を利用して発話内容を格構造で表現する。格構造の例として、表1に動詞「ある」の格構造の一部を示す。

3.5.2 絵の内容の蓄積

発話者毎の絵の内容を蓄積するのに、フレーム構造を利用している。例として、「右にある赤い太陽はおおき

[†]山梨大学大学院医学工学総合教育部，甲府市

い」という発話から得られる絵の内容は、発話者が“田村”とすると、表2のようになる。

表 1: 格構造の例「ある」

格の種類	意味素性と助詞
tai	+もの(が, は)
frT	+時(から)
toT	+時(まで)
atL	+場所(に)

表 2: 絵の内容の蓄積例

話者	田村
対象物	太陽
修飾	赤い, 大きい
場所	右

発話の中に対象物が存在しなかった時、または指示詞が発話中に存在した場合は、前の発話中にある対象物を指していると考え、このときの処理は下記のように分類して行う。

< 質問の場合 >

例: 「それは右ですか」, 「大きいですか」

この場合は、前の発話の対象物のことだと理解する。

< 前発話を部分的に言い換える場合 >

例: 「右にいます」

この場合は、前の発話の部分肯定とみなす。前の発話の適当な部分を絵の情報として採用する。

< 前発話を全面的に認める場合 >

例: 「はい」, 「あります」

この場合は、前の発話の全肯定とみなし、前の発話の内容を全て絵の情報として採用する。

3.5.3 割り込み発話

対話を円滑に進行させるためには、ロボットからの「割り込み発話」も重要な要素である。割り込みの内容やタイミングは、非常に難しい問題で今後詳しい検討が必要だと思うが、ここでは実験的に決めた基準で「割り込み発話」を行うことにする。

< ロボットが間違いを発見した時 >

例: 「雲の色が違いますよ」

< 片方の話者が一方的に話している時、相手に対して >

例: 「金丸君、質問をどうぞ」

< ランダムに、下記の割り込み発話をする >

例: 相槌「ええ」

: 冗談「すごいでしょ!」

: 情報不足を補う「金丸君、太陽の色は何ですか」

: これまで獲得した絵の情報の確認「すみません、確認します。田村君の絵は...」

「割り込み発話」の頻度は外部から設定できるが、これまでの経験から最後の「絵の情報の確認」の頻度を1番高くしてある。

音声認識は誤る可能性があるため、対話理解部分の評価のために、文字入力を実験を行った。簡単な絵を対象にしてはいるが、入力に誤りがない場合には適切な対話処理ができた。

4. 頭部の駆動

頭部の駆動は下記の2種類の処理時に行われる。

- 音声入力があると、ロボットは音源方向に頭を向ける。頭部の駆動と音声の入力を並列に処理する必要があり、外付けのモータ制御回路を用意している。
- ロボットは2人が対話中等にしばしば割り込みを行う。割り込みを行うときに、ロボットは頭部を対象者方向に向ける。そのため、ロボットは話者の位置を常に記憶しておく。この例のタスクでは話者はほとんど移動しないのでこの方法で対処できるが、話者が移動するようなタスクでは、画像を利用するなど別の手法を工夫する必要がある。

5. 評価

ロボットの総合的な評価段階には至っていないが、仮にロボットを構成して各処理部分の性能を調べてみた。

1. まず男性話者3名で、それぞれロボットの左右45度から「こんにちは」等の簡単な挨拶や「太陽はありますか?」等の発話を10回おこなう。話者とロボットの距離は1mである。1回ずつ確認しながらゆっくり対話した場合と、認識誤りがあってもかまわず連続発声した場合の結果を表3に示す。なお登録話者は7名で、6単語で話者識別学習を行っている。

表 3: ロボットの音響処理部性能評価 [%]

	ゆっくり対話	連続対話
音声認識率	68.3	41.7
話者識別率	50.0	35.0
話者方向識別率	90.0	70.0

2. 上記とは別の4名に2名ずつ2組で「間違い探し」のゲームをもらい、ロボットがその対話に加わった。人間はロボットの左右45度、1mの所にいる。ゲームが終わった後、5段階評価(5:良い, 4:普通, 3:やや悪い, 2:悪い, 1:使えない)を行ってもらった。4名の平均評価は、下記のものであった。処理速度: 3.2 返答の適切さ: 1.8 割り込みの適切さ: 3.0 対話の円滑さ: 1.5

6. おわりに

簡単なアルゴリズムで、複数話者と対話可能なロボットの作成を試みたが、まだ音声認識率をはじめ解決すべき問題が多く性能は低い。今後は様々な問題点を解決し、ロボットの性能向上を図る予定である。

参考文献

- [1] 松井 知子, 古井 貞熙: “音源・声道特徴を用いたテキスト独立形話者認識”, 電子情報通信学会論文誌 Vol.J75-A No.4 pp.703-709(1992年4月)
- [2] “汎用大語彙音声認識エンジン Julius/Julian”, <http://julius.sourceforge.jp/>
- [3] 京都大学言語メディア研究室: “日本語形態素解析システム JUMAN version3.61”, <http://www-lab25.kuee.kyoto-u.ac.jp/nl-resource/juman.html>