

HARK を用いたロボットクイズ司会者 HATTACK25 の開発

○西牟田勇哉 平山直樹 大塚琢馬 杉山治 糸山克寿 奥乃博 (京都大学 情報学研究科)

1. はじめに

近年の科学技術の発展により、人との対話を目的としたロボットが開発されている。このような対話ロボットの例として、AIBO [1] や PaPeRo [2], PALRO [3] などがある。これらの対話ロボットの目的は、人の作業を補助する、代わりに行うといった実用的な機能ではなく、コミュニケーションを通じて人を和ませる、楽しませることであり、このような目的を持つロボットは総じてエンタテインメントロボットと呼ばれている。エンタテインメントロボットは高齢者施設などにも導入されており、様々な方法でコミュニケーションの場を提供している。そのなかで代表的なものにクイズゲームがある。クイズゲームは人々にエンタテインメントを提供するコミュニケーションの一つであり、日常的なイベントやテレビ番組でも数多く取り上げられている。一般的なクイズゲームには、ゲームの司会を務めるクイズ MC (Master of Ceremonies) が不可欠である。クイズ MC は司会としてゲームの進行や時間管理を行うだけでなく、ゲームに参加している人々を楽しませる役割をもつが、そのような腕のよい司会者を常に確保することは難しい。

エンタテインメントロボットの研究としては、Dohsaka らの研究 [4] や藤江らの研究 [5] がある。これらの研究では、全員で一つの正解を目指す形式のクイズゲームを取り上げ、そこにロボットが介在することでゲーム参加者自身や、そのコミュニケーションの場が盛り上がることを主張している。しかしこれらは、多人数コミュニケーションの重要な要素である対話者の識別は行っていない。一方、回答者としての強い人工知能を目指す研究としては、クイズゲームに回答する人工知能 Watson [6] などがある。これらの研究は問題に素早く正確に回答することを目的としており、司会者や回答者とのインタラクションによるエンタテインメント性の向上を目指すものではない。

我々は、ゲームの参加者や視聴者を楽しませるためには、MC にも回答者に匹敵する強い AI が必要であると考え、状況に応じた適切な会話は機械的な司会進行よりも場を盛り上げることに貢献し、状況を正しく把握するためには MC もまた問題を正しく理解する必要があるためである。そこで本研究では、多人数対戦クイズゲームの“HATTACK25”を設定し、児玉清氏のような Intelligent な MC ロボットの開発を目指す (図 1, 2)。HATTACK25 は「パネルクイズアタック 25」(朝日放送) をもとにした、音声対話を用いたクイズゲームである。

ここで、既存の対話システム開発環境に、CSLU Toolkit [7] がある。これは、GUI 上で音声認識、音声合成などの基本要素を組み合わせることで容易に対話システムを設計できるツールである。しかし、このツールは 1 対 1 の対話システムしか構築できないうえ、既

存の機能を拡張することが難しい。一方、本研究で取り上げる“HATTACK25”は多人数が参加するうえ、状況に応じて入出力を柔軟に切り替える必要がある。そのため、この開発環境を本研究に利用することは難しい。本研究では、ロボットはマイクロホンアレイを装着しており、ロボット聴覚ソフトウェア HARK (Honda Research Institute Japan Audition for Robots with Kyoto University) [8] を用いて音源の定位と分離を行い、その結果を利用して MC を務める。

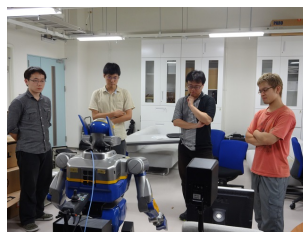


図 1 クイズゲームの司会者をするロボット



図 2 クイズのパネル

2. HATTACK25

2.1 概要

日本の代表的なクイズ番組に「パネルクイズ アタック 25」(朝日放送) がある。アタック 25 は 1975 年から続く日本のクイズ番組として最長寿番組である。従って本研究では、クイズゲームの MC に必要なタスクを洗い出すためにアタック 25 をケーススタディとして採用した。

本研究では、アタック 25 を音環境で再現した HATTACK25¹ を実装した。HATTACK25 は基本的に「アタック 25」と同じであるが、以下の点が異なる。

- 問題文読み上げによる一問一答もしくは選択方式のクイズのみを扱う。映像や音楽を用いたクイズは扱わない。
- 問題の読み上げは MC ロボットが行う。
- 問題に回答するときは、「はい」と発話する。早押しボタンは用いない

ゲームは 4 人でプレイする。ディスプレイ上に 1 から 25 の数字が 5 × 5 の格子状に並んでおり、プレイヤーはこのパネルをクイズによって取りあう。最終的にパネルを最も多く獲得したプレイヤーが優勝となる。また、その進行は基本的にクイズの出題、回答、パネル選択の繰り返しであり、これがいずれかの終了条件

1. 25 枚のパネルがすべて埋まった
2. 用意した問題をすべて出題した

¹HARK を用いた ATTACK25

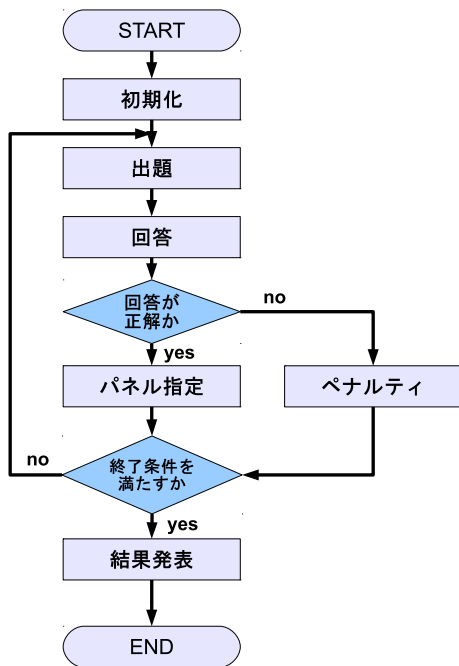


図3 HATTACK25 のフローチャート

を満たすまで繰り返される。回答を間違えた場合などの分岐が発生する場合もある。またゲーム開始に先立ち、ロボットが各プレイヤーの位置を確認するための前処理がある。図3にゲームのフローチャートを示す。

2.2 MC 視点で見たゲームの流れ

MC から見たゲームの流れは以下の通りである。

出題

クイズをランダムに選択して出題する。出題されるクイズとその正解はテキストの形式で保存されており、出題時は音声合成によってそのテキストを読み上げる。

回答

プレイヤーから合図があった場合、そのプレイヤーに解答権を与えて回答を促す。クイズの回答が正解の場合には、続いてパネルの選択に移る。不正解の場合は正解の発表が行われる。また、回答者はペナルティとして、続く2問の回答権を失う。(お立ち状態)

パネル選択

正解したプレイヤーは続いてパネルの選択を行う。パネルはオセロの要領で選択する。

2.3 タスク設定

上記の HATTACK25 の MC をロボットで構築するにあたって、そのタスクを明らかにする。

ゲームのタスクは、(1) 複数のプレイヤーから発話者を決定する、(2) クイズの正解・不正解を判断する、(3) 各プレイヤーのパネル枚数を記録する、(4) 状況に応じて適切な行動を行う、の4つである。

(1) は、ある状況で誰に発話権を与えるのかということを目指す。このタスクが必要とされる場面の例とし

て、クイズ回答者の決定がある。今回の問題への回答は、複数人が同時に合図を行うことが有り得る。そのため、ロボットが複数人の同時発話を聞き分ける必要がある。また、発話権を持っていないプレイヤーの発言についてはロボットが誤って受け付けないようにしなければならない。

(2) は、プレイヤーの話し方の違いによって、クイズの回答が複数の表現でなされるために必要となる。回答に複数の候補がある場合、単純一致では判断ができないため、判断に工夫が必要となる。

(3) は、HATTACK25 において各プレイヤーの得点を比較するために必要である。ロボットが多人数の得点を記録するためには、得点とプレイヤーの対応関係を把握しなければならない。

(4) は、ゲームの進行を円滑に進めるためだけでなく、ゲーム参加者を楽しませるために必要となる。ここでの行動とは、ゲームの進行に必要な発話だけでなく、プレイヤーとの受け答え、状況に応じて BGM を流すといったことを指す。

3. システムの設計

2. 章で解析した MC の役割にもとづいて、HATTACK25 の司会を行う司会ロボットを実装した。本章では、上述の MC ロボットの設計を、ハードウェア、ソフトウェアの両面から詳細に述べる。

3.1 ハードウェアの構成

本研究では、HATTACK25 の司会ロボットを、HRP-2 を用いて実装した。HRP-2[9] は、人の上半身を模したヒューマノイドロボットであり、頭部には、8ch のマイクロホンアレイを搭載している。外部には合成音声を出力するためのスピーカー、パネルを表示するためのディスプレイが設置されている。

3.2 ソフトウェアの構成

本研究で提案する MC ロボットは入力音響の定位、分離に HARK、認識に大語彙連続音声認識エンジン Julius² を用いている。この MC ロボットはプレイヤーの発話を入力として受け付け、合成音声を出力し、ゲームの状況に応じてパネルディスプレイを変化させる。そのソフトウェアの構成を図4に示す。

本システムにおけるシステムの流れは図5の状態遷移に従う。まず、ゲームを開始する前にプレイヤーの位置情報を登録するために必要な初期化を行う。各プレイヤーは図6のようにロボットの前方に 30° 以上(図では 45 度) 離れて立つ。続いてロボットの位置確認に対して返事し、その返事の定位結果をプレイヤーそれぞれの位置情報として登録する。初期化が終了するとクイズを出題する。プレイヤーから合図があった場合、音源定位を行い、その位置情報から回答権を与えるプレイヤーを決定する。そして回答を受け付け、その結果を判定し、正解であればパネルの選択に移り、不正解であれば回答者にペナルティを与えた後、次の問題を出題する。正解したプレイヤーはパネル番号をコールし、番号の選択によって再選択を促したり、パネル

²<http://julius.sourceforge.jp/>

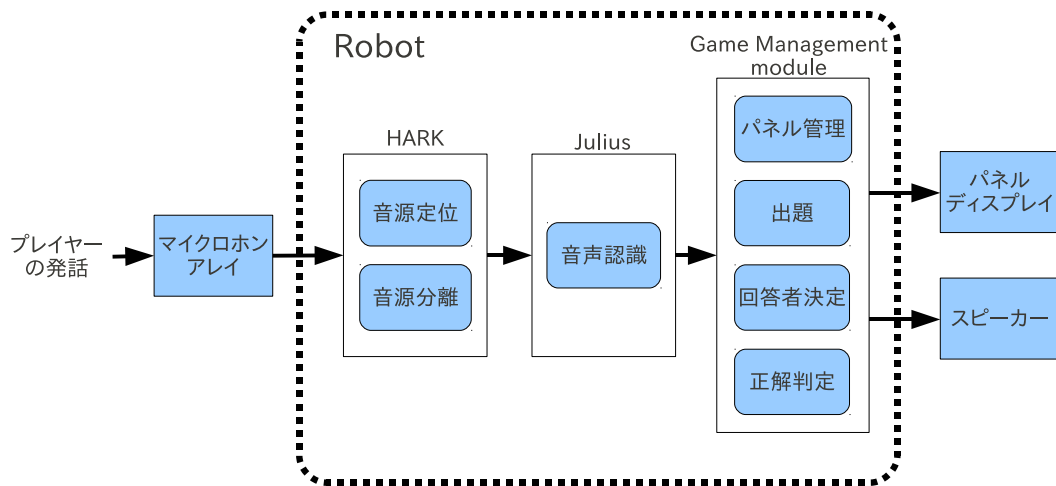


図4 システムの機能ブロック図

の色を変更する。これらを繰り返し、終了条件を満たせば結果発表を行い、ゲーム終了をアナウンスする。

3.3 課題と解決手法

MC ロボットを実装する上での主な課題は、話者の判定とそれ以外の雑音を区別し、雑音を除去すること、そして誤認識の抑制である。次節より、それぞれの詳細について述べる。

3.3.1 話者の判定

第一の課題は話者の判定である。本ゲームは対戦形式であるため、

- 複数の合図から一番早く言ったプレイヤーを判定しなければならない。
- クイズの解答やパネル番号の指定など、特定のプレイヤーが発話する場面で他のプレイヤーの発言に誤って反応しないようにしなければならない。

というように、常にどのプレイヤーが発言しているのかを判定する必要がある。本研究では、以下の方法によって発話者の判定を行った。

1. 3.2節で述べた初期化によって各プレイヤーの位置情報 ϕ_i ($1 \leq i \leq 4$) を求める。
2. 発話の定位結果 ϕ と ϕ_i が式 (1) の関係を満たすとき、そのプレイヤーが発話しているものとする。

$$|\phi - \phi_i| \leq 15^\circ \quad (1)$$

この結果を用いることで、出題時に合図を出したプレイヤーに回答権を与えたり、クイズの回答やパネルの指定に他のプレイヤーの発言が影響しないように除外するといったことが可能になる。

3.3.2 雑音の棄却

第二の課題は、発話以外の雑音の棄却である。HAT-TACK25 を実環境でプレイすると、周りの環境音やロボット自身のモーター音、笑い声や独り言などからも何らかの単語を認識してしまい、ゲームの進行において誤作動を引き起こす。本研究では音節タイプライタを利用することで、そのような雑音や本来の目的な

い発話による認識結果を棄却する。音節タイプライタとは、音節構造の知識のみを反映した文法であり、あらゆる候補仮説の尤度の上限を求める。その音節タイプライタの文法と、目的の記述文法と並行して認識を行わせる。その際の尤度の比が一定以上のしきい値より大きい小さいかで、発話として受理するか、そうでないかの棄却の判断をする。このしきい値は経験的に決められている。

3.3.3 誤認識の抑制

本研究では、誤認識による誤作動を抑制するために、音声認識をする際に以下のような工夫を行った。

1. 認識結果管理キュー

認識結果は、ゲームの進行とは別スレッドで常に受け付けられており、ゲームの状態にかかわらず送られてくる。そのため発話を受け付けていないときに、送られた結果がゲームで使用されないようにする必要がある。例えば、問題読み上げ前に誰かが「はい」と言った場合、遅延により読み上げ開始直後に「はい」が受理されてその人が回答権を獲得すると、ゲームの進行に不都合を生じる。そこで、認識結果管理のためのキューを用意した。このキューには認識結果が一時保存され、必要な時に取り出される。また、このキューは発話を受け付けるタイミングでクリアされ、それ以前の認識結果が発話の受け付けに影響しないようになっている。

2. 言語モデルの切り替え

今回の音声認識には、自分で記述した文法モデルを言語モデルとして使用している。ゲームの状況によって求められる発話は異なるため、ベースとなる3種類の記述文法を作成して、それらを状況に応じて切り替えながら用いることで想定外の発話が認識される確率を小さくしている。ベースの記述文法は以下の通りである。

(1) はいモデル

早言いの「はい」を認識するモデル

(2) 回答モデル

問題の回答候補を認識するモデル

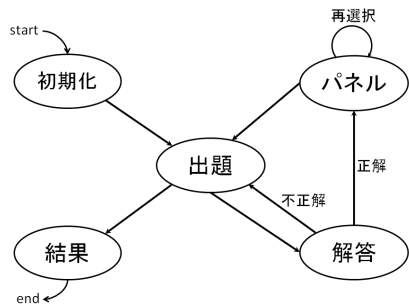


図 5 対話管理状態遷移図

MC : 次の問題, 4人です.
ブラジルの首都はどこでしょう.

PL : はい.

MC : 赤.

PL : ブラジリア.

MC : その通り, ブラジルだ.
さあ, 赤のかた, 何番.

PL : 15番.

MC : 15, 14, 13と赤に変わった.
(MC:MCロボット, PL:プレイヤー)

図 7 対話例

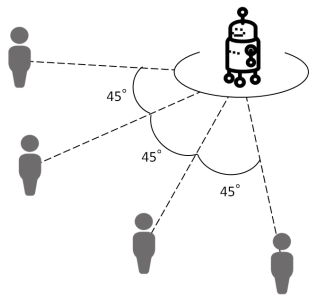


図 6 プレイヤーの位置関係

(3) 番号モデル

パネルの番号を認識するモデル

はいモデルや番号モデルは、発話される可能性のある語彙は限られている。しかし、クイズに対する回答の可能性は無限にあり、それらを網羅することは難しい。そのため、回答モデルでは各クイズの正解および可能性のある誤答を回答の候補として語彙を制限している。例えば「ブラジルの首都はどこでしょう」というクイズに対しては、正解の「ブラジリア」及び誤答としてあり得そうな「サンパウロ」、「リオデジャネイロ」などを候補に加えている。また、上記に加えて回答語彙の幅を制限するため、選択形式の問題を多く用意している。

この章では、最初に MC ロボットの構成について説明し、続いて実装する上での課題として話者の判定、雑音の除去、誤認識の抑制を取り上げ、それぞれの解決手法を示した。これらの問題を解決することで、実環境でのロボットの MC を達成した。その一方で、話者の判定については話者の動きに頑健でない、語彙の制限がプレイヤーの発話を制限している、などが原因で、プレイヤーの自由な行動、発話が阻害されるといった問題が起こっており、今後の課題として考えられる。

4. 実行例

HATTACK25 を実際にプレイした際に、ゲーム中になされた対話例を図 7 に示す。この対話は、ロボットの問題発表からプレイヤーの回答、パネル選択までの一連の流れを示している。

5. まとめ

本稿では、多人数で対戦するクイズゲームの MC をロボットで行うシステムを提案した。プレイヤーの識別を HARK による音源定位結果を用いることで達成した。また、音声認識の誤認識を音節タイプライタなどによって抑制した。

現在、ロボットはあらかじめ用意した応答を読み上げるだけである。今後は、ゲームの場を盛り上げるような気の利いた応答が出来るようにシステムを改良する。

謝辞

本研究の一部は科研費 24220006 の補助を受けた。

参考文献

[1] 藤田雅博: “Robot Entertainment System AIBO の開発 (< 特集 > 「実世界に近づくインタフェース技術」)”, 『情報処理』, vol.41, no.2, pp.146-150, 2000, 情報処理学会.

[2] 藤田善弘: “人工知能の現在と将来 パーソナルロボット PaPeRo の開発”, 『計測と制御』, vol.42, no.6, pp.521-526, 2003, 計測自動制御学会.

[3] “小型ヒューマノイド・ロボット PALRO”, <http://www.fsi.co.jp/company/news/100201.html>

[4] K. Dohsaka et al.: “Effects of Conversational Agents on Human Communication in Thought-Evoking Multi-Party Dialogues”, in Proceedings of SIGDIAL, pp.217-224, 2009.

[5] 藤江真也ほか: “人同士のコミュニケーションに参加し活性化する会話ロボット (対話生成,< 特集 > 人とエージェントのインタラクション論文)”, 信学論 (A), 基礎・境界, vol.95, no.1, pp.37-45, 2012.

[6] “IBM Watson: Ushering in a new era of computing”, <http://www-03.ibm.com/innovation/us/watson/index.shtml>

[7] S. Sutton et al.: “Universal speech tools: the CSLU toolkit.”, in Proceedings of ICSLP, pp.3221-3224, 1998.

[8] K. Nakadai et al.: “Design and Implementation of Robot Audition System ‘HARK’ — Open Source Software for Listening to Three Simultaneous Speakers”, Advanced Robotics, vol.24, no.5-6, pp.739-761, 2010.

[9] 五十棲隆勝ほか: “ヒューマノイドロボット HRP-2 の開発”, 日本ロボット学会誌, vol.22, no.8, pp.1004-1012, 2004.

[10] 河原達也ほか: 『知の科学 音声対話システム』, オーム社, pp.43-44, 2006.