The University of Osaka
Institutional Knowledge Archive

Title	Situation Understanding for Turn-Taking in Human-Robot Dialogue
Author(s)	杉山, 貴昭
Citation	大阪大学, 2017, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/61711
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

様式３

論　文　内　容　の　要　旨

氏　名　（　　杉山　貴昭　　）	
論文題名	Situation Understanding for Turn-Taking in Human-Robot Dialogue （人とロボットの音声対話における話者交替のための状況理解）

論文内容の要旨

本論文は、人とロボットの音声対話における話者交替のための状況理解に関する研究成果をまとめたものであり、5章より構成した。その内容を要約すると以下のようになる。

第1章では、本研究の目的と対象とする課題について述べ、本研究の課題解決に向けたアプローチを示した。具体的には、公共の場で音声対話可能な人型ロボットの実現を目的とし、この実現のためにロボットが対処すべき2つの課題として複数人対話処理と様々な入力音に対する処理を挙げた。また、これらの解決に向けて、社会的規範に基づくロボットによる対話状況の推定手法を提案した。

第2章では、ロボットによる対話状況の推定のために、ロボットの状態に着目した手法を提案した。具体的には、対話中にユーザが感じるロボットへの話しかけやすさを予測する手法について述べた。話しかけやすさを予測する枠組みとして、ロボットの状態を特徴に用いた2値判別モデルを提案した。話しかけやすさに寄与する特徴の設計と、学習データとして実際に感じるユーザの話しかけやすさを利用するために、ロボットの一連の動作に対して被験者が話しかけやすさを付与する実験を実施した。これにより、人間は話者交替時に相手の状況を考慮するという、社会的規範に基づく状況推定モデルを実現した。

第3章では、第2章で構築した、話しかけやすさの予測モデルが実際の音声対話で利用可能であることを示すために、このモデルを利用する以前に検討すべき2つの課題に取り組み、実際にこのモデルをロボット用音声対話システムに導入した。まず、2章で構築したモデルが特定のユーザに依存しないものであることを示すために、一般ユーザを対象に新たに被験者実験を実施し、話しかけやすさモデルを再構築した。再構築したモデルと2章で作成したモデルの性能を比較し、両モデルの性能が同等であることを確認した。また、このモデルが個人差やユーザの緊急度合に対して対応できることを実験的に確認した。最後に、このモデルを音声対話システムに導入し、有用な事例を紹介した。

第4章では、ロボットによる対話状況の推定のために、ユーザの状態に着目した手法を提案した。具体的には、公共の場でロボットが複数人のユーザと対話する際に、入力音に対してロボットが応答すべきか否かを推定する手法について述べた。この枠組みとして、ユーザの発話や動きに関する特徴を利用した機械学習手法を提案した。特徴として、人とロボットの音声対話時における人の特有の振る舞いを利用し、有用性を実験的に確認した。本手法により、公共の場で発生しうる様々な音に対して、ロボットが応答すべきか否かを適切に推定できることを示した。

第5章では、本研究全体を通して得られた主な成果をまとめ、本論文を総括した。

様式 7

論文審査の結果の要旨及び担当者

氏　名　（　杉山　貴昭　）			
論文審査担当者		（職）	氏　名
	主査	教授	馬場口 登
	副査	教授	佐藤　理史（名古屋大学大学院工学研究科）
	副査	教授	滝根 哲哉
	副査	教授	丸田 章博
	副査	教授	三瓶 政一
	副査	教授	宮地 充子
	副査	教授	井上　恭
	副査	教授	鷲尾　隆
	副査	准教授	新田 直子

論文審査の結果の要旨

ヒューマノイドロボットは、近年、大きな進歩を遂げ、その活躍場所は駅や大規模商業施設などの公共空間にまで広がりつつある。このようなロボットがさらに人間との距離を縮めるために不可欠なものは対話能力である。とりわけ、音声によるロボットと人間との対話は、最も重要なタスクと認識されており、様々な取り組みがなされている。本論文は、人間-ロボット間の音声対話能力を向上させるために、重要なモジュールとなる話者交替の推定機能に焦点を当て、ロボットや話者の状況理解に基づくアプローチを議論している。本論文では、ロボットが話者交替のタイミングを推定するために、人間側とロボット側のそれぞれの状況に基づき２種類のアプローチを考案している。第一のアプローチは、ロボットの挙動や状態に着目し、一方、第二のアプローチは、話者の発話状態に着目するものである。主たる研究成果を要約すると以下の通りとなる。

（1）第一のアプローチとして、対話状況を推定するために、対話中に話者がもつロボットへの話しかけやすさを予測するモデルを考案している。このモデルは、人間は任意のタイミングで話しかけるのではなく、相手の状況を考慮して話しかけるという社会的規範に立脚して構成される。ロボットの発話・挙動・視線などを特徴量としてロジスティック回帰を用いて予測モデルを定式化し、実験的検証によりモデルの有効性を明らかにしている。さらに、この予測モデルを現実の音声対話システムに導入して、モデルの汎用性とスケーラビリティを実験的に検証し、特に、ロボットが、話者の個人性に追随でき、緊急の問い合わせにも適応できることを確認している。考案した予測モデルが音声対話モジュールである話者交替推定機能の実現に大きく寄与することを、本研究結果は示唆している。

（2）第二のアプローチは、対話状況推定のためにロボットが話者の状態に着目するものである。ここでは、複数話者とロボットが対話する環境を想定し、ロボットが検出するすべての音（ロボットへの発話、他の話者への発話、話者自身への発話、環境雑音）に対し、ロボットが応答する義務があるか否かを推定する手法を考案、実現している。入力音区間中の話者の動き、音区間後の話者の動きや顔の向きなどの新たな特徴量を導入した結果、既存手法に比べて高い推定率を与えることを実験的に調べ、どのような特徴量が有効であるかを明らかにしている。

以上のように、本論文は人間-ロボット間の音声対話に関して数多くの有用な知見を与えており、情報通信工学、特に音声メディア処理、知的対話システムの発展に寄与するところが大きい。よって本論文は博士論文として価値あるものと認める。