様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　元 年　 ６ 月　 ６ 日現在

機関番号：１３９０１
研究種目：挑戦的萌芽研究
研究期間：2015～2018
課題番号：１５Ｋ１２０９５

研究課題名（和文）話し手の語る意欲を高めるパーソナルロボットの傾聴技術

研究課題名（英文）Attentive hearing with personal robot to motivate speakers

研究代表者
松原　茂樹（Matsubara, Shigeki）

名古屋大学・情報連携統括本部・教授

研究者番号：２０３０３５８９

交付決定額（研究期間全体）：（直接経費）　　2,700,000円

研究成果の概要（和文）：相槌生成タイミングを検出するための対話コーパスの作成とその利用方式を考案した。対話データに網羅的かつオフラインで相槌タグを付与することにより、揺れの小さいコーパスを作成できる．実験により，本方式の有効性を確認した。また、傾聴を示す多様な応答の収集を実施した。語りの音声に対し、表出するに相応しい応答の表現と生成タイミングを付与するという収集方式を採用した。本方式により自然で多様な応答を効率的に収集できることを確認した。

研究成果の学術的意義や社会的意義
ロボットとの対話では、対話の過程が快適であることも重要である。ロボットの認識や理解の状態を、ユーザに適宜フィードバックすることが有効であり、聴覚的フィードバック手段である相槌は重要な役割を果たす。本研究で、適したタイミングでの相槌生成の実現性を示した。一方、語ることは人間に備わる基本的な欲求といえる。人が語れる機会を増やすことは現代社会の重要な課題に対し、パーソナルロボットが語りを聞く役割を担うことが考えられる。本研究では、ロボットによる傾聴を実現するための効果的な応答データ収集を提案した。

研究成果の概要（英文）：We developed a method of creating and using a dialogue corpus to detect the timing of generating back-channel feedback. It is possible to create a coherent corpus by tagging dialogue data with back-channel feedback comprehensively. The effectiveness of this method was confirmed by experiments. In addition, we conducted a collection of responsive utterances showing attentive hearing. We adopted a collection method of giving appropriate response expressions and their generation timing to narrative speech. It was confirmed that natural responses can be efficiently collected by this method.

研究分野：自然言語処理

キーワード：音声言語処理　音声対話　応答生成　傾聴　コーパス　会話ロボット　語り　発話タイミング

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９、ＣＫ－１９（共通）

１．研究開始当初の背景

(1) 人との会話を目的とするコミュニケーションロボットの製品化が進んでいるものの、依然として我々の生活に広く浸透するには至っていない。人々を惹きつける存在であっても、家庭や職場などで日常的な話し相手を担うケースは稀である。

(2) 上述した状況は、現行のロボットが備える音声認識や言語理解などの性能が十分でないことのみに起因しない。会話とは、共通の話題に関して話したり聞いたりする行為であり、人にとって快適であることもあれば不快であることもある。例えば、「聞いていない人に話すこと」「関心のない話しを聞くこと」には苦痛を感じる。一方で、現行のコミュニケーションロボットの会話機能は、「ロボットとの会話は楽しい」という前提のもと実現されており、コミュニケーションを取る労力や時間､心的負担などはほとんど考慮されていない点に大きな原因がある。

(3) 代表者らは、本研究を開始した時点で、データに基づく対話研究の経験を有しており、音声対話システムの応答技術でいくつかの成果を得ていた。これらのノウハウを上述の問題の解決に活用できると考え、本研究を着想するに至った。なお、関連する研究動向として、ユーザの発声中に相槌や頷きを行う対話エージェントの研究が国内外で行われていたものの音声対話処理を円滑化する目的で実現されており、話し手の語る行為にアプローチするものではないという状況にあった。

２．研究の目的

話し手が語りたくなる傾聴方略をデータに基づき体系化し、そのような方略を備えたパーソナルロボットの実現性を明らかにすることを目指し､「傾聴に関わる応答として相槌を対象に、話し手の語る意欲を高める生成法を構築し、その有効性を実験的に検証すること」、および、「話し手の語り音声に対する聞き手の傾聴を示す応答を収録・整備し、相槌や共感等の傾聴に関わる応答方略を分析すること」の２点を目的に本研究を実施した。

３．研究の方法

(1) 音声対話システムによる相槌生成機能を実現するにあたり、相槌の種類の選定、及び、相槌の生成タイミングの判定、が課題となる。これらを解決する上で、実データを観察し、得られた知見を活用することは有用である。前者の課題に関して、これまでにいくかつの分析が行われ、相槌が打たれる状況と相槌の種類との関係について整理されている。例えば、聞き手が話し手の発話の聞き取りに成功している場合には、「はい」､「うん」などを下降調で発し、聞き取れない場合や理解できない場合は､「え」､「は」などを上昇調で発する。また､「はい」と「うん」など、同じ分類に属する相槌の使い分けは、話し手と聞き手の関係性が影響する。このため、システムは、自らの認識や理解の状況、並びに、ユーザに対する自らの立場に応じて、生成する相槌を選定することになる。一方、後者の課題についても分析が行われ、幾つかの知見が得られている。例えば、通常の対話では、相槌の中でも特に「はい」など、聞き取りの成功を伝える相槌が多数を占めることになるが、それらは音響的な観点からは､「ポーズが生じたとき」「音声のパワーが低下したとき」､また、言語的な観点からは「節のあと」「接続詞や終助詞のあと」などで打たれやすいといった特徴が知られている。しかし、相槌が発生しやすいタイミングを定性的に把握したとしても、それでもって任意のタイミングについて、そこが適切な相槌生成タイミングか否かを判断できるわけではない。システムが不適切なタイミングで相槌を打つと、ユーザはむしろ、システムの認識や理解に疑念を抱くことになり、相槌本来の役割を果たせなくなる。音声対話システムとしては、不適切なタイミングでは相槌を一切打たず、適切なタイミングではむしろ積極的に打つことが、ユーザに安心感を与える上で有効な戦略であるといえる。以上より、本研究では、相槌生成の適切なタイミングの網羅的な検出に焦点を当てた。また、相槌としては、聞き取りに成功したことを示す相槌を対象とし、システムとユーザの一般的関係を考慮し、「はい」を用いるものとした。

① 相槌は､話し手の発話を聞き取ったことを､聞き手が話し手に伝える合図として機能するため、その生成タイミングは、発話に含まれるある構成素の直後となる。このため、対話における話者交代（発話権の移動）から次の話者交代（発話権の移動）までを対話ターンと呼ぶとき、対話ターンを構成素の列で表現する。ただし、構成素としては、形態素や文節、節など様々な単位が考えられる。相槌タイミングを網羅的に検出するとは、任意の構成素が入力された時点で、その直後に相槌を打つことが可能か否かを判定することである。これを統計的手法により実現するには、上記の判定のためのモデルを、対話コーパスを用いて学習し、それを既入力の構成素の列に適用することとした。一方、統計的手法による相槌タイミング検出のための対話コーパスとして、これまで人間による自然な対話を収録した音声データが用いられてきた。この場合、データに含まれる相槌は適切なタイミングで打たれていることを仮定している。しかし、このような自然な対話に含まれる発話の各構成素の直後を観察すると、実際には、A）相槌に適したタイミングで打たれている、B）相槌に適さないタイミングで打たれている、C）相槌に適したタイミングで打たれていない､D)相槌に適さないタイミングで打たれていない、という四つのケースが存在する。このうち、A)、D）のケースが多くを占めるのであれば相槌

生成タイミング検出のためのデータとして相応しいといえる。しかし、実際にはB)、C）のケースが少なからず発生する。このうち、B)が生じるのは、聞き手は話し手の発話を予測できるわけではなく、打った相槌が結果的に不自然となることがあるためである。ただし、その発生頻度は必ずしも高くない。一方、C）は頻出する。たとえ相槌に適したタイミングであっても、人間による対話では、相槌の有無が対話の進行を大きく左右することは少なく、実際に相槌を打つか否かは聞き手の嗜好や意図、気分などに強く依存するためである。すなわち、相槌生成タイミングの検出に人間の対話データを用いることの問題は、C）とD）の区別が難しいことにあり、これは、相槌が打たれていないときに、それが相槌タイミングでないことによるのか否かはデータからは明らかでないことによる。

この問題に対して、相槌に適したタイミングと適さないタイミングが網羅的に明示された対話コーパスを作成することが考えられる。これは、対話コーパスにおける、構成素の列である対話ターンについて、そのあらゆる構成素の直後に、相槌生成タイミングか否かが記されたデータを作成することを意味する。本研究では、上述のデータをオンライン環境、すなわち、対話において打たれた相槌発生タイミングを対話コーパスにタグ付けするのではなく、オフライン環境、すなわち、対象となる対話データ上で、相槌生成に適したタイミングを選定し、それを対話コーパスにタグ付けするというアプローチを採った。オフラインでのタグ付けのために、タグ付け対象の対話の音声を聴取、また、その文字化テキストを閲覧しながら、相槌位置を定めること、ならびに、タグ付けられた位置で相槌が打たれる対話音声が生成され、その再生音声を聴きながら、タグを推敲すること、が可能な環境を設けた。これにより、タグ付け作業者は、話し手や聞き手とは異なる第三者の立場から、相槌位置の適切さを吟味することができ、網羅的なタグ付けを安定的に（作業者による揺れが少ない形で）実行することが可能となる。以上のアプローチに従い、既存の音声対話コーパスを用いてタグ付け作業を実施し、相槌生成タイミング検出のためのコーパスを作成した。

② 作成したコーパスを用いて相槌生成タイミングを検出する手法を考案した。本手法は、1 対話ターンの基本区間列中の基本区間が連続して入力されることを想定し、基本区間が一つ入力されるごとに、その基本区間の終了直後に相槌を生成できるか否かを推定する。推定には、Support Vector Machine を用いた。ある基本区間の終了直後に相槌を生成するか否かを決定する際に利用する素性を定めた。このうち、節境界の検出には、節境界解析ツール CBAP を用いた。CBAP には 147 種類の節境界ラベルが存在している。平均発話速度は、そのドライバが形態素より前に発話した全形態素に対する平均発話速度とした。変動パターンは、基本区間の終了時間から遡って 100 ミリ秒を三つの区間に分け（窓枠 50 ミリ秒、フレームシフト 25 ミリ秒)、ピッチまたはパワーの値をもとに各区間を「上昇」「平坦」「下降」のいずれかに分類し、その列（例えば、平坦-下降-下降）を変動パターンとした。ピッチとパワーは、音声分析ツールを使用して求めた。なお、利用した素性は全て、基本区間列から抽出可能な情報であり、各基本区間の言語情報や音響情報、時間情報は、その基本区間の入力が終わり次第得られるものとした。

(2) 傾聴的応答を適切に表出することにより、語りの傾聴を話し手に伝達する機能の実現を目指している。情報機器が応答を生成するにあたり、応答表現の選択と生成タイミングの決定という 2 点を解決する必要がある。生成の効果を高めるという観点から、「自然でかつ多様な応答表現であること」および「生成の頻度が高くかつ自然なタイミングであること」が要件となる。上記の要件を満たす生成方式が解明されていない現状では、傾聴的応答の実データを集め、観察・分析を通して有用な知見を蓄積することが肝要である。

① 語りに対する傾聴的応答の自動生成に向け、本研究では、自然でかつ多様な傾聴的応答データを収集した。収集の方針収集方法として、話し手と聞き手によるリアルタイムでのやり取りを記録し、聞き手の発話から傾聴的応答を取り出すことが考えられる。しかしこの方式では、聞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。そこで本研究では、語りデータに対して作業者が応答データを付与することとした。具体的には、語り音声の再生に同期して作業者が傾聴的応答を発声し、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た。双方向でのやり取りがない分、作業者は傾聴的応答の効果的な表出に集中できるという利点がある。

② 収録した応答の多頻度性、多様性、自然さを評価した。応答の多頻度性としては、傾聴的応答の発生間隔を調査した。傾聴的応答の多様性の評価では、応答の種類（文字列の異なり）に関する多様度指数、及び、相槌以外の傾聴的応答の出現割合を調べた。また、収集した傾聴的応答の表現及びタイミングの自然さを評価するために、被験者実験を実施した。

４．研究成果

(1) 相槌コーパスの作成と相槌生成タイミング検出の実験により、以下の成果が得られた。

① 相槌コーパスでは、5,416 の相槌タグが付与された。基本区間の 4.8%(5,416/113,172) の割合で相槌が出現していた。なお、タグ付けの対象となった車内音声対話コーパスを観察したところ、同一のドライバ発話に対して実際にオペレータが打った相槌は 73 回であり、基本区間に

対する割合は 0.065%(73/113,172) に過ぎなかった。両コーパスにおける相槌回数の差は著しく、既存のコーパスに相槌タグを網羅的に付与することの効果を示している。全ての相槌タグのうち、200msec 以内の無音区間に付与された相槌タグの割合は 33.1% (1,790/5,416)、200msec を超える無音区間は 44.2% (2,396/5,416) であり、残りの 22.7% (1,231/5,416) は形態素区間に付与されていた。また、文節境界の基本区間に付与された相槌タグの割合は 90.0% (4,872/5,416) であり、残りは文節の途中に付与されたものであった。ここで、文節境界の基本区間か否かは、直前の形態素区間が文節の最終形態素であるか否かで判定した。このように、無音区間以外、あるいは、文節境界でない位置でも、相槌生成に適したタイミングとなるケースが多く存在している。オフライン環境でのタグ付けによるコーパス作成により、作業者による揺れが少ないタグ付けが実施されていることを確認するために、タグ付け実験を実施した。実験では、コーパス作成対象の対話データに含まれる 358 対話ターンに対して、タグ付け作業者とは異なる被験者 3 名が各々、同様のタグ付けを実施した。作業者間での一致度が高ければ、それはコーパスが高い安定性を備えていることを意味する。Cohen の kappa 値を用いて、各 2 者間の一致度を測定した。比較のために、オンライン環境で打たれた相槌をタグ付けしたコーパス（以下、比較コーパス）を用いた。比較コーパスは、本コーパスと同じく、車内音声対話コーパスに収録されたドライバ音声の 297 対話ターンをタグ付け対象としている。具体的には、再生されたドライバ音声に対し、その聞き役である被験者が相槌を発声し、その発声がドライバ音声中のどの基本区間で開始されたかを対応付けることによりコーパスを作成している。4 名の被験者を設け、被験者には、不自然でないタイミングにできる限り網羅的に相槌「はい」を打つように指示し、また、事前に相槌の発声練習を実施している。なお、タグ付けにおけるドライバ音声の基本区間への分割は、本コーパスと同様に行った。実験の結果、比較コーパスでは、最も低い被験者間で kappa 値が 0.167、高い場合でも 0.536 であったのに対し、本コーパスでは、いずれの作業者間でも実質的に一致している水準にあり、オフライン環境での相槌タグ付与の優位性が示された。

②相槌生成タイミング検出の実験を実施した。実験は交差検定により実施した。具体的には、コーパスを、各グループのドライバ数がほぼ均等（34 名程度）になるように 10 分割し、そのうちの 1 グループをテストデータ、残りの 9 グループを学習データとした実験を 10 回繰り返した。ただし、10 グループのうち 1 グループは、デベロップメントデータとして使用したため、残りの 9 グループに対する実験結果のみを評価対象とした。構築した対話コーパスにおける相づち位置を正解とし、正解に対する再現率、及び、適合率により評価した。 評価対象の全データに対する本手法の適合率、再現率はそれぞれ、72.9%(3,146/4,316)、64.6%(3,146/4,870) であった。また、これらの調和平均である F 値は 68.5%であった。本手法による検出結果と正解データが全ての基本区間で一致した対話ターンは 8,009 ターン存在した（全対話ターンの 80.4%)。また、正解データにおいて相槌タグが 1 回以上付与された対話ターンは 3,162 個あり、その 47.1% (1,489/3,162)の対話ターンにおいて、本手法が全ての基本区間で正解した。使用した素性の効果を調査するため、全素性から一つの素性を取り除いて相槌タイミングを検出する実験を実施した。本手法と各素性を除去した場合との間で t 検定を行ったところ、文節境界、節境界、無音区間、直前の相槌からの経過時間に関する素性が、相槌生成タイミングの検出に寄与することを示した。

(2) 傾聴を示す応答発話を収集し、作成したデータを分析した結果、以下の成果が得られた。
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。コーパスには、30 名の高齢者による一人約 20 分の語りの音声が収録されている。全高齢者共通の 10 個の質問に対し、その回答を独話として語るという収録形式が採用されている。応答音声の発声は、高度な接客スキルを要する業務経験を有する作業者 1 名が担当した。作業者は、再生された語り音声に対しリアルタイムに応答した。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。なお、人が知覚できるポーズで分割された音声を発話単位と定めている。発話単位を形態素解析し、各形態素に発声時間を付与した。

② 応答の多頻度性として、収録データにおける傾聴的応答の発生間隔は、1.9 秒に 1 度という出現率であった。比較のため、雑談会話を収録した名大会話コーパスにおける傾聴的応答に相当する発話の出現率を調査したところ、12.4 秒に 1 度であった。このことは、本収集における傾聴的応答の多頻度性を示している。応答の多様性として測定した応答あたりのエントロピーは、収集データにおける応答あたりのエントロピーは 5.42 であった。名大会話コーパスに対しても同様に測定したところ、エントロピーは 4.86 であり、本データにおける応答の多様性の高さを確認した。次に、応答タイプの多様性を調べるために、データから 4,885 個の傾聴的応答を無作為に抽出し分類した。相槌以外の傾聴的応答も全体の 35.6%出現しており、多様なタイプの応答を収集できた。一方、被験者実験では、高齢者 5 名分の語り音声とその応答音声 (2,797 発話単位) のステレオ音声を使用した。被験者は、19～20 歳の男女計 5 名であり、各応答に対し自然か否かを判定した。自然でないと判定された応答は、被験者平均で 24.4 個であり、全体の 99.1%を自然な応答が占めた。

５．主な発表論文等

〔雑誌論文〕（計２件）

① 村田 匡輝, 大野 誠寛, 松原 茂樹、語りの傾聴を話し手に示す応答発話の収集、電気学会論文誌Ｃ、査読有、138 巻、5 号、2018、637-638
https://doi.org/10.1541/ieejeiss.138.637
② 大野 誠寛, 神谷 優貴, 松原 茂樹、対話コーパスを用いた相づち生成タイミングの検出、電子情報通信学会論文誌、査読有、138J100-A 巻、1 号、2017、55-65

〔学会発表〕（計６件）

① 村田 匡輝, 大野 誠寛, 松原 茂樹、系列変換モデルに基づく傾聴的な応答表現の生成、言語処理学会第 24 回年次大会、2018
② Tomohiro Ohno, Masaki Murata, Shigeki Matsubara, Collection of Responsive Utterances to Show Attentive Hearing Attitude to Speakers, The 11th International Conference on Ubiquitous Information Management and Communication, 2017
③ 村田 匡輝, 大野 誠寛, 松原 茂樹、話し手の語りに傾聴的な応答の収集、言語処理学会第 23 回年次大会、2017
④ 村田 匡輝, 大野 誠寛, 松原 茂樹、話し手への傾聴を示す応答発話の収集と分析、情報処理学会第１８３回知能システム研究会、2016
⑤ 田口 諒弥, 村田 匡輝, 松原 茂樹、対話ログからのマインドマップ生成のための検討、情報処理学会第 78 回全国大会、2016
⑥ 村田 匡輝, 大野 誠寛, 松原 茂樹、会話ロボットにおける繰り返し応答の生成、第 14 回情報科学技術フォーラム、2015

〔図書〕（計０件）

〔産業財産権〕
○出願状況（計０件）
○取得状況（計０件）

〔その他〕

６．研究組織

(1)研究分担者

研究分担者氏名：中村 剛士
ローマ字氏名：Tsuyoshi Nakamura
所属研究機関名：名古屋工業大学
部局名：工学研究科
職名：准教授
研究者番号（8 桁）：90303693

研究分担者氏名：大野 誠寛
ローマ字氏名：Tomohiro Ohno
所属研究機関名：東京電機大学・
部局名：未来科学部
職名：准教授
研究者番号（8 桁）：20402472

研究分担者氏名：村田 匡輝
ローマ字氏名：Masaki Murata
所属研究機関名：豊田工業高等専門学校
部局名：情報工学科
職名：准教授
研究者番号（8 桁）：30707807

(2)研究協力者

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。