研　究　報　告　書

「発話行動の階層的理解に基づく相互適応型音声インタラクション」

研究タイプ：通常型
研究期間：　平成22年10月～平成26年3月
研 究 者：　駒谷　和範

1.　研究のねらい

音声対話における発話の理解には、その言語表現（テキスト）の理解にとどまらず、発話という行為自体の包括的な理解が必要である。これまでの対話システムに関する研究の多くでは、入力されてくるテキスト（音声の場合では音声認識結果）のみを処理の対象としている。これは音声インタラクションの一面をモデル化しているに過ぎない。音声対話における発話という行為には、その言語情報だけではなく、それが行われたこと自体やそのタイミング、さらには発話が行われた状況も重要な情報である。

本研究では、従来の音声認識結果に偏重した音声対話理解を越え、頑健な音声対話システムの実現を目指す。特に、ヒューマノイドロボットとの音声対話を、ロボット自身に備え付けられたマイクを通じて実現する場合には、周辺雑音の混入や、話者の口元にマイクがないことなどにより、音声認識性能の劣化が著しい。音声認識誤りに頑健な音声対話を実現するには、発話の包括的な理解が不可欠である。つまり、発話行為理論（Speech Act Theory）における発語内行為レベルの情報だけでなく、発語行為レベルや発語媒介行為レベルの情報を、音声発話を解釈する場合に考慮することにより、頑健な音声対話の実現を目指す。

さらに、音声言語を用いたインタラクションではユーザ適応が本質的に必要である。なぜならどのようなユーザに対しても一様な応答を行うのは非協調的だからである。例えば、ユーザが既に知っている内容について、長々と説明を行うのは協調的ではない。つまりシステムがユーザに合わせて応答することが望まれるが、その一方で、ユーザがシステムの発話に応じて、自身の発話表現を変えるという現象も観測されている。本研究では、音声対話システムにおけるこのような双方向の適応に着目する。このようなユーザの適応に関する知見を蓄積することにより、相手に応じた音声対話システムの実現を目指す。

2.　研究成果

（1）概要

本研究ではまず、発話行動の階層的理解というコンセプトを提案した。発話の階層的理解の概念図を図1に示す。近年、Apple の「Siri」や NTT ドコモの「しゃべってコンシェル」など、スマートフォン上のアプリが多くのユーザにより使用され始めているが、これらにおける音声対話は、この階層の中の言語レイヤのみを扱っている。一方、ヒューマノイドロボットとの音声対話を、ロボット自身に備え付けられたマイクを用いた場合でも、頑健に実現するには、図 1 における言語レイヤ以外、つまり、社会レイヤや信号レイヤでも、ユーザとロボットの対話がかみ合っている必要がある。

本研究では、この各階層での情報を見出し、音声対話システムで考慮することで、ロボットとの音声対話を雑音に対してより頑健にできることを新たに示した。これは、実際にロボットを使った音声対話システムの開発を通して得られたものであり、対話システムの言語的側面のみに着目している従来研究では得られ難い知見である。具体的には、まず社会レイヤにおける制約を新たに「話しかけられやすさ」としてモデル化し、これを利用して雑音を棄却できることを示した。次に、信号レイヤでの齟齬に起因する誤りの存在を指摘し、これから生じる 2 つの誤りをそれぞれ修復する手法を提案して、これをシステムに実装した。さらに、ユーザとシステム間の適応は、このそれぞれのレベルでも観測される。本研究ではまずこのうちの言語レイヤにおける相互適応について調査を行った。

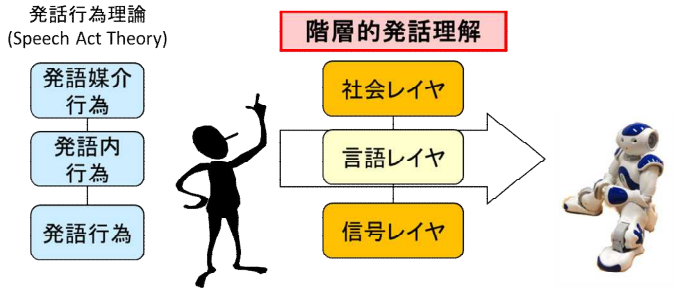

図１： 階層的発話理解の概略

（2）詳細

研究テーマ①:「社会レイヤにおける制約のモデル化と利用」

人間同士の対話には、対話者同士が無意識のうちに守っているルールが存在する。例えば、「言いたいことがあっても相手が話している間には話し始めない」ことや、「相手の話を聞く際には相手の方を向く」こと等である。我々は、このように人間が対話相手の状態を考慮して話しかけるという社会的規範が、ヒューマノイドロボットに話しかける際にも同様に成り立つと考え、ロボットの状態に基づき、ユーザが各時点で話しかけられると感じるか否かを予測するモデルを構築した（図 2）。

従来研究では、対話システムは常に受動的に、入力される情報を処理する。これに対して本研究は、システム（ロボット）の状態を考慮に入れ、これがユーザに与える影響を踏まえて、入力音を解釈する枠組みを新たに示したものである。

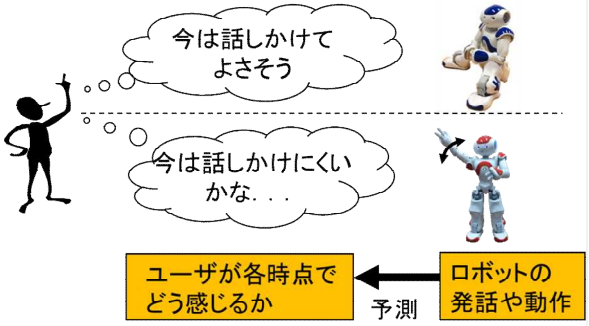

図２： 話しかけやすさの予測の概要

ロボットの状態は、その姿勢や動作、直前の発話内容を特徴として表現した。ロボットが任意の時点での話しかけられやすさを予測できれば、協調的なユーザがロボットに話しかけないタイミングを、ロボットが知ることができる。これにより、そのようなタイミングでの入力音は雑音等である可能性が高いとみなし、これを棄却するなど、社会レイヤにおける制約を、頑健な音声対話の実現に利用できる。

具体的には、ヒューマノイドロボットの挙動列に対し、本研究室の学生 3 名を被験者としてデータ収集を行い、ユーザが話しかけやすいか否かを予測可能であることを示した。まずユーザが実際に話しかけやすい状態か否かを付与した学習データを、マウスをクリックさせることで作成した。その後、機械学習により、各時点での話しかけられやすさを予測するモデルを

構築し、性能を検証した。この結果、ユーザが話しかけやすいと感じたタイミングを、87。4%の精度で予測できることを示した。

さらに、本モデルが、個人差や実験前に与えた教示に適応可能であることを示した。この成果は、本モデルを実際の対話システムに適用する際に必要である。

研究テーマ②:「信号レイヤにおける齟齬の修復」

音声対話システムでは、ユーザが話している最中にも関わらず、システムが話し始めてしまう場合がある。この問題は、ユーザの発話区間の誤検出により生じる。我々はこの現象を「発話の誤分割」と呼んでいる。これは、信号レイヤでの齟齬に相当する。

従来の音声対話システムでは、そもそもこのような問題を扱っていなかったり、スマートフォン上での push-to-talk(押してから話す)ボタンを使用して回避したりしている。これに対してロボットとの対話では、そのようなボタンは利用できないため、この信号レイヤの問題を解決する手法が必要である。本研究ではこの問題を明示的に指摘したうえで、これにより生じる 2 つの問題、つまり音声認識誤りとシステムの不適切な発話タイミングの両方を、解決する手法を新たに開発した。

音声認識誤りを修復するには、正しい発話区間に対して(再度)音声認識を行う必要がある。まず、修復が必要である場合に、誤分割された発話断片対を結合し、再度音声認識を行うことで正しい音声認識結果を得る手法を開発した。さらに、この修復が必要か否かの判定をより高精度に行うために、発話断片対から得られる様々な特徴を用いて、修復必要性の判定を二値分類問題として機械学習により判定する手法も開発した。この際に、有効かつドメインに依存しない特徴のみを用いるために、2 種類の特徴選択法を適用した。これにより、ベースラインでは 77。7%であった判定精度が、85。6%に向上した。

不適切なターンテイキングの修復にも取り組んだ。発話の誤分割時にシステムが誤って話し始めるのを防ぐルールを記述し、MMDAgent(名古屋工業大学で開発)上に実装した。発話の修復時に再度音声認識を行うため遅延が生じるが、システムにフィラーを生成させることで、不自然な間が生じることを防ぐようにシステムを実装した。

研究テーマ③:「対話システムにおける語彙の適応の調査」

人間同士の対話において、使用する表現を対話中に相手と同調する現象(lexical entrainment)が知られている。これは、同一の対象に対して複数の呼び方がある場合に観測できる現象である。このため本研究では、音声対話システムにおける語彙の同調を確認し知見を蓄積するために、まず簡略表現を認識・理解し、それを応答に用いる音声対話システムを構築した。ここでの簡略表現とは、例えば「ファミリーマート」に対する「ファミマ」のように、名称の一部を省略し短くした表現を指す。システムのタスクは、名古屋地区のコンビニやファーストフード店 3068 件に対する検索である。

従来の、システムを用いた語彙の同調の研究では、システムが用いた表現を、ユーザが使用するかどうかを調査するものが多い。これに対して本研究は、双方向の同調現象に着目した点に特色がある。つまり、ユーザがシステムの表現に同調する場合の調査に加えて、システムがユーザの表現に同調する機能を実装し、その場合にユーザの印象に与える影響も調査した。

被験者 33 名に対する評価実験の結果、システムが店の名前をチェーン名や支店名で参照した場合のうち 83%で、システムと同じ表現をユーザが用いることが確認された。この結果は、ユーザ発話を音声認識が容易な語彙に誘導したり、ユーザが使用する可能性が高い語彙に高い言語モデル確率を与えたりすることにより、音声認識誤りの防止に役立つ。さらに、ユーザが使用した語彙をシステムが使用した場合のユーザの印象を調査したところ、70%の被験者はシステムのこの機能に気づかず、残りの被験者にも概ね好感を持っていたことから、この機能は違和感なく受け入れられていたことが示唆された。今後、このような機能が、システムとの社会的な関係性の構築に有用となるかどうかを調査するうえでの糸口を得た。

3. 今後の展開

本研究では、音声対話システム、特にロボットとの音声対話を行う際に有効となる情報を新たに取得し、音声対話に用いる手法を提案した。近年、ロボティクス技術の進展に伴い、ロボットと音声対話を行うことに対するニーズや期待が高まっている。実環境での音声認識には様々な困難があることから、頑健なロボット対話を実現するためには、本研究により新たに提案したような、言語レイヤの情報を補う情報が不可欠である。

今後の展開として、以下が挙げられる。まず、本研究では、階層的発話理解を掲げ、社会レイヤや信号レイヤにおける制約の利用や問題の解決を提案したが、本研究で提案したものが、そのレイヤから得られる情報の全てというわけではない。階層的発話理解という枠組みの中で、他にも捉えるべき現象や、利用可能な情報がないかどうか、引き続き検討が必要である。

相互適応については、本研究では音声対話システムにおける語彙の双方向の適応を実現・検証するに留まった。ねらいとしては、語彙つまり言語レイヤのみに留まらず、社会レイヤや信号レイヤでの適応も捉えることを考えていたが、そこまでは至れなかった。今後、言語レイヤで得た知見や方法論を生かして、社会レイヤや信号レイヤでの、ふるまいの適応の検証や実現に展開したい。

最終的には、実環境で動作するロボットにおける実証実験へと展開する必要がある。雑音がある実環境下でも、ロボットと対話すること自体のニーズは大きい。音声認識技術そのものの発展も取り入れながら、より適切な要素統合を行い、最終的により頑健な対話システム構築を目指したい。

4. 評価

(1)自己評価

音声対話が必須となる状況設定のひとつとして、ヒューマノイドロボットとの音声対話に新たに着手した。新しいロボットの使用したシステム構築に試行錯誤が必要であったため、時間がかかってしまった。この結果として、階層的発話理解の部分については、社会レイヤ、信号レイヤの両方に関する新たな手法を開発できたものの、それらを相互適応という観点までは進められなかった。ただ、音声インタラクションに含まれる、言語情報以外も活用するという視点自体は、今後も持ち続けるべきものを提案できたと考えている。

本研究の問題点として、個別の要素に関しては提案手法による性能改善が示されているが、これらの性能改善が音声対話全体にどの程度の影響があるのかが明らかでない点が挙げられ

る。音声対話システムは様々なモジュールから構成されるため、一部のモジュールの性能変化が全体の性能にクリアに反映されるとは言い難い。例えば、別のモジュールの性能がボトルネックになり、全体の性能には反映されないということがしばしば生じる。またシステム構築に労力が必要であることから、適切なタスク設定も必要である。このような困難があるものの、提案手法のインパクトをより明確に示すには、統合されたシステムにおいて、その有効性を示す必要がある。

より印象的なデモの見せ方も工夫する必要がある。エラーハンドリングをテーマとしている以上、「賢い応答ができる」ではなく、「誤動作せずに普通に動く」ことが研究成果となるが、その中でも、研究の良さをどのように見せるのかについて、さらに工夫を講じたい。「困難な客観的評価に拘泥するのではなく、主観的に素晴らしいデモを見せるべき」というコメントは、これからも念頭に置いて研究を進めたい。

（2）研究総括評価

本研究は、これまで音声認識技術に頼ってきた音声対話の理解を、発話行為の分析技術を加えることによって、大きく前進させようとする試みである。そのために、音声対話を社会レイヤ、言語レイヤ、信号レイヤに階層化し、これまで研究の対象となり難かった社会レイヤに着目し、システムの話しかけられやすさを評価・向上させる技術を開発している。また不適切なターンテイキングを検出することで誤分割された発話断片を結合し、正しい発話区間を得る手法を開発している。さらに人のシステムに対する同調行動に着目し、音声認識が容易な語彙に発話を誘導する技術も提案している。このような、人の発話行為を分析することで音声対話の理解を進める研究には新規性があり、その方向性は高く評価できる。今後も、ヒューマノイドロボットとの対話など、将来の情報環境での音声対話を現実のものとするよう、研究を継続していくことを期待する。

5. 主な研究成果リスト

（1）論文（原著論文，査読付国際学会プロシーディング）発表

1. Kazunori Komatani, Kyoko Matsuyama, Ryu Takeda, Tetsuya Ogata, Hiroshi G. Okuno: Evaluation of Spoken Dialogue System that uses Utterance Timing to Interpret User Utterances. Proc. IWSDS, pp.315--325, 2011.
2. Kazunori Komatani, Mikio Nakano, Masaki Katsumaru, Kotaro Funakoshi, Tetsuya Ogata, Hiroshi G. Okuno: Automatic Allocation of Training Data for Speech Understanding based on Multiple Model Combinations. IEICE Transactions on Information and Systems, Vol.E95-D, No.9, pp.2298-2307, 2012
3. Kazunori Komatani, Akira Hirano, Mikio Nakano: Detecting System-directed Utterances using Dialogue-level Features. Proc. Interspeech, 2012.
4. Kazunori Komatani, Shojiro Mori, Satoshi Sato: Constructing Language Models for Spoken Dialogue Systems from Keyword Set. Proc. IEA/AIE-2013, Contemporary Challenges and solutions in Applied Artificial Intelligence, Studies in Computational Intelligence, Vol. 489, pp.69--76, 2013.

5. Kazunori Komatani, Naoki Hotta, Satoshi Sato: Restoring Incorrectly Segmented Keywords and Turn-Taking Caused by Short Pauses. Proc. IWSDS, pp.27-38, 2014.

（2）特許出願

なし

（3）その他の成果（主要な学会発表、受賞、著作物、プレスリリース等）

主要な学会発表

1. 駒谷 和範, 松山 匡子, 武田 龍, 高橋 徹, 尾形 哲也, 奥乃 博: 発語行為レベルの情報をユーザ発話の解釈に用いる音声対話システム. 情報処理学会論文誌, Vol.52, No.12, pp.3374-3385, 2011.
2. Taichi Nakashima, Kazunori Komatani, Satoshi Sato: Integration of Multiple Sound Source Localization Results for Speaker Identification in Multi-party Dialogue System. Proc. IWSDS, 2012. (also published from Springer: Natural Interaction with Robots, Knowbots and Smartphones, pp. 153-165, 2014).
3. Takaaki Sugiyama, Kazunori Komatani, Satoshi Sato: Predicting When People will Speak to a Humanoid Robot. Proc. IWSDS, 2012. (also published from Springer: Natural Interaction with Robots, Knowbots and Smartphones, pp. 187-198, 2014)
4. 駒谷 和範, 中島 大一, 杉山 貴昭: ロボット自身のマイクを介した Nao との音声対話. 情報処理学会研究報告, Vol.2013-SLP-095, No.9, 2013. （査読無）
5. 杉山 貴昭, 駒谷 和範, 佐藤 理史: ヒューマノイドロボットが話しかけやすさを予測するモデルの構築. 人工知能学会論文誌, Vol.28, No.3, pp.255-260, 2013.
6. Tsugumi Otsuka, Kazunori Komatani, Satoshi Sato, Mikio Nakano: Generating More Specific Questions for Acquiring Attributes of Unknown Concepts from User. Proc. 14th Annual SIGDIAL Meeting on Discourse and Dialogue, pp. 70--77, 2013.
7. 杉山 貴昭, 駒谷 和範, 佐藤 理史: ロボットへの話しかけやすさモデルの評価と個人差や教示による変動への対応. 人工知能学会論文誌, Vol.29, No.1, pp. 32-40, 2014.
8. Takaaki Sugiyama, Kazunori Komatani, Satoshi Sato: Evaluating Model that Predicts When People will Speak to Humanoid Robot and Handling Variations of Individuals and Instructions. Proc. IWSDS, pp.62-72, 2014.
9. 秋田谷 樹, 駒谷 和範, 佐藤 理史: 音声対話システムにおける簡略表現の使用とそのユーザ発話への影響. 人工知能学会研究会資料, SIG-SLUD-B303-03, pp.15-21, 2014. （査読無）

招待講演および依頼講演

1. 駒谷 和範: “音声対話システム技術の現状と課題”, 電気関係学会東海支部連合大会シンポジウム「ここまでできる言語処理技術ー音声・言語情報処理の最先端ー」, 2012 年 9 月 25 日
2. Kazunori Komatani: “Hierarchical Utterance Understanding for Robust Human-Robot Spoken Dialogues”, International Workshop on Spoken Dialogue Systems (IWSDS2014),

2014 年 1 月 18 日

受賞

1. 人工知能学会 研究会優秀賞（2011 年 6 月受賞）

新聞報道

1. 中日新聞・ジュニア中日「心が通じる！？会話ロボット」（2013 年 8 月 18 日掲載）