

顔表情認識による対話ロボットの相槌制御

三輪 大知^{†1} 上田 博唯^{†1}

概要：対話ロボットとユーザである人間とのコミュニケーションを円滑化するためには、対話ロボットが人間の話に合わせて頷くことが有効であることが知られている。本研究では対話の相手である人間の顔の表情を認識して、その表情のポジティブさに応じて相槌の振る舞いを制御することで、人間がより話しやすくなるという仮説について検証実験を実施したので報告する。

Nod Control of Conversational Robot by Face Expression Recognition

DAICHI MIWA^{†1} HIROTADA UEDA^{†1}

Abstract: It is known that nod tailored to the human story is effective for smooth communication between the interactive robots and humans. We propose a method to control the behavior of nodding in response to the degree of positive facial expression by recognizing the expression of the human face. In this paper, a verification experiment performed for the hypothesis that human beings is likely to talk more by this method is reported.

1. はじめに

本研究は報告者の一人の母親が介護ヘルパーの経験者であったことに端を発している。その方によると被介護者は訪問介護として限られた時間に来てくれる介護者には自分の話を聞いてもらうことを切望しているが、介護ヘルパーに許された時間は限られており、その間に洗濯・炊事・掃除などの家事もこなさなければならないので、離れたところから家事をしながら(顔を見合わせることもなく)、ときどき声をかけてあげることくらいが関の山ということであった。そんなことから、介護ヘルパーが家事をしている間、親身になって被介護者の話を聞いてくれるロボットがいれば、そのロボットと被介護者の会話を耳に挟みながら家事に専念でき、自分を必要としているような話題になったときには家事を中断して会話を合流することも可能となり、とても有用であろうという要求仕様がまとまってきた。

具体的な仕様については、介護ヘルパーが被介護者の話を聞くのにあたり気をつけている点を参考にして詳細を詰めていった。介護ヘルパー経験者の方から得たノウハウは次のようなものであった。対話の主体はあくまでも被介護者の側にある。従って、第一に重要なことは相手の話をできるだけ理解するように努め、そのことを聞き手が態度として表すことである。また、二番目に大事なことは、聞き手が相手の話の腰を折らないようにする

ことである。

一番目の相手の話の理解ということに関しては、現状の音声対話、特に自然言語処理の技術レベルはまだまだ不十分であり、中途半端にこれをロボットに実装することは、利用者に過大な期待を持たせてしまうことによる不満の要因となることが容易に予測される。そこで本研究では発話者の文の切れ目ごとに相槌を打つことによって、聞いていますよということを示すことにした。ひとの話の聞くときにロボットが頷き・手振り・身振りなどの身体動作を返すことによって、ひとが話しやすくなるということについては、既に岡山工科大学の渡辺によって明らかにされているとおりである[1]。本研究では、それに加えてオムロン社の顔認識技術 OKAOVision [2]を組み合わせることにより、話の内容がポジティブな気持ちになるものであるか、ネガティブな気持ちになるものであるかを判断して、それにあわせて相槌の言葉と頷きや手の仕草を変化させることを提案する。話の内容まではよくわからないが、すくなくともポジティブな気持ちかネガティブな気持ちかくらいは伝わっているということのフィードバックである。将来的には、オープンソースの大語彙連続音声認識エンジン Julius [3]などを上手に用いることで、もう少し意味的なところまで踏み込んで話の内容を感じ取れるようにしたいと考えている。

二番目の相手の話の腰を折らないということに関しては、上記の大語彙連続音声認識エンジン Julius を用いることで、確実に発話センテンスの切れ目を認識するようにして、話し手の話の切れ目までは相槌の発話や頷きを

^{†1} 京都産業大学
Kyoto Sangyo Univ.

入れないような制御を行うことにした。また先に筆者らの研究[4]で効果を認められた、語尾の単語によって、話し手が聞き手に同意を求める気持ちを判定して、相槌の発話や顔きの大きさや回数を変化させるという手法を組み合わせることも考えている。

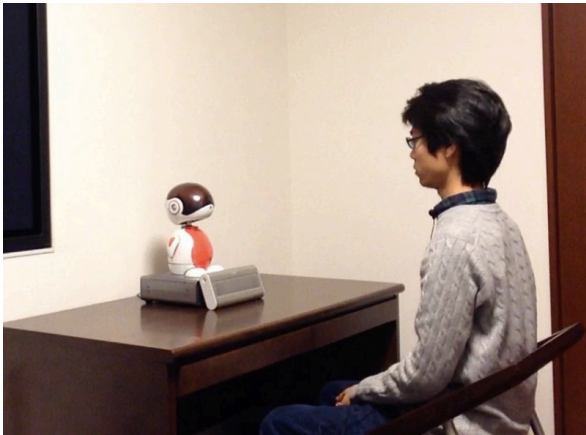


図 1 試作システムの外観
 Figure 1 Appearance of the System.

2. 実装

試作したシステムの利用状況を図 1 に示す。本システムでは、前述した顔認識モジュール OKAO Vision と大語彙連続音声認識エンジン Julius を利用している。Julius が音声入力を開始したというメッセージと同期して OKAOVision による顔表情認識を開始し、音声入力の終端を受信した時点(すなわち発話センテンスの区切り目)で対話ロボットの相槌を開始するように制御している。ただし、この時 Julius からパラメータとして得られる発話継続時間をチェックして、800ms 以下である場合には、ドアの開閉音などの外部ノイズや発話センテンス以外の咳払いや「えーと」などの相槌を必要としない短い発話であると判断して除外している。



図 2 ユーザ視点から見たロボット
 Figure 2 Appearance of the Robot by the User's Viewpoint.



図 3 顔表情認識中の顔画像
 Figure 3 Image under Face Expression Recognition.

ユーザから見たロボットの様子を図 2 に示す。表情認識には、このロボットが内蔵しているカメラを使っており、標準的な着座姿勢におけるカメラ視野に対する顔の大きさは図 3 に示すとおりである。OKAOVision は複数の感情に対応する表情の値と、感情がネガティブであるかポジティブであるかを表す-100 から+100 までの値(以下ポジティブ値と呼ぶ)を常時出力しているが、今回はポジティブ値だけを利用している。

相槌を実行する場合には、前述の発話区間中に顔表情認識モジュールから送られてきていた表情データを解析して、次のように三種類の相槌動作を使い分けるようにしている。

- 1) ポジティブ値が 0 より大きい場合 (ポジティブ)
 - セリフ: うん、うん
 - 手の動き: 約 60 度で上下、約 30 度で上下を早めに頭の動き(チルト): 10 度で上下、5 度で上下を早めに
- 2) ポジティブ値が 0 から -80 の場合 (真顔)
 - セリフ: うん、うん
 - 手の動き: なし
 - 頭の動き(チルト): 10 度で上下、5 度で上下を早めに
- 3) ポジティブ値が -80 より小さい場合 (ネガティブ)
 - セリフ: うん
 - 手の動き: なし
 - 頭の動き(チルト): 10 度で上下をやや遅めに

なお、顔表情の判定に際し、今回は話し始めてから 800ms 間のみという仕様で実装したが、今後は予備実験のログや記録映像の評価などに基づいて、以下に例示するような細かい使い分けを考えている。

- 1) 話し始めの笑み
 - 「その話す内容自体が面白かった」のであるから、聞いている間も継続的に楽しそうな動きをする。例えば身体を少し揺すったり、小さな顔きを繰り返す

- たりしながら聞く。
- 2) 発話の途中にくる笑み
 「何かがあったから笑ってしまった」のであるから、それに対応して反応する。例えば「ははは」と発話して、両手を上げる。
 - 3) 発話の終わりにくる笑み
 「話していておかしかった」のであるから、話し始めの笑みに対するより大きく反応する。

3. 評価実験

試作したシステムを使って以下のような評価実験を計画している。実験手順は以下のとおりである。実験は比較のために、今回提案する相槌制御のある対話ロボット、相槌制御のない対話ロボット、そして対話ロボットではなく生身の人間の三つの条件で対話を行うことによって評価を実施するが、ここでは対話ロボットの場合についてのみ示す。

- ① 被験者に以下の指示を与えて、約 6 分間の動画を見せよう
 - 後でロボットにビデオの筋書きを丁寧に説明してもらいます
 - 後の説明のために、必要ならメモを取っても構いません
- ① 被験者に以下の指示を与え、ロボットとの対話を行ってもらおう
 - 動画の内容を子供に説明するように、ゆっくりとロボットに向かって説明してください
 - 次に、動画の面白かったところを、ロボットに向かってできるだけたくさん話してください
 - 最後に、動画の中の場面のどこが一番気に入ったかについて話してください
 - 話しかけるときにはロボットの顔を見て話すようにしてください
- ② ロボットとの対話の終了後、5 段階の印象評価アンケートに答えてもらう



図 4 ログ映像
 Figure 4 Logging Video.

- ③ 記録しておいた実験中のログ映像を見つつ、ここでどんなことを感じていたかについて思い出したことをすべて用紙に記入してもらう

被験者に見せる動画は、いわゆるドッキリと呼ばれるビデオで、笑いや同情、そして驚きなどを誘う内容となっている。印象評価アンケートにおいては、ロボットの喋り方や仕草への違和感、相槌動作の単調さ、相槌動作タイミングの適不適、ロボットによる話し手への理解度、話の腰を折らなかつたか、楽しく会話できたか、などの項目について 5 段階で記入させるようにした。実験中のログ映像の一例を図 4 に示す。被験者には、この被験者の表情とロボットの仕草とを同時に確認することができるアングルで撮影されたログ映像を自由に再生(巻き戻し早送りも可とした)しながら、実験中に感じたことを、時間軸メモリのついた用紙に書き留めてもらうようにした。したがって実験者はこの時刻情報に基づいて、ログ映像の該当箇所を参照しながら詳しく分析することが可能となる。

4. 予備実験の結果と考察

本稿執筆時点までに完了した相槌制御のある対話ロボットの評価実験 5 人分(文系女性 1 名、文系男性 2 名、理系男性 2 名)について述べる。それぞれの被験者についての発話継続時間とロボットの相槌の種類ごとにカウントしたものを表 1 に示す。ロボットに話しかけた発話の継続時間は最も長かった被験者の場合でおよそ 3 分半、短い被験者は 1 分程度となったが、どの被験者に対しても、ロボットは 1 分あたり 10 回程度の相槌を返している。また、ロボットに話しかけている時にポジティブな表情が出る(笑顔になる)かどうかということについては、被験者ごとのバラツキがかなり大きいことがわかった。被験者の数を増やしつつ、このようなアプリケーションにおける表情認識の問題点(そもそもどんな人でも表情豊かに話すものなのか、ロボットのカメラでうまく表情を捉えられるのか etc.)の抽出と解決が必要である。それと並行して表情認識の結果による相槌制御の有効性をより深く考察していきたいところである。

表 1 相槌の種類ごとの回数

Table 1 Set of Style in MS-Word template file.

番号	継続時間	相槌の種類ごとの回数		
		ポジティブ	真顔	ネガティブ
1	3分 28 秒	5	17	1
2	2分 54 秒	0	22	0
3	1分 17 秒	2	7	0
4	1分 11 秒	0	12	0
5	1分 42 秒	1	22	0

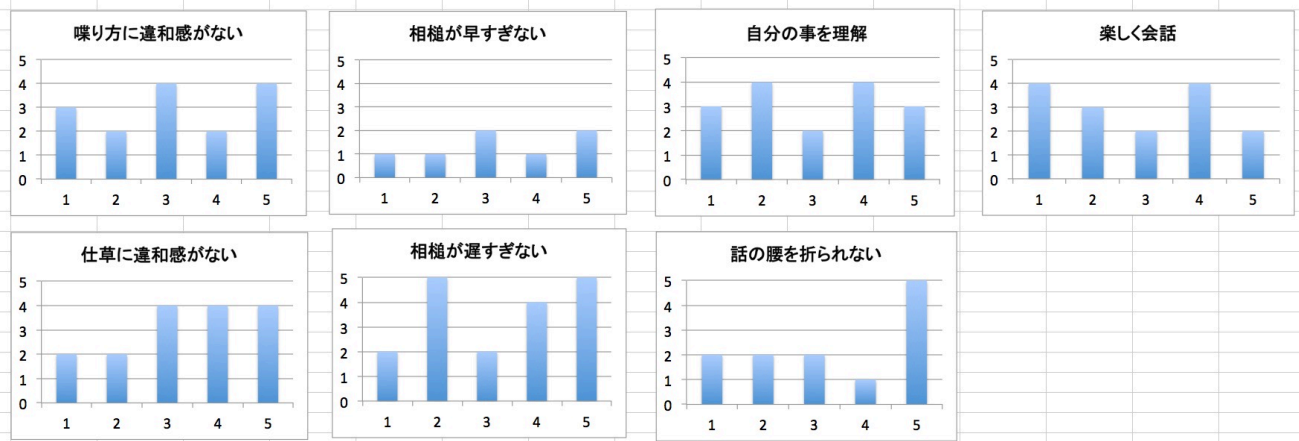


図 5 印象評価の結果

Figure 5 result of the impression evaluation.

表 2 被験者のコメントとその分析結果

被験者	被験者のコメント	分析結果
1 文系 女性	1 相槌が少し早いと思った	前の発話の終端の判定の遅れとロボット動作開始のディレイとにより、次の発話の終端の前で相槌を打った
	2 言葉に詰まった時に頷いてくれて先が話しやすくなった	言葉につまった時に発話の終端と正しく判定していた「言葉に詰まった時の相槌」を好意的に受け取っている
	3 手をパタパタしているのを見て、もっと話してあげたいと思った	ポジティブ表情が正しく判定されていたロボットのポジティブな対応が期待通りの効果を得た
	4 自分が少しキョロキョロしていた時も反応してくれて嬉しかったし、可愛いと思った	Juliusのメッセージがバッファにたまっていて、それが遅れて処理されて相槌を実行していたプログラムのバグでさえ好意的に受け取られていた
2 文系 男性	1 相槌が早い	声が小さくなったところで発話の終端と誤って判定していた
	2 相槌を行うタイミングじゃない所での相槌	前の発話の終端の判定が遅れ、次の発話の終端の前で相槌を打った
	3 少し遅れたタイミングでの相槌	発話の終端の判定の遅れとロボット動作開始のディレイ
	4 早い相槌	被験者1の1と同じ
	5 話終わりに相槌がない	対話の最後の発話が短すぎてノイズとして破棄された

それぞれの被験者の印象評価値を図 5 (評価値の数値を好意的な方が上に来るようにして表示) に示す。横軸が被験者番号であり縦軸が印象評価値である。この印象評価の結果についても、被験者ごとのバラツキはかなり大きい。概ね好意的な印象を持たれていると言えるのではないだろうか。ただ、全体的に相槌のタイミングが早すぎるという印象になっている。ログ映像を見つ、どこでどんなことを感じていたかを記入してもらった結果の二人分を表 2 に示すが、この解析から、その原因を伺い知ることができる。表 2 において、被験者 1 の 1 のコメント「相槌が少し早いと思った」についてログ映像を分析した結果、ひとつ前の発話の終端判定の遅れに加えてロボット動作開始のディレイがあるために、また該当する発話の長さが短かったために、結果として次の発話が終わり切る直前で相槌を打ったように感じられて

いるということが判明した。被験者 2 の 4 のコメントも同様である。この逆に、システムの動作としては同じ現象なのに、それを被験者 2 の 2 や 3 のコメントのように、「相槌を行うタイミングじゃない所での相槌」や「少し遅れたタイミングでの相槌」と受け止められるケースもあった。これらの被験者の受け止め方は該当する発話の長さがどれくらいであるかに依存して変化する。また被験者 2 の 1 のように声が小さくなったところで発話の終端と誤って判定してしまったことが「相槌が早い」と受け止められるケースもあった。これらのことから、相槌のタイミングの遅れに関しては、今後なんらかの改善が必要である。

被験者 2 の 5 のコメント「話終わりに相槌がない」は、対話の最後の発話が短すぎてノイズとして破棄されたもので、なかなか厄介な問題である。発話継続時間が短い

場合(800ms 以下)には、ドアの開閉音などの外部ノイズや発話センテンス以外の咳払いや「えーと」などの相槌を必要としない短い発話であると決めてしまった設計は実用的ではなく、ノイズをノイズとして認識するように改善する必要がある。

被験者1の3のコメント「手をパタパタしているのを見て、もっと話してあげたいと思った」を分析すると、それ時 OKAO Vision がまさしくポジティブ表情を検出しており、それによってロボットがポジティブな対応をしていたケースで、ここでは期待通りの効果が得られたと言える。被験者1については4のコメント「自分が少しキョロキョロしていた時も反応してくれて嬉しかったし、可愛いと思った」を見ても、ロボットに対して愛着とでもいべき関係を築きつつあると思われる。この時、ロボットはプログラムのバグによる誤動作をしていたが、それすら好意的に受け止められているということに注意すべきである。

もう一つ注目すべきは被験者1の2のコメント「言葉に詰まった時に頷いてくれて先が話しやすくなった」である。これは、言葉に詰まって空白時間ができたところを Julius が発話の終端と認識したために、ロボットが頷いたものであった。それをこの被験者は自分への思い遣りと受け止めたようである。しかし、別の被験者は同様のことが起きたケースについて、自由記述欄に「言葉に詰まった時に相槌を打たれて不満であった」と書いていた。被験者1は3のコメント「手をパタパタしているのを見て、もっと話してあげたいと思った」にも見られるようにロボットに対して大変好意的であり、それゆえにこの一つの頷きが自分への思い遣りのように思えるのに対し、ロボットに対して特に好意的ではなく別の被験者にとっては、そのロボットの一つの頷きが自分への軽蔑や嫌味のようなものに見えてしまうのであろう。

被験者数がまだ少ないので、断定的なことは言えないが、文系の学生はロボットに対して懐疑的であり、文系の学生はロボットに対して好意的であるという傾向があるように見える。そして前者はロボットに話しかける発話の継続時間が短く、印象評価の評価値も低いという傾向にあった。いずれにしても、このような聞き手役のロボットを設計する際に、このロボットと人間の関係をどのようにモデル化することができるのか、またそのモデルに基づいてどのようにロボットの対応を個人の性格や感性に適応させていくことができるのか、そういうことが今後の大きな課題であると思われる。

5. まとめ

対話ロボットが対話の相手である人間の顔の表情を認識して、その表情のポジティブさに応じて相槌の振る舞いを制御することで、人間がより話しやすくなるという仮説に基づく対話ロボットの構想について提案した。本稿では、介護ヘルパーが家事をしている間、親身になって被介護者の話を聞いてくれるロボットを具体的なターゲットとしてシステムを試作して計画している検証実験について報告した。また、予備的な実験で得られた幾つかの興味深い結果についても述べ、それらについて考察した。今後、ロボットと人間の関係をどのようにモデル化していくことができるのかが大きな課題である。

参考文献

- 1) OKAO Vision | 技術紹介 | オムロン人画像センシングサイト : +SENSING: <http://plus-sensing.omron.co.jp/technology/>
- 2) 渡辺富夫: 身体的コミュニケーションにおける引き込みと身体, ベビーサイエンス, Vol.2, pp.4-12 (2003)
- 3) 大語彙連続音声認識エンジン Julius: <http://julius.osdn.jp/>
- 4) 松原 大, 上田 博唯: 愚痴を聞くロボットの提案, 電子情報通信学会技術研究報告 (MVE) 111(38), 45-50 (2011)