

# 複数ロボット間連携による対話破綻回避効果の分析

## Avoiding dialogue breakdown through the coordination of multiple robots

杉山弘晃 \*1  
Hiroaki Sugiyama

目黒豊美 \*1  
Toyomi Meguro

吉川 雄一郎 \*2  
Yuichiro Yoshikawa

大和 淳司 \*3  
Junji Yamato

\*1 NTT コミュニケーション科学基礎研究所  
NTT Communication Science Laboratories

\*2 大阪大学  
Osaka University

\*3 工学院大学  
Kogakuin University

Current chat-oriented dialogue systems have a difficulty in understanding the continuity between a wide range of topics that are observed in chatting. Due to this difficulty, such systems sometimes utter sentences that are not related with dialogue context and cause dialogue breakdown. In this paper, we analyze the effects of multiple robots in conversation on avoiding such dialogue breakdown.

### 1. 序論

近年、従来のタスク指向の対話システムとは異なる、雑談を行う対話システムに注目が集まっている [大西 14, Ritter 11, Wong 12, 東中 14]. 雑談対話は、エンタテインメントやカウンセリング目的のみならず、ユーザの潜在的な要求を引き出したり、ユーザと良好な関係を構築する上で重要である。

雑談対話システムを構築する上での課題の一つに、対話破綻の回避がある。雑談対話システムは、ユーザ発話に含まれる幅広い話題に回答する必要がある。そのため、適切な応答を出力し続けることは難しく、現在の対話システムでは、対話を破綻させるような発話がしばしば生成される問題がある [東中 15]. 対話の破綻を引き起こす不適切なシステム発話を自動的に検出する技術も盛んに研究されているが、推論が必要な話題の遷移を適切に評価できないなど、未だ課題も多い [杉山 16a].

これに対し、Fig. 1 のように複数の対話ロボットを協調動作させることで、人とロボットが1対1で対話するよりも、対話の破綻を回避できる可能性が示唆されている。例えば飯尾らは、公共のイベントの来場者に対し、対話ロボットを複数体化することで、単体ロボットの場合に比べて、対話の成立感が向上することを示している [飯尾 16]. しかし飯尾らの研究では、ユーザからシステムへの入力ボタンインターフェースによるポジティブ・ネガティブの表出に限定されており、本質的に対話破綻が生じない設計となっている。そのため、複数ロボット化が対話破綻回避に有用であるかまでは示されていない。また、どのような要因で対話の成立感が向上したかについて、詳細な分析はなされていない。一方有本らは、ロボットを複数体にすることで、発話が無視された印象や会話の難しさを軽減できることを示している。この研究では、まずロボットからユーザに質問し、ユーザからの応答に対してロボットは「すっか」などの相槌を打ち、その後あらかじめ質問に紐付けられた自己開示発話を行う、というプロトコルで対話が進められた。すなわち、ユーザからロボットへの入力は自由発話となったものの、依然ユーザ発話を認識せずにあらかじめ人手で紐付けた文を発話している。音声認識を用いない、非常に限定された状況についての検証に特化しているため、一般的な雑談対話システムにおける対話破綻回避に有用であるかは示されていない。

一般的な雑談対話における、ロボット複数体化の対話破綻回避効果を直接調べた研究として、白石ら [白石 16] がある。白石らは、対話破綻を引き起こすような文をロボットが発話する連絡先: 杉山弘晃, NTT コミュニケーション科学基礎研究所, 〒 619-0237 京都府相楽郡精華町光台 2-4, sugiyama.hiroaki@lab.ntt.co.jp



図 1: 複数ロボット雑談対話システム

際に、その発話をするロボットとは別のロボットを介入させることで、対話全体の破綻感が有意に緩和されることを、ビデオ視聴実験を通して示している。彼らは実験に用いる対話ビデオを、人-システム間の対話で発生した破綻を分析する、対話破綻検出チャレンジのデータを基に構成している。ビデオでは、ロボットがユーザ発話に基づいた発話をしているため、有本らが検証した音声認識なしの実験条件に比べ、より自然な状況での対話破綻回避効果が示されているといえる。しかしながら、2体目のロボットによる介入発話が入手で作られており、本来の破綻検出チャレンジで用いられた対話システムの性能を超えた介入がなされている可能性がある。また、ビデオ視聴を通じた第三者評価によって検証がなされており、直接対話者の主観評価が向上するかについては扱われていない。

本研究では、雑談対話におけるロボットを複数体としたときの対話破綻回避効果について、実際に人とロボットが対話する状況下で検証する。実際に複数ロボットでの対話システムを構築し、展示会などで来場者の反応を観察したところ、様々な要因で対話の破綻感が軽減され、対話の満足度が改善される可能性が示唆されていた。そのため、対話破綻および対話の満足度に関連する要因について、ロボットの複数体化に付随して起こる差異との関連を詳細に検証する。

### 2. システム

#### 2.1 対話のフロー

Fig. 2 に、複数ロボットが対話する際の基本フローを示す。各ボックスがあるタイミングでのいずれかの話者の行動を示しており、ボックス内上段の A→B は話者 A が話者 B に向けて発話することを示す。話者は R1 が1体目のロボット、R2 が

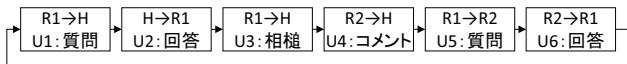


図 2: 対話フロー

2 体目のロボット, H がユーザを表す. 対話は R1 から H に向けての質問 U1 から始まり, それに対する H の回答 U2 を R1 が相槌 U3 で受け, その後 R2 が H に対して, H の回答 U2 に関連する追加のコメント U4 を発話する. このコメント U4 に関連する質問応答 U5, U6 が R1 と R2 の間でなされた後, R1 から H への質問 U1' に戻る, という流れである. ここで, 追加コメント U4 はユーザの発話に対してなされるため破綻の可能性はあるが, ロボット間の質問応答 U5, U6 は不確定要素がないため, 確実に成立するように応答対をあらかじめ用意できるように注意する. このようなフローとすることで, ユーザの幅広い話題の発話を相槌と追加コメントで受け止めつつ, ロボット間で関連する質問応答を行うことで, ロボットが対話の主導権を保持し続けることができる. さらに, もし上記追加コメントがやや文脈からそれた内容になっていたとしても, 即座に確実に成立するロボット間のやりとりをはさむことで, 対話全体の破綻を回避できると期待できる.

## 2.2 システムの実装

本節では, Fig. 2 中の各ボックスで用いる構成モジュール, および全体の実装について説明する.

R1 から H への質問には, 杉山らによって収集された, 話者の嗜好や経験を問う, パーソナリティ質問を利用する [杉山 16b]. 杉山らは, 初対面の人に尋ねる質問・応答対を大規模に収集・分類し, それらを Person DB (PDB) としてまとめている. PDB に含まれているパーソナリティ質問は, 一般的に発話されやすい質問であるため, 対話の話題を切り出す上で適切な質問であると考えられる. PDB から実際に質問を選択する際は, まだ発話されていない質問の中から, 直前 2 発話との文間類似度が高い話題に属している中で最も PDB 中の出現数が多い質問を選択する. 文間類似度の計算には Word Mover's Distance (WMD) を利用する. WMD で考慮する単語は内容語に限定し, かつ名詞が存在する場合には名詞のみを考慮することとする. また, 類似話題が見つからなかったとき, 対話の開始時, および同じ話題が 3 回繰り返された場合には, 20 種類の First Question から最も類似度が高いものを発話する. First Question とは, 人同士の雑談対話の開始時に発話された質問のうち, 文脈依存性が少なくいつ発話しても受け入れやすいと思われる質問を, あらかじめ人手で選択したものである.

相槌は, 「そうですねですかー」や「うんうん」などの曖昧な応答を発話し, それに加えてユーザ発話に名詞が含まれている場合には, その名詞を「○○なんですね」のようにリピートする. 追加コメントは, 東中らの雑談対話システム [Higashinaka 14] を利用して得られた発話候補文のうち, 質問や挨拶などを除外したものを用いる. ロボット間の質問応答は PDB に含まれている質問応答を利用し, ユーザに提示する質問と同様の方法で直前 2 発話と類似する質問を選択し, それに紐付いた応答と合わせて発話する.

全体の実装として, 対話ロボットには, CommU (Fig. 1 \*1) を用いる. 音声を受け取るマイクはインテリジェントマイク \*2 を, 音声認識エンジンには NTT 製の SpeechRec \*3 を用いる. 音声合成は, CommU の外見に合わせた子供の声として, AI Talk 社製のこうたろうとあんず \*4 を用いる.

\*1 <https://www.vstone.co.jp/products/sota/index.html>

\*2 <http://www.ntt.co.jp/ntt-tec/application/2016ap002.html>

\*3 <http://www.ntt-it.co.jp/product/v-series/speechrec/>

\*4 <http://www.ai-j.jp/>

表 1: 各実験セットの対話フローの差異

| ID | セット名         | 台数 | 相槌 | レスポンス | コメント    | ロボ QA |
|----|--------------|----|----|-------|---------|-------|
| 1  | 2 台通常        | 2  | ✓  | 早     | ✓       | ✓     |
| 2  | 2 台レスポンス遅    | 2  | ✓  | 遅     | ✓       | ✓     |
| 3  | 2 台相槌なし      | 2  |    | 遅     | ✓       | ✓     |
| 4  | 1 台通常・相槌あり   | 1  | ✓  | 早     | ✓       |       |
| 5  | 1 台相槌のみ      | 1  | ✓  | 早     |         |       |
| 6  | 1 台相槌なし      | 1  |    | 遅     | ✓       |       |
| 7  | 2 台横領き QA なし | 2  | ✓  | 早     | ✓ (非切替) |       |
| 8  | 2 台コメントなし    | 2  | ✓  | 早     |         | ✓     |
| 9  | 2 台 QA なし    | 2  | ✓  | 早     | ✓       |       |
| 10 | 2 台横領き       | 2  | ✓  | 早     | ✓ (非切替) | ✓     |

表 2: 評価項目: 5 が最良・ちょうどよい

| ID  | 評価項目               | ラベル  |
|-----|--------------------|------|
| Q1  | 対話のテンポ (発話間の間) の遅さ | テンポ  |
| Q2  | ロボットが特定の話題に固執していたか | 固執   |
| Q3  | ロボットの話題遷移は飛びすぎていたか | 飛び   |
| Q4  | ロボットからの質問の物足りなかったか | 深掘り  |
| Q5  | ロボットからの質問は煩わしかったか  | 問煩雑  |
| Q6  | あなたの話を受け止めてもらったか   | 無視   |
| Q7  | 対話が成立していたか         | 破綻   |
| Q8  | 対話を続けやすかったか        | 継続   |
| Q9  | 対話は盛り上がったか         | 盛り上り |
| Q10 | また対話したいか (対話満足度)   | 満足度  |

## 3. 実験

### 3.1 実験設定

本実験では, 対話フローおよび対話ロボットの台数を変化させ, Table 1 の各セットについて評価を行う. フロー上で変化させる部分は, 相槌の有無, レスポンスの速度, コメント発話の有無, コメント発話時の話者切替の有無, ロボット間 QA の有無, およびロボットの台数である. なお, コメント発話時に話者切替を行わない場合, コメント発話をしないロボットは, 発話するロボットを見ながら頷く動作 (横領き) を行う. また, ロボットが 1 台のみの場合, ロボット間 QA は不可能となる.

実験参加者は日本語話者 10 名 (男 6 名, 女 4 名, 20 代 ~ 50 代でバランス, 複数ロボットとの対話経験のある者は含まない) で, 各話者はロボットとの音声対話を計 20 回 (10 セットと 2 回ずつ) 行う. 各対話時間は 4 分で, 各対話後に, Table 2 に示す各評価項目について, 5 段階 Likert 尺度で評価を行う. これらの評価値を Table 3 に示す各セットの対について比較する. 有意性の検定は, Wilcoxon の符号順位検定を用いる.

表 3: 比較要素, およびそれらに対応するセット対

| ID  | 比較要素                    | セット対 (あり-なし) |
|-----|-------------------------|--------------|
| D1  | 相槌の有無 (レスポンス改善なし)       | 2-3          |
| D2  | 相槌の有無 (レスポンス改善あり・2 体)   | 1-3          |
| D3  | 相槌の有無 (レスポンス改善あり・1 体)   | 4-6          |
| D4  | レスポンス改善 (相槌効果なし・2 体)    | 1-2          |
| D5  | 追加コメントの有無 (2 体)         | 1-8          |
| D6  | 追加コメントの有無 (1 体)         | 4-5          |
| D7  | コメント話者切替の有無 (ロボ間 QA なし) | 9-7          |
| D8  | コメント話者切替の有無 (ロボ間 QA あり) | 1-10         |
| D9  | ロボ間 QA の有無 (コメント話者切替なし) | 10-7         |
| D10 | ロボ間 QA の有無 (コメント話者切替あり) | 1-9          |
| D11 | ロボット多数派 (コメント話者切替なし)    | 7-4          |
| D12 | ロボット多数派 (コメント話者切替あり)    | 9-4          |

表 4: 各セットの評価値。ラベルの極性によらず、大きい値が良い評価を表す。最大値を太字、2 番目に大きい値を下線付きイタリックで表す。

| ID | セット名         | テンポ         | 固執          | 飛び          | 深掘り         | 問煩雑         | 無視          | 破綻          | 継続          | 盛り上り        | 満足度         |
|----|--------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 1  | 2 台通常        | 4.25        | 3.90        | 3.4         | 3.80        | 3.90        | <u>3.95</u> | <u>3.85</u> | <u>3.95</u> | 3.45        | 3.85        |
| 2  | 2 台レスポンス遅    | 2.95        | 4.00        | <u>3.65</u> | 3.85        | 4.1         | 3.80        | 3.65        | 3.75        | 3.1         | 3.75        |
| 3  | 2 台相槌なし      | 3.2         | 3.85        | 2.95        | 3.5         | 3.75        | 3.35        | 3.1         | 3.1         | 2.85        | 3.45        |
| 4  | 1 台通常・相槌あり   | <u>4.65</u> | 3.8         | 3.5         | 3.55        | 4.15        | 3.90        | 3.70        | 3.90        | <u>3.50</u> | <u>4.05</u> |
| 5  | 1 台相槌のみ      | <b>4.85</b> | 3.6         | 3.0         | 3.05        | 3.65        | 3.75        | 3.70        | 3.65        | 3.20        | 3.75        |
| 6  | 1 台相槌なし      | 3.2         | 3.55        | 3.1         | 3.6         | 3.75        | 3.25        | 3.2         | 3.2         | 3.15        | 3.4         |
| 7  | 2 台横領き QA なし | 4.4         | <u>4.15</u> | <b>3.95</b> | <u>3.90</u> | 4.15        | 3.70        | 3.70        | 3.9         | 3.35        | 3.8         |
| 8  | 2 台コメントなし    | 4.60        | <b>4.20</b> | 3.55        | 3.6         | <u>4.25</u> | 3.90        | <b>4.2</b>  | <b>4.1</b>  | <b>3.55</b> | 4.0         |
| 9  | 2 台 QA なし    | 4.25        | 3.95        | <b>3.95</b> | 3.80        | <u>4.25</u> | <b>4.05</b> | 3.80        | <u>3.95</u> | 3.35        | <b>4.1</b>  |
| 10 | 2 台横領き       | 4.4         | 4.10        | 3.35        | <b>4.0</b>  | <b>4.45</b> | <b>4.05</b> | 3.75        | 3.90        | 3.35        | 3.85        |

### 3.2 比較要素

本研究では、2.1 節で示した対話フロー中の各動作や、対話ロボットを複数体化がユーザの対話への印象に与える影響を調べるため、Table 1 の各セットを比較する。本研究では特に、Table 3 および以下に示す、6 種類 12 項目の要素に着目して比較検証する。

#### 相槌の有無

相槌を打つことで、それが曖昧な応答であっても、ユーザの発話を受け止めている印象 Q6 が改善されると期待される。特にリフレーズをすることで、単なる曖昧な応答のみの場合に比べ、より受け止めてもらっている感が向上すると予想される。

なお、相槌を発話する際、曖昧な発話であれば、発話自体を解析する前に発話することができる。その場合、ユーザ発話解析を必要とする通常の発話を生成するよりも短い時間で応答を生成することができるため、対話のテンポ感 Q1 を改善できる可能性がある。本実験では、上記レスポンスの改善を含む条件と、あえてレスポンスを遅くし相槌単体の効果を検証する条件 (2 体, 1 体) の計 3 パターンについて検証する。

#### レスポンスの上昇

相槌を発話しつつ、レスポンスを遅くするセットを用意することで、相槌の有無によらない形で、レスポンス上昇の効果を検証する事ができる。レスポンスの向上により、テンポ感 Q1 の向上が見込めるとともに、それによる対話の盛り上がり Q9 も改善されると予想される。

#### 追加コメントの有無

単純なりピートや相槌よりも、ユーザの発話に適切に対応する発話となるため、ユーザの受け止めてもらった感 Q6 が向上すると予想される。また、適切に話題が遷移することで、特定の話題への固執 Q2 が改善される効果も期待される。一方、直接ユーザ発話に応じる形の発話となるため、話題が不連続となったり (Q3)、対話が続けにくさ Q9 が悪化する可能性がある。本検証では、ロボットが 2 体の場合、1 体の場合の双方について検証する。

#### コメント話者切替の有無

ロボットがユーザ発話に対してコメント発話を行う際、話題が不連続となる場合がある。そのとき、コメントを発話するロボットを、相槌を打つロボットと別ロボットとすることで、話題の不連続感 Q3 を軽減できる可能性が示されている [有本 16]。一方、R1 による質問の内容に関連した話題について R2 が再び発話することになるため、やや話題への固執感 Q2 に悪影響が出る可能性もある。ここでは、コメント発話後に QA を発話する条件、発話しない条件の 2 パターンについて検証する。

#### ロボット間 QA の有無

ロボットが複数体いる場合、ロボット間で確実に破綻しない対話を行うことができる。コメント発話が不自然な発話であったとしても、そうしたコメント発話後にロボット間の対話を挟み込むことで、対話全体の成立感 Q7 が向上する可能性がある。本実験では、ロボット間の対話に PDB を利用した質問応答を用いる。なお、この効果はコメント発話の話者切り替えと重複しているため、本実験では、コメント発話の話者切り替えを行う場合と行わない場合の 2 パターンについて検証する。

#### ロボット多数派

通常のユーザとロボットが 1 対 1 で対話する場合は、ユーザの対話能力が場の基準となり、ロボットがその水準に満たない場合、ロボットに対して失望してしまうことがある。しかし、ロボットを複数体とすることで、ロボットが多数派、ユーザがマイノリティになるため、ロボットの対話能力が場の基準となると予想される。そのため、話題の固執 U2 や飛び U3 が改善されると期待される。

### 3.3 結果と分析

Table 4 に各実験セットの結果を示す。各項目について最大値を太字、2 番目に大きい値を下線付きイタリックで表す。全体的に 2 台条件のほうが最大の評価値を得ることが多く、特にセット 8 の 2 台コメントなしが 4 項目、次いでセット 9 の 2 台 QA なしとセット 10 の 2 台横領きが 3 項目で最大の評価値を得ている。

一方、テンポの項目に限ると、セット 5 の 1 台相槌のみ条件が特に高い評価を得ている。セット 5 は、相槌をはさみレスポンスを向上させた中で最もロボットの発話数が少ないセットである。一方、レスポンスが早いセットの中では、セット 1 の 2 台通常や、セット 9 の 2 台 QA なしが低めの評価となっており、コメント発話を話者を切り替えて提示する場合の評価が低くなっている。これは、コメント発話が話題とややずれた発話となりがちであり、そうしたやや不自然な発話を別ロボットが強引に割り込んで発話することで対話のリズムが崩れ、少なくともテンポに対しては悪影響を及ぼしていたと考えられる。

Table 5 に、Table 3 のセット対ごとに比較した際に、有意差のあった評価項目を示す。P は各比較要素で正の効果があった評価項目 (Table 3 のセット対で左にあるセットが高い項目) を、N は負の効果があった評価項目を表す。なお、() がついていないものは  $p < .01$ , ついていないものは  $p < .05$  の有意水準を表す。D1 から D3 を見ると、相槌を挟むことで全体的に改善されていることがわかる。レスポンスの改善がない状態でも話題の飛びや破綻感の軽減に有用であり、特に 2 体条件で幅広く改善していることがわかる。一方、レスポンスのみを改善させることで、テンポが改善する他、盛り上がりが改善されることがわかる。

表 5: セット間比較で有意差のあった比較要素. P は正の効果, N は負の効果を表す. ( ) なしは  $p < .01$ , ( ) つきは  $p < .05$  を表す.

| ID  | 比較要素                    | テンポ | 固執  | 飛び  | 深掘り | 問煩雑 | 無視  | 破綻  | 継続  | 盛り  | 満足度 |
|-----|-------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| D1  | 相槌の有無 (レスポンス改善なし)       |     |     | P   |     | (P) |     | P   | (P) | P   |     |
| D2  | 相槌の有無 (レスポンス改善あり・2 体)   | P   |     | P   | (P) |     | P   | P   | P   | P   | P   |
| D3  | 相槌の有無 (レスポンス改善あり・1 体)   | P   |     |     |     | P   | P   | (P) | P   | (P) | P   |
| D4  | レスポンス改善 (相槌効果なし・2 体)    | P   |     |     |     |     |     |     |     | P   |     |
| D5  | 追加コメントの有無 (2 体)         | N   |     |     | P   | N   |     |     | N   |     | N   |
| D6  | 追加コメントの有無 (1 体)         | N   | P   | (P) | P   | P   |     |     | (P) |     | P   |
| D7  | コメント話者切替の有無 (ロボ間 QA なし) | N   | N   |     |     |     |     |     |     |     | P   |
| D8  | コメント話者切替の有無 (ロボ間 QA あり) | N   | N   |     | N   | N   |     |     | P   |     |     |
| D9  | ロボ間 QA の有無 (コメント話者切替なし) | (N) | N   | (N) |     | P   | (P) | (P) |     |     | (P) |
| D10 | ロボ間 QA の有無 (コメント話者切替あり) |     | (N) | N   |     | (N) |     |     |     |     | N   |
| D11 | ロボット多数派 (コメント話者切替なし)    | N   | P   | (P) |     |     |     |     |     |     | N   |
| D12 | ロボット多数派 (コメント話者切替あり)    | N   |     | (P) |     | (P) |     |     |     | (N) | (P) |

また, コメント発話を追加することで, ロボット質問の物足りなさを改善することができている。これは, 本実験での実装では, ロボットの質問はユーザ応答に関連はするものの話題を深掘りする質問ではないため, ロボットがコメント発話を挟まずに次々と質問した場合に, かえって聞いてほしい欲求が高まるためと考えられる。特に, Table 4 で 1 台相槌のみ (コメント発話・ロボ間 QA ともに無い唯一のセット) の深掘りが非常に低い評価であることも, この仮説を支持していると考えられる。一方, コメント発話の追加によって, 2 体, 1 体条件のいずれにおいてもテンポが低下している。1 体条件ではそれ以外は概ね改善されているものの, 2 体条件では逆に印象が悪化している。1 体条件では, 多少話題がずれたとしても, 相槌のみを打つ場合に比べコメント発話を追加することで幅広い話題に対応できる印象をあたえることができる。しかし 2 体条件では, 元々ロボット間で確実に成立する QA 対話を挟んでいるため, コメント発話の正の効果があまり感じられず, むしろ話題のずれが強調されたものと考えられる。

コメント話者の切り替え (D7, D8) を調べると, ロボット間 QA あり・なしの両条件で, テンポと固執の評価が低下したが, ロボット間 QA なしでは満足度が向上した。有意差は出ていないものの, 無視された感など他の項目での改善傾向が見られているため, そうした複数の要因の組み合わせが満足度向上につながったものと考えられる。関連して, ロボット間 QA の有無 (D9, D10) を調べると, コメント話者切り替えがない場合では問煩雑, 無視, 破綻, 満足度に改善傾向が見られたものの, テンポ, 固執, 飛びは悪化した。また, コメント話者切り替えがある場合は, ロボット間 QA を行うと固執, 飛び, 問煩雑, 満足度が悪化した。双方で固執, 飛びに悪影響が見られるため, ロボットが 2 体で 1 つの話題について話し続けることで固執感が生まれ, かつ大きく話題が飛んだときにその印象が強まるものと考えられる。一方, ロボット間 QA をせず, かつコメント話者切り替えを行わないと, 2 体目のロボットがほぼ何も発話しないことになるため, 違和感が大きくなったものと考えられる。

ロボットが多数派になる影響を調べたところ, テンポには悪影響が出ているものの, 逆に話題の飛びに対しては改善傾向が見られる。また, コメント話者の切り替えを伴う場合には満足度も向上している。セット 9 が最も高い満足度を得たことを考えると, ロボットを多数派にし適切に発話を振り分けることによる満足度の改善効果は大きいものと思われる。

#### 4. 結論

本研究では, 雑談対話ロボットを複数体化することによる効果を検証した。検証の結果, 相槌を除きロボットの発話数が大きいほどテンポ感は悪化していたが, ロボットを複数体にすることで, 満足度や他の尺度が改善されることが示された。

本実験では, ロボットの発話順などを完全に固定して実験していたため, 自由なターンテイクを許した場合の印象は多少変化すると予想される。特に, 本来 1 対 1 で質問を続けると圧迫感や不自然感が増すと考えられるが, 予想に反してそうした影響はあまり見られなかった。これは, 実験を 2 体と 1 体で頻りに切り替えて行っており, かつロボットから質問を受け続けていたため, 1 対 1 での一方的な質問にもあまり違和感を覚えなくなっていたものと考えられる。こうした分析結果を踏まえ, 今後実験プロトコルの改良および自由なターンテイクの実現に取り組んでいきたい。

#### 謝辞

音声認識器, インテリジェントマイクおよび雑談対話システムをご提供いただいた NTT メディアインテリジェンス研究所 音声言語メディアプロジェクトの皆様へ感謝いたします。

#### 参考文献

- [Higashinaka 14] Higashinaka, R., Imamura, K., Meguro, T., Miyazaki, C., Kobayashi, N., Sugiyama, H., Hirano, T., Makino, T., and Matsuo, Y.: Towards an open-domain conversational system fully based on natural language processing, in *Proc. COLING*, pp. 928–939 (2014)
- [Ritter 11] Ritter, A., Cherry, C., and Dolan, W. B.: Data-Driven Response Generation in Social Media, in *Proc. EMNLP*, pp. 583–593 (2011)
- [Wong 12] Wong, W., Cavedon, L., Thangarajah, J., and Padgham, L.: Strategies for Mixed-Initiative Conversation Management using Question-Answer Pairs, in *Proc. COLING*, pp. 2821–2834 (2012)
- [杉山 16a] 杉山弘晃: 発話生成における誤りパターンの分析に基づく対話破綻検出, 第 7 回対話システムシンポジウム, pp. 81–84 (2016)
- [杉山 16b] 杉山弘晃, 目黒豊美, 東中竜一郎: 対話システムのパーソナリティを問う質問の大規模な収集と分析, 人工知能学会論文誌, Vol. 31, No. 1, pp. 1–9 (2016)
- [大西 14] 大西可奈子, 吉村健: コンピュータとの自然な会話を実現する雑談対話技術, NTT DoCoMo テクニカル・ジャーナル, Vol. 21, No. 4, pp. 17–21 (2014)
- [東中 14] 東中竜一郎: 雑談対話システムに向けた取り組み, 第 70 回言語・音声理解と対話処理研究会 (SIG-SLUD) (2014)
- [東中 15] 東中竜一郎, 船越孝太郎, 荒木雅弘, 塚原裕史, 小林優佳, 水上雅博: Project Next NLP 対話タスク: 雑談対話データの収集と対話破綻アノテーションおよびその類型化, 言語処理学会 (2015)
- [白石 16] 白石達, 五味輝, 大和淳司, 杉山弘晃: 対話破綻における複数ロボットの利用による破綻緩和効果の分析, HCG シンポジウム, pp. 445–448 (2016)
- [飯尾 16] 飯尾尊優, 吉川雄一郎, 石黒浩: 展示会におけるボタン入力対話体験の評価, 人工知能学会全国大会, pp. 203–3 (2016)
- [有本 16] 有本庸浩, 吉川雄一郎, 杉山弘晃, 目黒豊美, 大和淳司, 石黒浩: 対話焦点の不連続性を隠蔽する複数ロボット間連携の生成, 第 7 回対話システムシンポジウム, pp. 51–52 (2016)