

ヒューマノイドロボットへの 話しかけやすさのモデル化

杉山 貴昭[†]駒谷 和範[‡]佐藤 理史[‡]
[†]名古屋大学 工学部電気電子・情報工学科 [‡]名古屋大学大学院 工学研究科 電子情報システム専攻

1 はじめに

ヒューマノイドロボット（以下、ロボット）との音声インタラクションの実現には、周辺雑音による誤動作が問題になる。従来このような誤動作回避は入力音の判別に基づき行われる [1, 2]。これに対して本研究では、入力音ではなく、その受け手であるロボットの挙動に着目する。具体的には、任意の時点で、話しかけられやすい状況にあるかどうかを、ロボット自身がその発話や動作に基づき予測するモデルの構築を目指す（図1）。これにより、協調的なユーザが話しかけるであろうタイミングを、ロボットが知ることができる。逆に、話しかけられにくい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せる。つまり入力音の解釈時に、その時点での話しかけられやすさの状態を考慮可能なモデルを構築する。またこのモデルを用いて、ロボットにとって都合の悪い状況（内部ファンの動作中や周囲のキャリブレーション中など）に、話しかけられにくい挙動を生成・選択することも可能である。

本研究では、まずロボットの様々な挙動に対し、ユーザが実際に話しかけやすいかどうかを付与した学習データを作成する。この後、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する。本稿では、このうちの学習データの収集までを報告する。まず様々な挙動の作成について2章で述べ、データ収集について3章で述べる。このデータについて、実際に複数のユーザが話しかけられるとした区間が一致するかどうかを4章で確認する。これにより、収集したデータが学習データとして利用可能かどうかを検証する。

2 話しかけやすさに寄与するロボットの動作や発話

本研究では、ロボットに対してユーザが話しかけやすい状態かどうかを考える。この話しかけやすさは、まずユーザが話しかけようとしている内容に依存する。例えば「救急車を呼んでほしい」など非常に急ぎの内容の場合は、ロボットの状態に関わらずユーザは話しかけると予想される。ここでは、緊急でない、例えば「部屋の温度を上げてほしい」のような内容を、割り込んで話しかけやすいかどうかを考える。

この話しかけやすさにどのような動作や発話が寄与するかを考え、これらを含んだロボットの挙動を作成して、学習データ収集に用いる。

2.1 話しかけやすさに寄与する要素

ロボットに対する話しかけやすさに寄与する要素として、ロボットの動作や姿勢、発話を扱う。これらは、後に機械学習の特徴として用いるため、ロボット内部で自動的に取得可能なもののみを考える。本研究ではヒューマノイドロボットを使っているため、ユーザがロボットに社会性を感じると仮定し、人間同士の場合を参考にこ

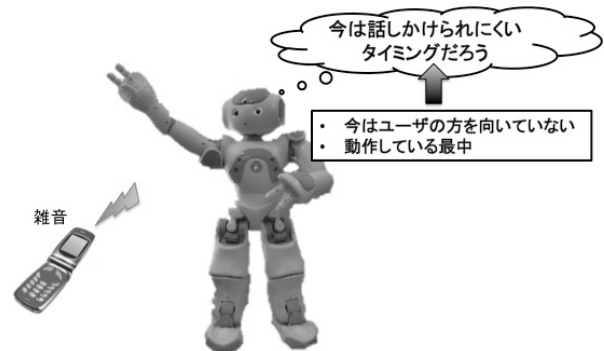


図1: ロボットが話しかけられやすさを予測する

表1: 話しかけやすさに寄与する要素

要素	話しかけやすい	話しかけにくい
動作	動作中	静止
ロボットの目線	ユーザ方向	それ以外
発話の文末の韻律	上がる	それ以外
発話の文末表現	疑問・要求	それ以外
発話間間隔	長い	短い

の要素を設計した。概要を表1に示す。例えば動作や姿勢については、ロボットが静止している時や自分の方を向いている時は、話しかけやすいと考えられる。発話については、次々と説明を続けている時には割り込みにくい、発話が終わってしばらく間が空いた場合には話しかけやすいと考えられる。

これらの要素を個々に議論することはできるが、実際のロボットの挙動では、これらが連続的かつ複合的に起こる。このため、任意の時点での話しかけられやすさの予測に用いるために、これらの要素を複合的に含んだ挙動を作成する。

2.2 ロボットの挙動の作成

表1の要素を含んだロボットの挙動を作成した。ヒューマノイドロボットとして Aldebaran Robotics 社製の NAO[§]を使用し、音声合成には VoiceText[¶]を使用した。内容はロボットの自己紹介であり、長さは約2分30秒である。ユーザの位置はロボットの正面であるとしている。

作成した挙動の一部を図2に示す。区間(A)ではロボットは動作中である。また顔は右方向を向いており正面に位置するユーザと目は合っていない。区間(B)では、ロボットは正面を向き、発話をせず静止している。区間(C)では、1秒程度の発話があり、周りを見回している。このように表1の要素を複合的に含むように挙動を作成した。

Modeling Ease of Talking to Humanoid Robot: Takaaki Sugiyama, Kazunori Komatani, and Satoshi Sato (Nagoya Univ.)

[§]<http://www.aldebaran-robotics.com/>

[¶]<http://voicetext.jp/>

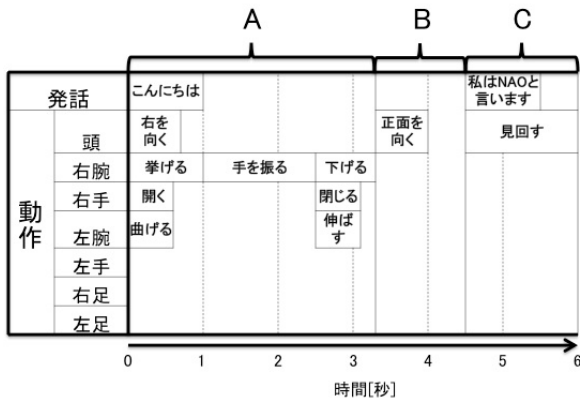


図 2: 作成した挙動の一部

3 データ収集

2章で作成した挙動に対して、複数の被験者により実際に話しかけやすさを付与させる。具体的には、被験者がロボットの挙動に対して話しかけられると感じた区間を記録する。話しかけやすさは前の発話や挙動に関係すると考え、ロボットの一連の挙動を通して見ながら付与された。

話しかけやすさの付与は、計算機のディスプレイに表示された GUI を用い、マウスをクリックすることで行った。何も操作していない時は「話しかけにくい」、マウスをクリックすると「話しかけやすい」と表示され、 t 秒後に「話しかけにくい」に戻る。クリックから t 秒以内にもう一度クリックした場合、その時点から再度 t 秒間「話しかけやすい」が表示される。予備実験の結果、 $t = 0.5$ 秒とした。この「話しかけやすい」が表示されていた区間を、ユーザが話しかけられると感じた区間とした。

本研究室の学生 3 名を被験者として実験を行った。まず被験者に実験について教示を行い、本実験で使用する GUI の使用方法の説明とそれを使用する練習を行った。次に、初めてロボットの挙動を見る場合には見入ってしまうことが多いため、被験者にロボットの挙動をまず数回見させた。その後、データ収集を、被験者 1 名について続けて 3 回行った。

被験者には、事前教示として、ロボットに「もう少し大きい声でしゃべってください」と伝えるのに、「ねえ」と呼びかけるタイミングをうかがっている、という状況を想定させた。さらに、話しかけられると感じる区間が続く間は、続けてマウスをクリックするように指示した。

4 収集したデータの分析

被験者実験により、3名の被験者が話しかけられると感じた区間を記録した 3 回ずつのデータを得た。これを用いて、被験者内および被験者間の比較を行う。

4.1 被験者内の比較

3名の被験者 (A, B, C) ごとに、まず 1名の被験者内の比較を行う。この結果を表 2 に示す。表 2 は、被験者ごとの 3 回の実験から得られた、話しかけられると感じた区間の長さ [秒] の合計と平均、標準偏差である。

各被験者の平均を比較すると、被験者 C に比べて、被験者 A と被験者 B は話しかけられると感じた区間が短い。つまり、話しかけやすさはユーザによって感じ方が異なることを確認した。また、各被験者の標準偏差を比較すると、被験者 B の値が他の被験者に比べて大きい。

表 2: 各被験者が話しかけられるとした区間長の合計 [秒]

被験者	1 回目	2 回目	3 回目	平均	標準偏差
A	15.6	16.2	17.3	16.4	0.70
B	16.7	17.8	8.6	14.4	4.10
C	27.8	27.5	27.9	27.7	0.17

表 3: 話しかけられるとした区間長とその重なり

x	y	$t(x)$	$t(y)$	$t(x \cap y)$	重複度
A	B	16.2	17.8	14.4	0.889
B	C	17.8	27.5	16.5	0.927
C	A	27.5	16.2	15.6	0.963

これは被験者 B の 3 回目の実験結果で、1 回目、2 回目に比べて話しかけられると感じた区間が短かったことによる。実験の様子から、同じ実験を 3 回続けて行ったことで、被験者が疲労していたのが原因と考えられる。

4.2 被験者間の比較

被験者間の比較を行い、被験者間で共通して話しかけられる／られないと感じる区間が存在するかどうかを調べる。なお、被験者間の比較には、各被験者の 3 回のデータのうち、2 回目の実験結果を用いた。

各被験者が話しかけられるとした区間長と、被験者間でのその重なりを表 3 に示す。 x と y は被験者を表し、例えば、A B は被験者 A と B の比較である。 $t(x)$ と $t(y)$ はそれぞれの被験者が話しかけられると感じた区間の長さ、 $t(x \cap y)$ はこの 2 名の被験者が共通して話しかけられると感じた区間 [秒] の合計である。さらにこれらの結果に対して重複度を定義し、 $t(x \cap y) / \min(t(x), t(y))$ で計算する。これは話しかけられると感じた区間が短い被験者の結果が、どの程度共通して話しかけられると感じた区間を包含しているかを表す。4.1 節で示されたように、話しかけられると感じる度合いには個人差があるが、その共通部分の割合により、話しかけられるとした区間の一致度を示す。

表 3 より、全ての被験者間において重複度が高いことを確認した。同様に、ABC3 者間の重複度を計算すると、0.877 であった。これらより、共通して話しかけられると判定した区間が存在することが確認された。さらに、話しかけられないと感じた区間についても同様の比較を行い、重複度は A B, B C, C A の順に、0.986, 0.989, 0.995 であった。3 者間では 0.986 であった。これより、ユーザによらず、共通して話しかけられる／られないと感じる区間が存在することを確認した。

5 おわりに

4.2 節の結果より、ユーザによらず、共通して話しかけられる／られないと感じる区間が存在した。このことから、収集したデータが、ロボットが話しかけられやすさを予測するモデルの学習データとして有用であることを示した。今後は、ロボットの発話やロボットの API から取得した関節角度を入力特徴とした機械学習を行い、話しかけられやすさの予測を行う。

参考文献

- [1] A.Lee et al. Noise Robust Real World Spoken Dialogue System using GMM Based Rejection of Unintended Inputs, International Conference on Spoken Language Processing, in INTERSPEECH, pp.173–176, 2004.
- [2] 野村行弘, 他. 雑音量に依存しない音声領域と雑音領域との判別を用いた音声強調. 信学技報 SP2004-6, Vol.104, No.30, pp.29–34, 2004.