

K_005

ロボットの状況判断モデル構築のための人物動作分析

An Analysis of Human Movement for Constructing a Model of Situation Assessment of Robots

松井 和教[†] 市川 貴浩[†] 菅田 雅彰[†] 村上 真[‡] 白井 克彦[†]
 Kazunori Matsui Takahiro Ichikawa Masaaki Honda Makoto Murakami Katsuhiko Shirai

1. はじめに

近年、ロボット工学の進歩や社会的なニーズの高まりから、様々なロボットの開発が進んでいる。特に技術の進歩や社会的なニーズから人間とコミュニケーションを図り、オフィスでの仕事の補助や家庭での遊び相手として生活空間を共有するロボットが研究されている。中でもヒューマノイドロボットはその身体性から、人間と円滑なコミュニケーションを実現できると期待され、注目を集めている。しかし、従来のコミュニケーションロボットはいずれも人間の働き掛けに単に応答するか、人間の状況を考慮せずに一方的に働き掛けているものであるため、人間同士が行う円滑なコミュニケーションを実現しているとは言い難い。

そこで、我々は円滑なコミュニケーションを実現するロボットとして、人間の動作や表情・音声などから相手の状況を判断し、相手の作業大きさを妨げることなく自ら対話を開始、終了するロボットの構築を目標とする。

ロボットが相手の作業を妨げずに対話を開始するには、話しかけるか否かを判断する状況判断モデルが必要となる。本研究では、この状況判断モデルを構築するために人物動作分析を行う。

2. 分析手法

相手の作業を妨げずに話しかけるには、相手が作業中か否かを判断すれば良い。そこで、被験者に何らかの作業を課し、別の被験者に話しかけるか否かを判断してもらう実験を行うことで、状況判断と人物動作の関連性を分析する。

実験は一般的に人間が作業を行う場所、将来ロボットが普及すると予想される環境を考慮しオフィス内を想定する。今回は初段階として座って作業する人物の動作分析を行う。

3. 予備実験

人物動作を分析するには、人が話しかける際、話しかける相手のどの身体部分を見て判断しているかを調べる必要がある。そこで、どの身体部位を見ているかを特定するために予備実験を行う。予備実験を行うために人間にオフィス内で想定される作業を課し、その様子をビデオカメラで撮影したものを実験データとして準備した。

3.1 実験方法

実験データを用いて予備実験を行った。データは15秒の長さに編集したものを35個用意して被験者にランダムで提示した。ただし、15秒の間に作業者に極端な動きが起こらないよう編集した。被験者には実験データの映

像を見てもらい映像の人物に「話しかける」「話しかけない」の2択で判断してもらった。その際に人物の身体どの部分を見て状況判断したかをアンケートで調べた。

ただし、相手の作業を妨げずに話しかけてもらうために、そこで被験者に映像内の人物に聞きたい些細な用事があるという状況設定を設けた。

また、話しかける人間と、話しかけられる人間の関係は話しかけに影響する可能性がある。そこで、一般的にロボットと人間の関係ではロボットの方が下であることから、話しかけられる映像内の人物をオフィスの上司、話しかける被験者をその部下としてお互いの関係を設定した。この実験を5人の被験者に行った。

3.2 予備実験結果&考察

被験者に行ったアンケートには、話しかけると判断した理由として「椅子にもたれている」「余所見をしている」「手持ち無沙汰だった」といった意見が多く存在した。逆に話しかけないと判断した理由として「姿勢が前屈み」「顔が机を向いている」「物を持っている」といった意見が多くみられた。このことから、作業者の「姿勢」「顔の動き」「手の状態」といった身体動作が、話しかけるか否かの状況判断に重要であると考えられる。

4. 三次元位置情報を用いた人物動作分析

ロボットの状況判断モデル構築のために、人物動作をステレオカメラを用いて三次元位置計測する実験を行った。

4.1 三次元位置計測

三次元位置情報は、複数のカメラによるステレオ法の原理により求めることが可能である。まず、左右画像間のキャリブレーションを行う。キャリブレーションは左右画像間のブロックマッチングにより可能である。左右画像間のブロックマッチングは、次の式で求められる。

$$\min_{d=d_{\min}}^{d=d_{\max}} \sum_{i=-\frac{M}{2}}^{\frac{M}{2}} \sum_{j=-\frac{M}{2}}^{\frac{M}{2}} |I_{\text{right}}[x+i][y+j] - I_{\text{left}}[x+i+d][y+j]| \quad (1)$$

左右画像間の輝度を用いて、 $M \times M$ ブロックのブロックマッチングを行い、ブロック内の輝度 I_{right} と I_{left} の差が最小となる対応点を検出する。

奥行き z は、カメラ間の距離（基線長） D と2画像間の視差 $|x_2 - x_1|$ および焦点距離 f を用いて、次の式で求められる（図1参照）。

$$z = D * f / |x_2 - x_1| \quad (2)$$

[†]早稲田大学

[‡]東洋大学工学部

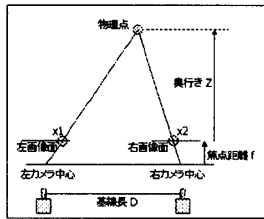


図 1: ステレオ法の原理

入力デバイスとして、PointGrey社の2眼ステレオカメラ Bumblebeeを用いる。解像度はXGA (1024 × 768 ピクセル) で、奥行きは15FPSで測定可能である。

5. 実験方法

オフィス内で想定される作業を課した人物の体にマーカーをつけ、その様子をステレオカメラで撮影したものを実験データとして用意した。マーカーは頭部に3個、腰に2個、首・肩・肘・手・背中にそれぞれ1個ずつ配置した(図2参照)。

予備実験と同様に撮影映像を15秒の長さに編集したものを被験者にランダムで提示し、話しかけるか否か判断してもらった。映像の編集方法や話しかける状況は予備実験と同様のものを設定した。この実験を7人の被験者に行った。



図 2: ステレオカメラによる撮影

5.1 実験結果&考察

被験者によって状況判断された映像の結果と、ステレオカメラで得た人物の身体部位の三次元位置情報を用いて、「話しかける確率」と「姿勢」「顔の動き」「手の状態」の関連性を調べた。姿勢については腰のマーカー2つと首のマーカーの三次元位置情報から背中の傾きの角度を求め、前屈みか否かを調べた。また、顔の向きは頭部につけた2つのマーカーを用いて y 方向 z 方向それぞれの顔の角度を求め、2つの角度の分散値の平均をとることで、顔が動いているか否かを調べた。「手の状態」については、物を持っているといった状態が三次元情報に現れないので、手を机に置いていない状態と置いた状態を分けて分析を行った。

結果として、「話しかける確率」が高い映像ほど「背中の傾き」「顔の動き」の値が大きく、逆に「話しかける

確率」が低い映像ほど「背中の傾き」「顔の動き」の値が小さいという傾向が出た。また、手を机に置いていない状態の方が、置いてある状態に比べて「話しかける確率」が高かった。このことから「姿勢」「顔の動き」「手の状態」が状況判断に有効であることが確認できた。状況判断モデル構築を想定して、各映像の「話しかける確率」を従属変数とし、「背中の傾き」「顔の動き」独立変数として重回帰式を求め、予測値との比較を行った。結果として、両方の状態で誤差が生じた。手を机に置いた状態では「話しかける確率」と「予測値」の誤差が特に大きいという結果が出た。このことから、手を机に置いた状態での状況判断には別の要素が必要であると考えられる。

表 1: 手を机に置いていない状態

映像	話しかける確率	予測値	姿勢の角度	顔の動き
A	100.0 %	85.7 %	110.703	70.990
B	100.0 %	79.7 %	110.495	60.623
C	100.0 %	114.0 %	91.647	154.963
D	100.0 %	84.0 %	85.489	112.107
E	57.1 %	55.1 %	82.235	66.011
F	14.3 %	34.3 %	93.112	9.640
G	0.0 %	24.5 %	71.534	29.843
H	0.0 %	16.9 %	62.453	32.041

表 2: 手を机に置いた状態

映像	話しかける確率	予測値	姿勢の角度	顔の動き
I	57.1 %	73.3 %	119.168	109.689
J	57.1 %	26.2 %	93.209	22.696
K	28.6 %	20.5 %	76.872	72.914
L	14.3 %	49.3 %	116.893	14.995
M	0.0 %	23.0 %	93.358	8.097
N	0.0 %	5.1 %	76.611	6.628
O	0.0 %	8.2 %	68.610	57.246
P	0.0 %	-1.4 %	66.802	23.206

6. まとめと今後の予定

本研究では、相手の作業を妨げずに、話しかけるロボットの状況判断モデル構築を目的とし、話しかけの判断と人物動作の関連性の分析を行った。結果として、「姿勢」「顔の動き」「手の状態」が状況判断に有効だと確認された。しかし、手を机に置いた状態での状況判断には別の要素が必要であると考えられる。今後は、別の要素として手の詳細な状態・顔の表情などを用いることで、より高い精度で話しかける確率を予測可能にしていく予定である。

参考文献

- 小阪朋也, 米村俊一, "暇度評価指標による人間の稼働状況推定法の検討", 2001年電子情報通信学会総合大会, A-14-4, pp. 280, 2001.
- 松井俊浩, 麻生英樹, J. Fry, 浅野太, 本村陽一, 原功, 栗田多喜男, 速水悟, 山崎信行, "オフィス移動ロボット Jijo-2 の音声対話システム", 日本ロボット学会誌, Vol. 18, No. 2, pp300-307, 2000.