

令和 4 年 6 月 15 日現在

機関番号：13901

研究種目：挑戦的研究（萌芽）

研究期間：2018～2021

課題番号：18K19811

研究課題名（和文）リスナーとしてのスマートスピーカーが表出する共感的応答

研究課題名（英文）Conversational response generation by smart speakers as attentive listeners

研究代表者

松原 茂樹（Matsubara, Shigeki）

名古屋大学・情報連携推進本部・教授

研究者番号：20303589

交付決定額（研究期間全体）：（直接経費） 4,900,000円

研究成果の概要（和文）：本研究では、語りのリスナーとしての役割を担うスマートスピーカーが、話し手に対して共感的応答を生成する方式について検討した。聴き手による応答を大規模に収集し、共感の深さと応答の種類の関係について定量的な分析を与えた。繰り返し応答や不同意応答など共感的応答を表出すべき語句の検出技法、及び、表出されやすさの推定技法を開発した。実験により検出及び推定性能を評価し、語りに傾聴を示す応答の生成可能性を確認した。

研究成果の学術的意義や社会的意義

語ることは人間に備わる基本的な欲求である。高齢化や独居化が進む現代社会では、人が語る機会の減少が問題となっている。スマートスピーカーなどの情報機器が語りを聴く役割を担うことにより、語りのある生活が身近になることが期待される。スマートスピーカーが語りのリスナーとしての役割を担い、話し手にとって語りたいたいと思える存在になるための応答とその表出技法にアプローチした。実験により、傾聴を示す応答生成の実現性を示した。

研究成果の概要（英文）：This study investigated methods for generating sympathetic responses to the speaker as an attentive listener of narratives. A large-scale conversational response corpus was collected, and quantitatively analyzed from the viewpoint of the relationship between the depth of sympathy and the type of response. The techniques for detecting words and phrases for which empathetic responses such as repeated responses and disagreement responses should be expressed and estimating the generation timing of responsive utterances were developed. Experimental results have shown the feasibility of appropriately generating attentive listening responses was confirmed.

研究分野：自然言語処理

キーワード：音声言語処理 会話システム コーパス 応答生成 談話理解

1. 研究開始当初の背景

(1) 語ることは人間に備わる基本的な欲求である。語るという行為は、聴き手がいて初めて成立する。社会の個人化が進み、聴き手不在の生活シーンが増加している。人が語れる機会を増やすことは現代の重要な課題といえる。

(2) スマートスピーカー（音声アシスタント機能と無線通信機能を備えた情報機器）が語りを聴く役割を担うことにより、語りのある生活を身近にすることが考えられる。スマートスピーカーが聴き手（リスナー）として認知されるには、「語りを聴き取る機能」および「語りを傾聴していることを話し手に伝達する機能」を備える必要がある。前者の機能は音声認識や言語理解の技術によって実現される。一方、後者の機能を実現する明示的な方法は、発話という手段で語りに応答することである。

2. 研究の目的

(1) 本研究では、「語りのリスナーとしての役割を担うスマートスピーカーなどの情報機器が、話し手にとって語りたくなる存在になるような傾聴を示す応答とその表出法とは？」という問いに回答するという動機のもと、それに関わる科学的知見と技術的解法を発見することを目指す。この課題にアプローチするため、話し手の語りに対して共感的応答を生成する会話エージェントの実現性を示すことを目的とする。共感的応答とは、語りの途中で表出された、語りに対する同意や称賛等を示す発話を意味する。

(2) 語りの音声データを対象に共感的応答を収録し、「文字化」「発声時刻」「談話現象」「語り」と応答の対応」が付与された共感的応答コーパスを作成する。収集した共感的応答を、「共感度」と「発話タイプ」で分類し、共感的応答を体系化する。共感的応答を生成する技法を開発し、コーパスをデータセットとした応答生成実験において、その効果を検証する。これらを達成目標とし、共感的応答の表出を応答表現の生成と応答タイミングの検出に分解してアプローチする。

3. 研究の方法

(1) 語りの聴き手を担う会話エージェントの実現に向け、傾聴応答を収集する。あらかじめ収録された語りの音声に対し、産出するに相応しい傾聴応答の表現と産出タイミングを付与するという収集方式を採用する。

自然でかつ多様な傾聴応答データの収集方法として、話し手と聴き手によるリアルタイムでのやり取りを記録し、聴き手の発話から傾聴応答を取り出すことが考えられる。しかしこの方式では、「聴き手の反応が話し手の振舞いに影響を及ぼす可能性があり、収集データの汎用性が損われる」「ある語りに対し、収集できるのは1名の聴き手の応答に制限されるため、収集する応答の表現やタイミングの多様性が損なわれる」という問題がある。

本研究では、語りデータに対して作業者が応答データを付与する。具体的には、語り音声の再生に同期して作業者が傾聴応答を産出し、その文字化データ及び時刻データを、語りデータに注釈付ける。双方向でのやり取りがない分、作業者は傾聴応答の効果的な産出に集中できるという利点がある。また、同一の語りデータに対して、新たな聴き手を設けて収録することも容易であり、収録データの網羅性の向上にも対応できる。

(2) 語りに示す共感に応じた傾聴応答の生成に向けて、傾聴応答の種類と共感度の関係について考察を与える。傾聴応答の種類を共感度の観点から5段階に分類する。傾聴応答データにおける16種類の傾聴応答を分類対象とする。

続いて、語りに対して傾聴応答を付与したデータを用いて、この分類の正当性を検証する。まず、応答の具体性を用いて、分類の正当性を評価する。応答の具体性はその種類によって異なり、例えば、具体性の高い応答を発話するには、語りの内容をより深く考慮する必要があり、語りへの共感度も高くなる。一方、具体性の低い応答は、語りの内容を深く考慮せずとも発話可能であり、共感度は低くなる。応答の具体性として2つの特徴量「応答を構成する形態素数」及び「応答の情報量」を用いる。次に、本節では、応答の汎用性の観点から、分類の正当性を評価する。例えば、繰り返し応答や評価応答が発話可能なタイミングは限られるため、汎用性が低い応答と言える。これらの応答を適切なタイミングで発話するには語りの内容を深く理解する必要があり、共感度が高くなる傾向にあると考えられる。一方、感心や相槌が発話可能なタイミングは多数存在し、これらは汎用性が高い応答と言える。これらの応答の発話に必ずしも語りの内容を深く理解する必要はなく、共感度が低い傾向にあると考えられる。以上から、応答の汎用性と共感度はトレードオフの関係にあると言えるため、傾聴応答の分類を応答の汎用性によって評価する。

(3) 語りにおける傾聴的応答を付与した応答データに基づき、繰り返しの対象となる語句の検出手法について検討する。繰り返し応答は、対話における機能の観点から、「発話権が移動する繰り返し応答」と「発話権が移動しない繰り返し応答」に分類できる。このうち、本研究では、発話権が移動しない応答を対象とする。まず、語りと応答データを用いて、繰り返し応答データを作成する。続いて、本データの一部（42,153個の形態素）を用いて、繰り返し応答466個を

分析する。これらの知見をもとに、節境界単位内の文節集合から繰り返すべき語句を含んだ文節の検出手法を開発し、その検出性能を実験により評価する。

(4) 適切なタイミングでの傾聴応答の表出は、話し手の語る意欲を促進する効果が期待できるが、不適切なタイミングでの表出は逆効果になりうる。そのため、傾聴応答の表出では、その表出タイミングが重要となる。本研究では、適切なタイミングでの傾聴応答の自動生成の実現に向けて、あるタイミングが傾聴応答の生成にどの程度適するかを推定する手法を開発する。本研究では、この適切さを表す指標として、傾聴応答の表出率を定義し、これを予測する。あるタイミングでの傾聴応答の表出率は、そのタイミングで傾聴応答を表出した聞き手の割合とする。開発手法は、語りの音響情報とテキスト情報を用いて、傾聴応答の表出率を予測する。具体的には、これらの情報を transformer ベースの手法でエンコードし、エンコード結果を 1 次元に変換して表出率を算出する。提案手法の予測性能の評価のために、表出率の予測実験を実施する。

4. 研究成果

(1) 語りのデータとして、高齢者のナラティブコーパスを使用した。傾聴応答の産出は、高度なコミュニケーションスキルを備えた作業員 11 名が担当した。11 名の作業員がすべて同一の語りデータに対して独立した環境で応答を産出した。作業員は、再生された語り音声に対しリアルタイムに応答を産出した。語り音声の再生は 1 回限りとした。応答音声は接話マイクを通して収録した。

すべての傾聴応答に対して、その応答のタイプを人手で付与した。応答タイプは、相槌や感心など、16 種類定めた。収録データにおける応答タイプにおいては、傾聴応答の代表である相槌の占める割合が、全体の 67.96% と最も大きかった。相槌以外の応答タイプは、全体の 32.04% を占めており、感心、繰り返し、評価の順に多く出現していた。

収集した応答の分析により、収録データの多頻度性、多様性、網羅性を評価した。応答の多頻度性として、収録データにおける傾聴応答の出現頻度を集計した。作業員平均で 2.47 秒に 1 度という高い出現率であった。作業員によってばらつきはあるものの、収録データにおける傾聴応答の多頻度性が示された。傾聴応答の多様性の評価として、応答の種類（文字列の異なり）に関する多様性指数を測定した。指数としては、応答あたりのエントロピーを採用した。収集データにおける応答あたりのエントロピーは、作業員平均で 4.43 であり、応答の多様性の高さを確認した。応答の網羅性では、語りデータに 11 名の作業員が独立に応答を付与していることに注目し、収集された傾聴応答データが、傾聴応答の産出タイミングを網羅しているかを評価した。語りデータに含まれる 29,969 個の産出タイミング候補のうち、25,523 個、つまり全体の 85.16% でいずれかの作業員が傾聴応答を産出していた。本データが傾聴応答の産出に適したタイミングのほとんどを網羅できていることを確認した。

(2) 傾聴応答の種類を、応答の役割を参考に共感度の観点から 5 段階に分類した。この分類の正当性を、傾聴応答データを用いて定量的に評価した。まず、評価データの全応答について 2 つの特徴量を算出し、傾聴応答の種類ごとの平均値を用いて、共感レベルごとの特徴量の平均値を算出した。レベルが大きくなると、いずれの特徴量の平均値も高くなる。これは、傾聴応答の分類の正当性を支持している。次に、傾聴応答の分類を、エントロピーを用いて評価した。評価データの全応答についてエントロピーを算出し、傾聴応答の種類ごとに平均値を用いて、レベルごとの平均値を算出した。レベルが低いほどエントロピーの値が高く、傾聴応答の分類の正当性を支持する結果となった。

(3) 傾聴応答データに繰り返し対象語句のタグ付けを行った。作業の結果、753 個の繰り返し応答に対して、830 個の形態素列にタグが付与された。タグ付け数が繰り返し応答の数よりも多いのは、1 つの繰り返し応答に対して、隣接しない複数の形態素列にタグが付与される場合があるためであった。この一部を分析データとした。分析の結果、数詞や固有名詞、情報量の大きい語句、及び、フィラーと共起する形態素列が繰り返されやすいことが明らかになった。

語りの語句から繰り返すべき語句を含んだ文節の検出手法を開発した。本手法は、節境界単位を入力とし、節境界単位を構成する文節集合から 1 つ以上の繰り返し対象文節を検出する。実験では、検出の再現率、適合率を計測し、本手法の効果を確認した。

(4) 本研究では、あるタイミングが傾聴応答の生成にどの程度適するかを表す指標として、傾聴応答の表出率を定義し、これを予測する方式を定めた。あるタイミングでの傾聴応答の表出率は、そのタイミングで傾聴応答を表出した聞き手の割合とした。表出率が高いタイミングほど、そのタイミングで応答を表出した聞き手が多いことを意味する。予測された表出率は、適切な応答生成タイミングの判断に利用されることを想定する。表出率の予測は、従来の二値の予測に比べ、より細かい粒度での明確な値の予測である。そのため、表出率を予測するシステムは、二値の予測を行う従来のシステムと比べ、移植性が高く、他の要素と組み合わせやすい。

開発した手法では、応答は語りの言語的境界の直後で発話されやすいことを踏まえて、語りの文節終端、あるいは、語りの文節終端から 200ms ポーズが継続した点を傾聴応答の表出率を予測するタイミングとした。語りの音響情報とテキスト情報から、表出率を予測する。音響情報としては、対象の応答タイミング直前のフレームの系列を用いた。テキスト情報としては、対象の応答タイミング直前のトークンの系列を用いた。トークンは、語りのサブワードと応答タイミングを表現する 4 種のタグから構成される。トークン単位の特徴量には、トークンの埋め込み表現

を用いた。手法では、音声分析ツールを使用して、対象の応答タイミング直前のフレーム系列から、音響情報の特徴量行列を抽出する。一方、対象の応答タイミング直前のトークン系列をトークン埋め込み層に入力して、テキスト情報の特徴量行列を得る。次に、これら特徴量行列に対して、positional encoding の加算と、線形変換、ReLU 関数を適用したのち、1 層の transformer encoder による変換を行う。その後、変換後の行列に対して、要素ごとに平均値をとり、音響情報とテキスト情報それぞれのエンコード結果であるベクトルを得る。最後に、これら二つのベクトルを連結させたのち、線形変換と sigmoid 関数を適用し、0 以上 1 以下の 1 次元の値に変換することで、表出率を得る。

実験では、テストデータ中の全応答タイミングから算出される RMSE (Root Mean Squared Error) により、表出率の予測性能を評価した。さらに、表出率の正解値ごとに RMSE を算出し、これも評価指標とした。提案手法に加え、学習データにおける表出率の出現分布に従ってランダムに表出率を予測する手法、音響情報のみを使用して表出率を予測する手法、テキスト (Text) 情報のみを使用して表出率を予測する手法を実装した。実験の結果、提案手法とランダム手法を比較すると、提案手法が全ての評価指標で上回っており、提案手法の有効性を確認できた。音響情報のみ手法と言語情報のみ手法についても、同様の結果が得られた。このことは、音響情報とテキスト情報がいずれも表出率の予測に寄与することを示している。一方、提案手法と両手法を比較すると、提案手法が全 12 個の評価指標のうち 9 個で最良の結果を示しており、音響情報とテキスト情報の両方を使用して表出率を予測することの有効性を確認できた。また、分析の結果、提案手法は、正解の表出率が高いほど予測値も高くなる傾向にあることも確認した。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計12件（うち招待講演 0件 / うち国際学会 4件）

1. 発表者名 Koichiro Ito, Masaki Murata, Tomohiro Ohno, Shigeki Matsubara
2. 発表標題 Relation between Degree of Empathy for Narrative Speech and Type of Responsive Utterance in Attentive Listening
3. 学会等名 The 12th Language Resources and Evaluation Conference (LREC-2020) (国際学会)
4. 発表年 2020年

1. 発表者名 伊藤滉一朗, 村田匡輝, 大野誠寛, 松原茂樹
2. 発表標題 語りに傾聴を示す応答の表出されやすさの推定
3. 学会等名 言語処理学会第27回年次大会
4. 発表年 2021年

1. 発表者名 村田 匡輝, 大野 誠寛, 松原 茂樹
2. 発表標題 語りの傾聴において表出する応答データの拡充
3. 学会等名 言語処理学会第26回年次大会発表論文集
4. 発表年 2020年

1. 発表者名 伊藤 滉一朗, 村田 匡輝, 大野 誠寛, 松原 茂樹
2. 発表標題 語りの傾聴における共感の深さと応答の種類の関係
3. 学会等名 言語処理学会第26回年次大会発表論文集
4. 発表年 2020年

1. 発表者名 小杉 駿介, 伊藤 滉一朗, 村田 匡輝, 大野 誠寛, 松原 茂樹
2. 発表標題 傾聴において表出する評価応答に対応した語りの語句
3. 学会等名 情報処理学会第82回全国大会
4. 発表年 2020年

1. 発表者名 堀上 七瀬, 村田 匡輝, 松原 茂樹
2. 発表標題 話し手の語りに傾聴を示すシステムの開発
3. 学会等名 電気・電子・情報関係学会東海支部連合大会
4. 発表年 2019年

1. 発表者名 小杉 駿介, 伊藤 滉一朗, 村田 匡輝, 大野 誠寛, 松原 茂樹
2. 発表標題 語りに傾聴を示す応答の表現とその表出効果
3. 学会等名 電気・電子・情報関係学会東海支部連合大会
4. 発表年 2019年

1. 発表者名 伊藤 滉一朗, 村田 匡輝, 大野 誠寛, 松原 茂樹
2. 発表標題 傾聴を示す応答で繰り返される語りの語句の検出
3. 学会等名 言語処理学会第25回年次大会
4. 発表年 2019年

1. 発表者名 Koichiro Ito, Masaki Murata, Tomohiro Ohno, Shigeki Matsubara
2. 発表標題 Estimating the Generation Timing of Responsive Utterances by Active Listeners of Spoken Narratives
3. 学会等名 2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU 2021) (国際学会)
4. 発表年 2021年

1. 発表者名 Shunsuke Kosugi, Koichiro Ito, Masaki Murata, Tomohiro Ohno, Shigeki Matsubara
2. 発表標題 Identification of Important Utterances in Narrative Speech Using Attentive Listening Responses
3. 学会等名 16th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP 2021) (国際学会)
4. 発表年 2021年

1. 発表者名 小杉駿介, 伊藤滉一朗, 村田匡輝, 大野誠寛, 松原茂樹
2. 発表標題 傾聴を示す応答を用いた語りの重要箇所の同定
3. 学会等名 第20回情報科学技術フォーラム講演論文集 (FIT2021)
4. 発表年 2021年

1. 発表者名 Koichiro Ito, Masaki Murata, Tomohiro Ohno, Shigeki Matsubara
2. 発表標題 Construction of Responsive Utterance Corpus for Attentive Listening Response Production
3. 学会等名 The 13th Language Resources and Evaluation Conference (LREC-2022) (国際学会)
4. 発表年 2022年

〔図書〕 計0件

〔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
研究 分担者	大野 誠寛 (Ohno Tomohiro) (20402472)	東京電機大学・未来科学部・准教授 (32657)	
研究 分担者	村田 匡輝 (Murata Masaki) (30707807)	豊田工業高等専門学校・情報工学科・准教授 (53901)	

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
---------	---------