5T-06

早言いクイズ司会者ロボットの開発と評価

西牟田 勇哉[1]　糸山 克寿[1]　吉井 和佳[1]　奥乃 博[2]

1. 京都大学 大学院情報学研究科 知能情報学専攻　2. 早稲田大学 実体情報学博士プログラム

1. はじめに

クイズゲームはマルチパーティ人・ロボットインタラクションにおいて重要なトピックである [1, 2]．その司会者のタスクは，(1) 適切なゲームの司会進行，(2) 参加者・聴衆を盛り上げる言動であるが，我々はまず (1) に着目し，クイズゲームを適切に進行するロボット司会者を開発している [3, 4]．本研究では，一般的な「早押し」クイズをケーススタディとした「早言い」クイズを取り扱う．「早押し」クイズには以下の 2 種類の形式がある．

1. **教室型**: 教室で行われる授業のように事前に声による合図を行い，回答者を決定してから回答する [3].
2. **せり型 (オークション型)**: せりやオークションのように事前合図を行わず，直接回答する [4].

このようなクイズインタラクションでは，複数プレイヤーが同時に回答したり，ロボットの出題中にプレイヤーが割り込み回答することがある．本研究のロボットは，自身に装着したマイクロフォンアレイを用いた音源定位・分離といったロボット聴覚技術に基づいて複数人の音声を処理し，クイズゲームの進行を管理する．本稿では，せり型の「早言い」クイズについて開発したロボットの概要と評価について述べる．評価ではロボットの性能評価に加えて被験者実験を行い，人とロボットの聴覚能力について比較する．

2. 「早言い」クイズとロボットクイズ司会者

本節では，本研究で取り扱う「早言い」クイズとそのロボット司会者について述べる．「早言い」クイズとは，出題に対し直接回答するタイプのクイズゲームであり，ロボットが問題を読み上げている間であっても回答が可能である．そのため，プレイヤーは自身の音声のみでゲームに参加可能であり，押しボタンなどの事前合図のための特別なデバイスを必要としない．

ロボット司会者のシステム構成を図 2 に示す．ロボットは自身に装着したマイクロフォンアレイでプレイヤーの音声を受け付ける．入力音響信号はロボット聴覚ソフトウェア HARK [5] によって音源定位・分離され，分離音は Julius [6] によって認識される．発話の音源定位結果と事前に登録したプレイヤーの位置情報を比較することでゲーム中の発話とプレイヤーの同定を行い，それぞれの回答である分離音を全て認識する．認識結果が正解であった発話については，オンセット時刻を比較することで発話の順序を検出する．音声認識部では，分離音の音声認識精度を向上するためにインタラクションの状況に応じた適切な言語モデルの切り替えによる対話のルール外発話や誤認識の抑制 [7]，音節タイプライタと通常の記述文法の認識尤度比較による雑音の棄却を行う [8].

3. 評価実験

ロボットの性能を評価するために，同時発話からの最速発話者検出と最速発話者の音声認識の成功率について評価実験を行った．本実験では，ロボットの評価実験に加えて被験者実験を行い，その結果を比較し考察する．

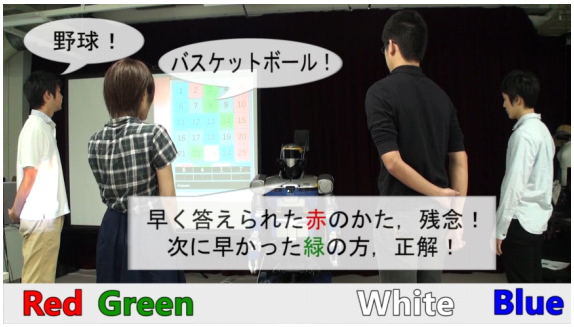


図 1: ロボット司会者とクイズインタラクション．2 人のプレイヤーの同時回答を処理し，進行を管理している．

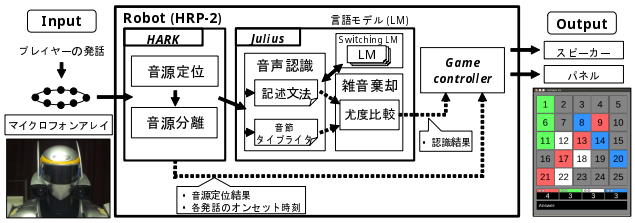


図 2: ロボット司会者 システム構成

3.1 実験設定

発話内容を変更しつつ繰り返し実験を行うため，図 3 上部に示すように人の代わりにスピーカを用いて環境を構築した．ロボット (被験者) は，各スピーカから 1.5m の距離に位置する．それぞれのスピーカの間隔は 40° であり，スピーカの地上からの高さは 1.5m である．スピーカからは事前に録音した選択式クイズの回答 (20 代，男女比 3 : 1) である単語の同時発話を出力した．本実験では最速スピーカは必ず他のスピーカと異なる発話を出力する．ロボット後方のカーテン仕切りの背後に設置された計算機・ファイルサーバからはファンノイズが常に発生している[1]．被験者実験では図 3 左下のように，被験者がロボットと同じ位置で実験を行う．被験者とスピーカの位置関係は，評価実験におけるロボットとスピーカの位置関係と同じである．

3.2 実験内容

本実験では，3，4 人の同時回答について評価した．先に 1 台のスピーカから回答を出力し，残りのスピーカは 20–200msec の同じ時間差を与えて同時に回答を出力する．同時回答からの最速発話者同定成功率 R_{fp}，最速発話音声認識成功率 R_{sr} は式 (1) に従い求めた．

$$R_{fp} = \frac{M_{fs}}{N_{all}}, \quad R_{sr} = \frac{M_{sr}}{N_{all}}, \tag{1}$$

ここで，N_{all} は同時回答の総発話回数，M_{fs} は最速発

Development and Evaluation of a Quizmaster Robot for the Fastest-Voice-First Quiz Interaction: *Izaya Nishimuta, Katsutoshi Itoyama, Kazuyoshi Yoshii (Kyoto Univ.), and Hiroshi G. Okuno (Waseda Univ.)*

[1]実験室: 63.0[dB] (ファンノイズのない環境では 43.2[dB])

話者を正しく検出した回数，M_{sr} は最速発話者の発話を正しく認識した回数である．ロボットの音声認識には回答の候補を複数含む記述文法を使用し，被験者実験における音声認識は記述文法の回答候補一覧からの選択とした．本実験では，同時発話出力中におけるロボット (被験者) 後方に設置されているスピーカからの音[2]出力 (バージイン状態) の有無を変更して実験し，結果を比較した．バージインが有る場合は無い場合に比べてマイク周辺の SNR (Sound-Noise-Ratio) が高くなり，最速発話者の同定や認識が困難となる．

3.3 実験結果

図 4, 5 にそれぞれ与えた時間差ごとの最速発話者の同定成功率，最速発話の音声認識成功率の結果を示す．被験者実験の結果は被験者 6 人 (20 代，男女比 5:1) の平均である．ロボットの評価結果について，同定成功率はバージインの有無による差は生じず，与えた時間差が大きくなるほど成功率が向上した．音声認識成功率は与えた時間差に関係なく常に低く，バージインが有る場合は無い場合に比べて更に低下した．被験者実験では，バージインの影響を受けていなかった最速発話者の同定については，同時発話人数が 4 人であるときのバージイン状態における成功率が低いが，その音声認識精度は高かった．

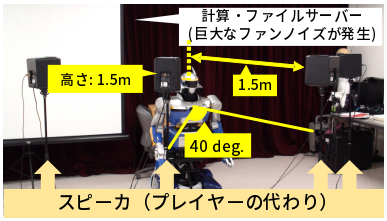
3.4 考察

図 4 から，ロボットはバージインの有無に関わらず被験者実験よりも高い成功率を獲得したことが確認できる．これはロボットが人よりも発話者順序の検出を正確に行うことを示唆している．また図 5 に示すように，最速発話の音声認識について，ロボットの成功率が人のそれよりも大幅に低い．そのため現状のロボットでは，せり型のクイズの司会を務めることは難しく，教室型のアプローチが適していると考えられる．ただし，被験者実験の際に行ったアンケートでは,「他の発話を捨て，最も早いと思った発話に集中すると認識できた.」といった特定の音に注目するカクテルパーティ効果を現すような回答があったのに対し，ロボットは同時発話に含まれる全ての発話順序を検出し，認識している．それにより特定の音に注目するよりも，同時発話に対する応答の幅を広げることができると考えられる．その実現には音声認識成功率の向上が課題となるが，この課題については既知音や残響，反射の影響を抑圧することで改善が期待できる．

4. おわりに

本稿では,「早言い」クイズその司会を行うロボットを構築した．同時発話からの発話とプレイヤーの同定や最速発話者の検出・認識は音源定位・分離や発話のオンセット時刻の比較によって実現した．性能評価と被験者実験により，多人数かつ短い時間差の同時発話についてロボットは人よりも発話順の検出を正確に行えることを確認した．今後は，音声認識率の向上や，認識を全発話に拡張した被験者実験などを行う．

謝辞 本研究の一部は，科研費 基盤研究 (S) No.24220006 の支援を受けた．



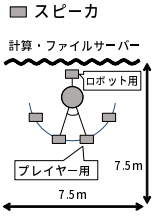


図 3: 実験環境 (上: ロボットの評価実験の様子，左下: 被験者実験の様子，右下: 実験環境俯瞰図)

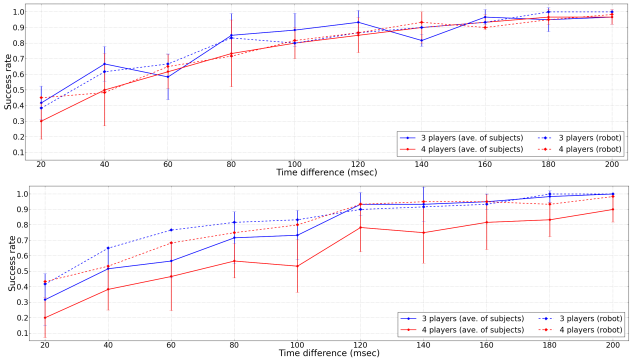


図 4: 最速発話者同定成功率 (上: バージイン無し，下: バージイン有り)

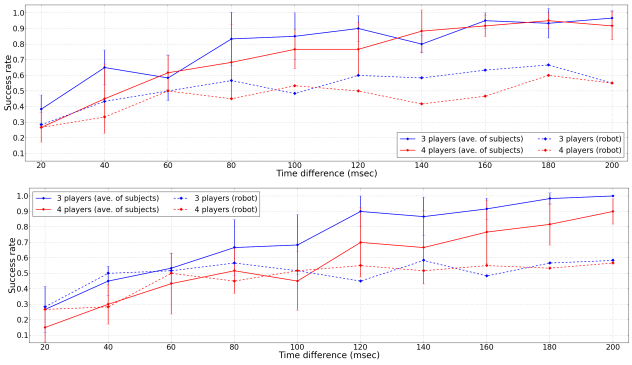


図 5: 最速発話音声認識成功率 (上: バージイン無し，下: バージイン有り)

参考文献

[1] Y. Matsuyama *et al.*, “Framework of communication activation robot participating in multiparty conversation,” *AAAI Fall Symposia*, 2010, pp. 68–73.

[2] R. Looije *et al.*, “Help, I need some body the effect of embodiment on playful learning,” *Proc. of IEEE-RO-MAN*, 2012, pp. 718–724.

[3] I. Nishimuta *et al.*, “A robot quizmaster that can localize, separate, and recognize simultaneous utterances for a Fastest-Voice-First quiz game,” *Proc. of IEEE-HUMANOIDS*, 2014, pp. 967–972.

[4] I. Nishimuta *et al.*, “Development of a robot quizmaster with auditory functions for speech-based multiparty interaction,” *Proc. of IEEE/SICE-SII*, 2014, pp. 328–333.

[5] 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.

[6] A. Lee *et al.*, “Recent development of open-source speech recognition engine Julius,” *Proc. of APSIPA-ASC*, 2009, pp. 131–137.

[7] M. Santos-Pérez *et al.*, “Topic-dependent language model switching for embedded automatic speech recognition,” *Ambient Intelligence - Software and Applications*, 2012, vol. 153, pp. 235–242.

[8] T. Jitsuhiro *et al.*, “Rejection of out-of-vocabulary words using phoneme confidence likelihood,” *Proc. of IEEE-ICASSP*, vol. 1, 1998, pp. 217–220.

[2]回答を促すサイン音．実際のクイズゲームでの BGM に相当