

令和 5 年 6 月 27 日現在

機関番号：14301

研究種目：若手研究

研究期間：2020～2022

課題番号：20K19821

研究課題名（和文）対話理解および発話生成と連動するターンテイキングシステム

研究課題名（英文）A turn-taking system linked with dialogue understanding and utterance generation

研究代表者

井上 昂治（Inoue, Koji）

京都大学・情報学研究科・助教

研究者番号：10838684

交付決定額（研究期間全体）：（直接経費） 2,500,000円

研究成果の概要（和文）：音声対話システムにおけるターンテイキング（発話権取得）の新たなモデルを実現した。人間どうしのターンテイキングに倣い、対話データセット中の各発話の「意図」と「内容」が理解可能か否かをアノテーションした。そして、はじめに「意図」または「内容」が理解可能かを判定し、その結果に基づいて発話権を取得するか否かを判定する2段階のターンテイキング予測モデルを構築した。加えて、音声対話システムの基礎的な機能として、同調笑いの生成についても取り組んだ。笑いの検出、同調笑いの予測、笑いの種類の選択という3つのモデルで構成されるシステムを提案し、その有効性を確認した。

研究成果の学術的意義や社会的意義

音声対話システムは、会話ロボットやスマートスピーカに展開されている。しかし、これらのシステムによるやりとりは機械的であると言わざるを得ない。その要因の一つとしてターンテイキングが挙げられる。現在のシステムでは、発話権を取得するに際して、不自然に長い間や割り込みが生じることが多く、これにより対話の円滑さを低下させている。その一方で、人間どうしの対話では、特に意識することなく、円滑なターンテイキングが実現されている。本研究により、人間どうしのターンテイキングのメカニズムの解明に向けて、構成論的な一つのアプローチを示すことができた。

研究成果の概要（英文）：A novel model for turn-taking, predicting the right to speak in spoken dialogue systems, has been pioneered. To mirror human turn-taking, annotations were applied to discern the 'intent' and 'content' of each utterance within a dialogue dataset. Subsequently, a two-step turn-taking prediction model was developed. It first determines if the 'intent' or 'content' is intelligible and then decides whether to take the turn. Additionally, to enhance the functionality of spoken dialogue systems, the generation of shared laughter has been realized. A system composed of three modules for laughter detection, shared laughter prediction, and laughter type selection was proposed, demonstrating its efficacy.

研究分野：音声対話システム

キーワード：音声対話システム 会話ロボット ターンテイキング 発話権取得 話者交替

1. 研究開始当初の背景

近年、人間と音声でやりとりをする「音声対話システム」が、会話ロボットやスマートスピーカにおいて実用化されている。ただし、扱われる対話は一問一答がほとんどである。また、そのやりとりは機械的であり、あくまでツールとしての側面が大きい。将来的には、長くて深い対話を円滑に進め、人間社会において社会的な役割を担うことが期待される。

音声対話システムが対話を円滑に進めるためには、ターンテイキングの高精度化が重要である。ターンテイキングとは、現在の話し手から次の話し手へ発話権(ターン)が移行する過程を指す。音声対話システムの文脈では、ユーザのターン終了を予測することに相当する。ターン終了を正しく予測することができれば、ターンが真に終了する場合には素早く、そうでない場合にはしばらく待つといった、適応的で円滑なふるまいが実現できる。これまでに、ターン終了の予測モデルを構築する数多くの研究がなされており、主に先行するユーザ発話の韻律や言語の情報が特徴量として用いられている。ただし、ターンテイキングのふるまいは任意性が高いため実用的なシステムへの予測モデルの搭載はあまりなされておらず、一定時間の沈黙を基準とするアドホックな方法がとられることが多い。

本研究では人間どうしの会話のメカニズムを音声対話システムに応用することを目指す。人間どうしの会話におけるターンテイキングでは、より多面的な情報が考慮されているはずであり、現在の音声対話システムのそれとはギャップがあるといえる。したがって、このギャップを埋めることで、より自然で円滑なターンテイキングを音声対話システムが実現できることが期待される。

2. 研究の目的

本研究の目的は、音声対話システムにおけるターンテイキングに関して、人間どうしの会話のメカニズムに倣い、複数の機能が連動しながらターンテイキングを予測する新たな枠組みを実現することである。

3. 研究の方法

はじめに、本研究で構築するターンテイキングシステムの動作を検証するための基盤となる音声対話システムを構築する。そして、このシステム上において、基本的なターンテイキングシステムを実装し、その動作を確認する。続いて、人間どうしのターンテイキングに倣った学習が可能なデータセットを構築する。具体的には、人間どうしの対話における各発話について、その発話の「意図」と「内容」が理解可能か否かをアノテーションする。そして、このアノテーションデータを用いてターンテイキングモデルを学習し、その効果を検証する。

4. 研究成果

はじめに、提案するターンテイキングシステムの効果を実証するための音声対話システムの構築に取り組んだ。具体的には、傾聴と就職面接のシステムを実装した。傾聴システムは、相槌、掘り下げ質問、評価応答(「素敵ですね」など)などの聞き手応答を生成し、ユーザの話に耳を傾けて聞くものであり、独居高齢者などの話し相手となることを想定している。構築したシステムの評価では、人間による傾聴(Wizard-of-OZ)との比較を行い、「真面目に話を聞く」、「集中して話を聞く」、「積極的に話を聞く」といった基本的な傾聴スキルに関しては、提案システムと人間とで有意な差がないことが確認された。その一方で、「相手の話に興味を示す」、「相手の話を理解する」、「相手に対して共感を示す」といった高度な傾聴スキルについては人間のほうが有意に高く、提案システムと人間との違い(ギャップ)も明らかになった。この結果は今後の研究開発の方向性を定める上で重要な知見となり得る。就職面接については、掘り下げ質問を動的に生成する機能を実装した。既存のシステムは予め定められた質問を順番に行うものであり、面接志願者から見て想定しやすいものである。したがって、面接のリアルさや実用性が欠けてしまう。そこで、志願者の回答の内容に応じて、さらなる掘り下げ質問を動的に生成することが重要となる。ここで提案した掘り下げ質問は2種類であり、充足度判定に基づくものとキーワード抽出に基づくものである。この掘り下げ質問の有無による変化を調べたところ、面接のリアルさや緊張感などが掘り下げ質問により有意に向上することがわかった。さらに、システムのインタフェースにおいて、アンドロイド(人間酷似型ロボット)とCGエージェントを比較したところ、アンドロイドにおいて、その存在感が掘り下げ質問により有意に向上することがわかった。これは掘り下げ質問とアンドロイドの組み合わせによる相乗効果を示唆するものである。さらに、前述の傾聴対話システムについては、多人数会話の設定にも対応できるように拡張した。ここでは、複数のユーザの話に対して、順番に傾聴を行うが、話をしていない他のユーザの会話への参加度を

高めるために、必要に応じて、他のユーザの聞き手応答を引き出すための発話を行う。具体的には、従来は「素敵ですね」とシステムが単独で反応していたのに対して、他のユーザに対して「素敵ですよ？」と同意を求めるような発話を行う。これにより、会話に参加している全員がその度合いを高め、会話の場の全体が盛り上がるようにする。多人数会話はターンテイキングの今後の研究課題になり得るものであり、そのための基盤を整備することができた。

そして、上記のシステムにおいて、従来のターンテイキングシステムを実装した。従来の機能は2種類で、ユーザが一定時間以上沈黙した場合にシステムがターンを取得する「タイムアウト方式」と、ユーザ発話の言語情報からターンが終了する確率を予測し、それを基にタイムアウトの時間を動的に決定する「予測方式」である。これらのシステムが上記のいずれのタスクにおいても実時間で動作することを確認した。さらに、提案するターンテイキングシステムの初期検討として、簡単な言語理解に基づいて、上記の2つの方式を使い分ける方法も実装した。具体的には、ユーザ発話が短い、即ち音声認識結果から得られる単語数が一定数以下の場合には、動作が頑健なタイムアウト方式を採用し、一定数以上になったタイミングで予測方式に切り替える。また、別の方法として、各対話タスクにおいて実装する言語理解モジュールの出力結果に応じて切り替えることも検討した。例えば、就職面接タスクでは、掘り下げ質問を生成するために、ユーザ発話からそのキーワードを抽出するが、キーワードが何も抽出されていない場合はタイムアウト方式を用い、そして何等かのキーワードが抽出された時点で予測方式へ切り替える。以上の切り替え方式を実装し、被験者との対話実験においてその効果を確認した。特に、就職面接でのターンテイキングが頑健になることを確認した。特に、ユーザが発話を十分に終えていないタイミングでシステムが誤ってターンを取得してしまうことを減らすことができた。さらに、十分に長い発話をするユーザに対しては、予測方式に切り替えることで、ターンの移行がより円滑になる場面が観察された。

続いて、本研究課題において提案するターンテイキングシステムのモデルを学習するためのデータセット構築に取り組んだ。ここでは、収録した初対面会話データにおいて、発話の「意図」と「内容」が理解可能か否かを、各発話に対してアノテーションした。「意図」はその発話の対話行為(dialog act)がその時点で理解することができるか否か、という判断基準を設けた。対話行為とは「質問」、「回答」、「言明」などのように発話の種類を表すものである。「内容」はその発話の内容が理解できるかであるが、より客観性をもたせるために、その時点でその次の発話を生成することが可能か否かを判断基準とした。例えば、「あなたの」、「趣味は？」という連続した2つの発話がなされている場合、最初の発話では「意図」も「内容」も理解できないが、次の発話がなされた時点で両者とも理解可能となる。ここでの狙いは、ターンテイキングのふるまい(ターンが移行するか保持されるか)の手掛かりとなる、より客観的な現象を発見することである。したがって、上記の意図や内容のラベルとターンテイキングのふるまいとの関係性を調査した。その結果、「意図」や「内容」が理解可能であるときに、ターンが移行する確率が高くなることが明らかになった。

そして、上記のターンテイキングのデータセットを用いて、ターンテイキングモデルを構築した。ここで提案したターンテイキングモデルは、はじめに前述の「意図」または「内容」が理解可能か否かを判定し、その判定結果に基づいてターンが移行するか否かを判定する2段階の予測を行うものである。前半で述べた従来のシステム(予測方式)は「意図」や「内容」を考慮しない段階のモデルと捉えることができ、この段階の予測と比べて、提案システムではターンテイキングの予測精度が向上することを確認した。

また、上記のターンテイキングの研究と並行して、音声対話システムの基礎的な機能についても研究を実施した。ここでは、同調笑いの生成について取り組んだ。同調笑いとは、相手が笑ったときに、その笑いに合わせて自身も笑うものであり、結果的に両者が笑うことになる。音声対話システムにおいては、ユーザの笑いに対してシステムが適切に笑い返すことで、システムの共感スキルを向上させることが期待される。これは冒頭で述べた傾聴対話システムの比較実験で明らかになったシステムと人間とのギャップを埋める取り組みであるとも解釈できる。ここで提案したシステムは、ユーザの笑いの検出、同調笑いの予測、システムの笑いの種類の選択の3つのモジュールで構成される。システムの笑いの種類については、先行研究を参考にして、「大笑い」と「社交笑い」の2種類とした。この3つのモジュールについて、それぞれの機械学習モデルを前述の初対面会話データを用いて学習した。評価実験では、「全く笑わないシステム」および「社交笑いのみを行うシステム」と比較したところ、提案システムではシステムの共感性や人間らしさ、さらには対話の自然さが向上することを確認した。このシステムは冒頭で述べた音声対話システム上に実装されており、リアルタイムで動作することができる。

以上、本研究課題全体を通じて、ターンテイキングの新たなモデルの構築に取り組み、人間の認知処理により近いものを探索的に検討することができた。加えて、音声対話システムの基礎的な課題にも取り組み、多人数会話や同調笑いといった新たな課題を提案することができた。

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 4件）

| | |
|---|--------------------------|
| 1. 著者名 井上 昂治, ラーラー ディベッシュ, 山本 賢太, 中村 静, 高梨 克也, 河原 達也 | 4. 巻 36 |
| 2. 論文標題 アンドロイドERICAの傾聴対話システム--人間による傾聴との比較評価-- | 5. 発行年 2021年 |
| 3. 雑誌名 人工知能学会論文誌 | 6. 最初と最後の頁 H~L51_1-12 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1527/tjsai.36-5_H-L51 | 査読の有無 有 |
| オープンアクセス オープンアクセスとしている(また、その予定である) | 国際共著 - |
| 1. 著者名 井上 昂治, 原 康平, ララ ディベッシュ, 中村 静, 高梨 克也, 河原 達也 | 4. 巻 35 |
| 2. 論文標題 掘り下げ質問を行う就職面接対話システムの自律型アンドロイドでの実装と評価 | 5. 発行年 2020年 |
| 3. 雑誌名 人工知能学会論文誌 | 6. 最初と最後の頁 D~K43_1-10 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1527/tjsai.35-5_D-K43 | 査読の有無 有 |
| オープンアクセス オープンアクセスとしている(また、その予定である) | 国際共著 - |
| 1. 著者名 井上 昂治, 河原 達也 | 4. 巻 76 |
| 2. 論文標題 アンドロイドを用いた音声対話研究 | 5. 発行年 2020年 |
| 3. 雑誌名 日本音響学会誌 | 6. 最初と最後の頁 236~243 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.20697/jasj.76.4_236 | 査読の有無 無 |
| オープンアクセス オープンアクセスとしている(また、その予定である) | 国際共著 - |
| 1. 著者名 Inoue Koji, Lala Divesh, Kawahara Tatsuya | 4. 巻 9 |
| 2. 論文標題 Can a robot laugh with you?: Shared laughter generation for empathetic spoken dialogue | 5. 発行年 2022年 |
| 3. 雑誌名 Frontiers in Robotics and AI | 6. 最初と最後の頁 - |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.3389/frobt.2022.933261 | 査読の有無 有 |
| オープンアクセス オープンアクセスとしている(また、その予定である) | 国際共著 - |

〔学会発表〕 計9件（うち招待講演 1件 / うち国際学会 3件）

| |
|--|
| 1. 発表者名 Koji Inoue, Hiromi Sakamoto, Kenta Yamamoto, Divesh Lala, Tatsuya Kawahara |
| 2. 発表標題 A multi-party attentive listening robot which stimulates involvement from side participants |
| 3. 学会等名 SIGdial Meeting on Discourse and Dialogue (SIGDIAL) (国際学会) |
| 4. 発表年 2021年 |

| |
|--------------------------------------|
| 1. 発表者名 井上昂治, Lala Divesh, 河原達也 |
| 2. 発表標題 共感を表出する音声対話システムのための共有笑い生成 |
| 3. 学会等名 人工知能学会 言語・音声理解と対話処理研究会 |
| 4. 発表年 2021年 |

| |
|---|
| 1. 発表者名 井上昂治 |
| 2. 発表標題 アンドロイドERICAの音声対話システム～マルチモーダルチューリングテストへの挑戦～ |
| 3. 学会等名 音学シンポジウム2021 (招待講演) |
| 4. 発表年 2021年 |

| |
|---|
| 1. 発表者名 Koji Inoue, Kohei Hara, Divesh Lala, Kenta Yamamoto, Shizuka Nakamura, Katsuya Takanashi, Tatsuya Kawahara |
| 2. 発表標題 Job interviewer android with elaborate follow-up question generation |
| 3. 学会等名 International Conference on Multimodal Interaction (ICMI) (国際学会) |
| 4. 発表年 2020年 |

| |
|--|
| 1. 発表者名 Koji Inoue, Divesh Lala, Kenta Yamamoto, Shizuka Nakamura, Katsuya Takanashi, Tatsuya Kawahara |
| 2. 発表標題 An attentive listening system with android ERICA: Comparison of autonomous and WOZ interactions |
| 3. 学会等名 SIGdial Meeting on Discourse and Dialogue (SIGDIAL) (国際学会) |
| 4. 発表年 2020年 |

| |
|--|
| 1. 発表者名 井上昂治, Lala Divesh, 河原達也 |
| 2. 発表標題 ヒューマンロボットインタラクションのための相槌・笑いのリアルタイム検出 |
| 3. 学会等名 日本音響学会2021年春季研究発表会 |
| 4. 発表年 2021年 |

| |
|---|
| 1. 発表者名 井上昂治, Lala Divesh, 河原達也 |
| 2. 発表標題 ヒューマンロボットインタラクションにおける音響特徴に基づく共有笑いの予測 |
| 3. 学会等名 日本音響学会2021年春季研究発表会 |
| 4. 発表年 2021年 |

| |
|---|
| 1. 発表者名 井上昂治, Lala Divesh, 山本賢太, 中村静, 高梨克也, 河原達也 |
| 2. 発表標題 アンドロイドERICAの傾聴対話システムにおけるWOZとの比較評価 |
| 3. 学会等名 人工知能学会 言語・音声理解と対話処理研究会 (SIG-SLUD) |
| 4. 発表年 2020年 |

| |
|---|
| 1. 発表者名 井上昂治, Lala Divesh, 山本賢太, 中村静, 高梨克也, 河原達也 |
| 2. 発表標題 WOZとの比較による自律型アンドロイドERICAの傾聴対話システムの評価 |
| 3. 学会等名 日本音響学会2020年秋季研究発表会 |
| 4. 発表年 2020年 |

〔図書〕 計1件

| | |
|-----------------------|-----------------|
| 1. 著者名 井上 昂治、河原 達也 | 4. 発行年 2022年 |
| 2. 出版社 オーム社 | 5. 総ページ数 272 |
| 3. 書名 音声対話システム | |

〔産業財産権〕

〔その他〕

-

6. 研究組織

| 氏名 (ローマ字氏名) (研究者番号) | 所属研究機関・部局・職 (機関番号) | 備考 |
|---------------------------|-----------------------|----|
| | | |

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

| 共同研究相手国 | 相手方研究機関 |
|---------|---------|
| | |