

姿勢ミラーリング効果を伴う 3D 視マルチモーダル傾聴対話システムの開発

Development of a 3D-Visible Multimodal Attentive Listening System with Posture Mirroring Effect

情報工学専攻 水谷 林太郎

Rintaro MIZUTANI

概要

傾聴はラポール形成に効果的なコミュニケーション技法である。本研究では、裸眼立体視ディスプレイに 3D アバターを投影するマルチモーダル傾聴対話システムを開発した。本システムは相槌、共感、オウム返し、モダリティに応じた応答、相互の質問などの傾聴対話をベースにした音声コミュニケーションに加えて、表情、視線、瞬き、頷き、姿勢ミラーリングなどの非言語情報からなる多角的な伝達手段を有する。大学の学部生と大学院生からなる 60 人の参加者が、アバターと一対一の密室において、趣味、好き/嫌いな食べ物、将来の目標、悩み、休日の過ごし方、子供の頃の思い出、部活やサークル、アルバイト、失敗談その他について自由に 3 分間程度の会話をおこなったのち、22 項目の感性評価に回答した。アバターの話の聞き方や共感の示し方などの傾聴姿勢が参加者から高い評価を得た。加えて、アバターが親身だと感じ、またアバターと話したい、と感じる参加者が多数派であることが確かめられた。

キーワード: 自然言語処理, 対話システム, マルチモーダル, 傾聴, ミラーリング。

1 はじめに

音声やテキストによって対話をおこなう対話システムは、広く実用化されており、注目を集めている。音声対話のタスクの一つに傾聴対話がある。日本語の傾聴対話では、アンドロイドの ERICA を用いたマルチモーダルな傾聴対話システムを実現した井上らの研究 [1] や、犬型ロボットを用いた下岡らの研究 [2] などの例がある。

非言語情報の一つである姿勢の傾聴技法にミラーリングがある。ミラーリングとはクライアントの身振りや仕草、姿勢、表情などの視覚的なものや、声のトーン、リズム、話のスピードなどの音声的なものなどの、ノンバーバル行動やよく用いる言葉のパターンを注意深く観察し、それに合わせていくことである [3]。ミラーリングはラポール形成につながるポジティブな効果、印象

を与えることが、人対人のデータから示されている [3]。吉田^{*1}らは 3D アバターの姿勢ミラーリングを用いた傾聴対話システムを開発し、印象評価に効果があることを報告している [4]。しかし、本来立体的であるべき姿勢ミラーリングの情報が、通常のディスプレイで伝えられているとは考えにくい。そこで、本研究では裸眼立体視ディスプレイを用いて 3D アバターを立体的に表現し、物理的なロボットでは難しい繊細な動きを実現する。

本研究では姿勢ミラーリングや裸眼立体視ディスプレイを用いることによる、ユーザーとアバターとのラポール形成に対する効果の検証、並びに人間と人工知能の間に信頼関係を構築し、人の話し相手になれるような対話システムの開発を目指す。

2 システム設計

2.1 システム全体の構成

システムの全体構成を図 2 に示す。相槌生成システムは香曾我部らの研究 [5] を参考に開発した。音声認識には GCP Speech-to-Text、音声合成には GCP Text-to-Speech を用いる。

2.2 対話システムのアバター

アバターは若者に親しみを持ってもらうために、ポピュラーなキャラクターである図 1 のユニティちゃん (@UTJ/UCL) を使用する。ユニティちゃんには表情や瞬きといった機能がついている。背景は当研究室の有志が作成した、当研究室を模したマップを使用している。

2.3 傾聴対話システムの構成

傾聴対話システムのフローチャートを図 3 に示す。

2.3.1 応答文生成

応答文は相槌、オウム返し、共感、用言に関する深掘り質問、モダリティに応じた応答の五つで構成されている。共感応答には日本語評価極性辞書や日本語感情表現辞書を用いる。用言に関する深掘り質問では京都大学格フ

^{*1} 吉田裕. 姿勢一致エージェントの開発と対話におけるラポール形成過程の検証. 大阪工業大学情報科学部情報メディア学科卒業研究, 2013.

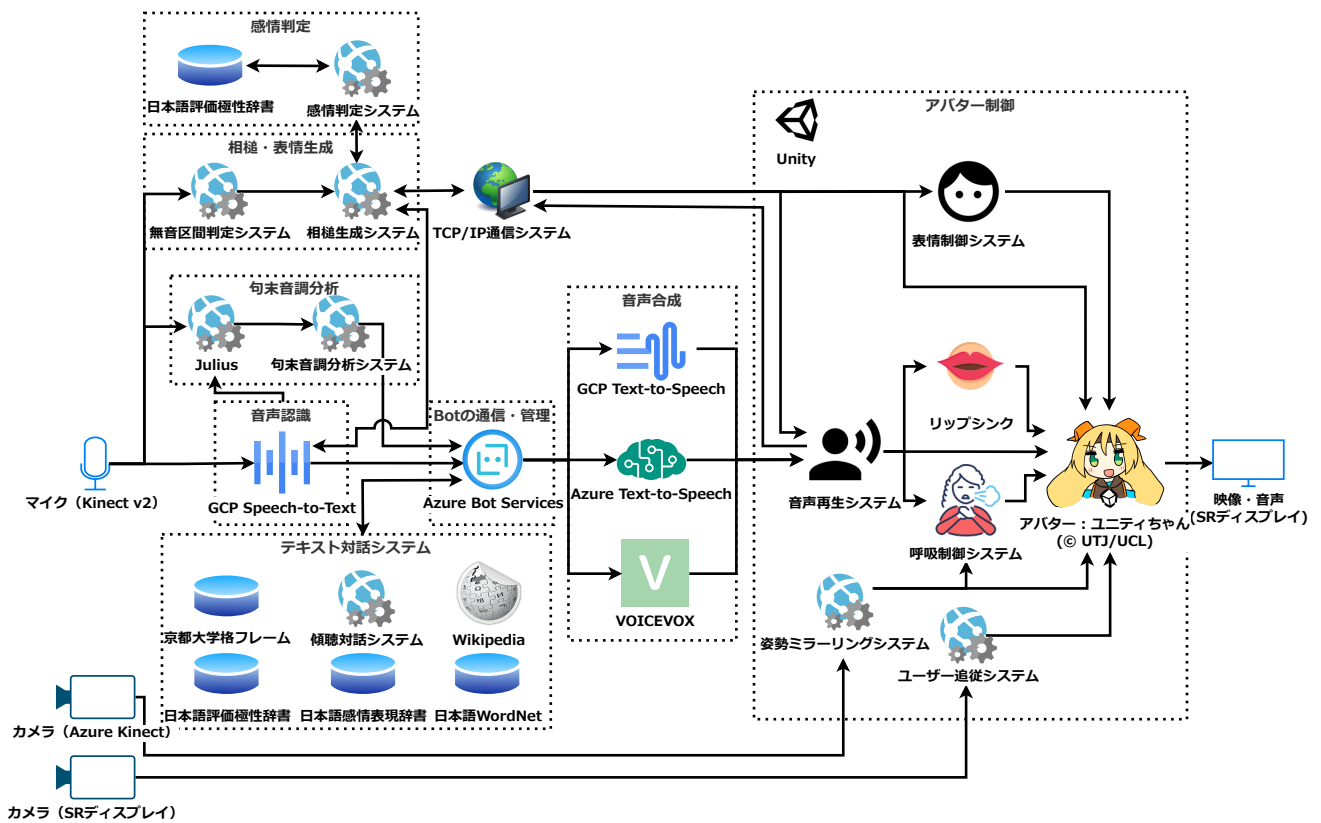


図 2 マルチモーダル傾聴対話システムの全体構成

レーム [6] による質問生成手法を, 古川らの研究 [7] を参考に開発した. モダリティに応じた応答では, KNP の解析結果に含まれるモダリティを利用する. 各モダリティに対し, あらかじめ作成した応答を返す.

2.3.2 質問応答

オープンクエスチョンでは WordNet[8] や Wikipedia の情報を用いて応答する. クローズドクエスチョンでは日本語評価極性辞書の情報を用いて, ポジティブな内容であれば肯定, ネガティブな内容であれば否定する.

2.4 句末音調分析

句末音調分析システムは, Julius[9] を用いて句末の音調を分析することにより, テキストからは判別できない

疑問文を判別するシステムである. 本システムは波多野らの研究 [10] を参考に開発した.

2.5 姿勢ミラーリング

吉田らの研究 [4] を参考に, 姿勢ミラーリングシステムを開発した. アバターの頭に関しては, ユーザー追従システムにて操作しているので, 姿勢ミラーリングの対象にはならない. 姿勢ミラーリングには鏡像的姿勢一致と同一姿勢の 2 種類が存在する. 鏡像的姿勢一致の方がラポール形成に有効という実験結果があるので*1[11], 本システムは鏡像的姿勢一致をおこなう.

2.6 対話のデモ動画

筆者との対話デモ動画は YouTube にある*2. 360 度ステレオ VR 版も存在する*3.

3 対話システムの評価実験

3.1 実験設定

対話システムと実験参加者が密室にて一対一の状況において対話をおこなったのち, 参加者は評価アンケートに記入した. 対話時間は 3 分である. 参加者からシステムへの質問は可能としたが, 質問攻めはしないように注意した. また, 参加者が話題に困らないように最初の話



図 1 アバター: ユニティちゃん

*2 <https://youtu.be/SvsgBJJ7kwg>

*3 <https://youtu.be/W23AkfgFe-s>

表 1 アンケート結果

番号	評価項目	最大値	最小値	平均値	中央値	標準偏差
Q1	アバターが話した言葉は自然だった	7	1	4.4	5	1.3
Q2	アバターはタイミングよく反応していた	7	2	4.9	5	1.3
Q3	アバターはこまめに反応していた	7	2	5.5	6	1.2
Q4	アバターの反応は人間らしかった	7	1	4.5	5	1.4
Q5	アバターの反応はあなたの話を適切に促していた	7	1	4.1	4	1.5
Q6	アバターの反応の頻度は適切だった	7	2	5.1	5	1.3
Q7	アバターの動きは自然だった	7	2	5.4	6	1.4
Q8	このアバターとまた話したい	7	1	5.1	5	1.5
Q9	このアバターは話しやすい	7	1	4.5	5	1.4
Q10	アバターは親身だと感じた	7	2	5.4	5.5	1.2
Q11	アバターは真面目に話を聞いていた	7	2	6.0	6	1.0
Q12	アバターは集中して話を聞いていた	7	2	5.9	6	1.1
Q13	アバターは積極的に話を聞いていた	7	1	5.4	5.5	1.3
Q14	アバターは話を理解していた	6	1	3.8	4	1.4
Q15	アバターは話に対する関心を示していた	7	1	4.9	5	1.3
Q16	アバターはあなたに対して共感を示していた	7	2	5.3	5	1.1
Q17	アバターは裏で人間が操作していたと思う	6	1	2.4	2	1.4
Q18	アバターは会話の間の取り方がうまい	7	1	3.8	4	1.4
Q19	実験で使用したディスプレイは、アバターの印象に良い影響を与えていた	7	1	6.0	6	1.2
Q20	会話について満足した	7	1	4.5	5	1.4
Q21	会話でのやりとりはスムーズだった	6	1	4.1	4.5	1.3

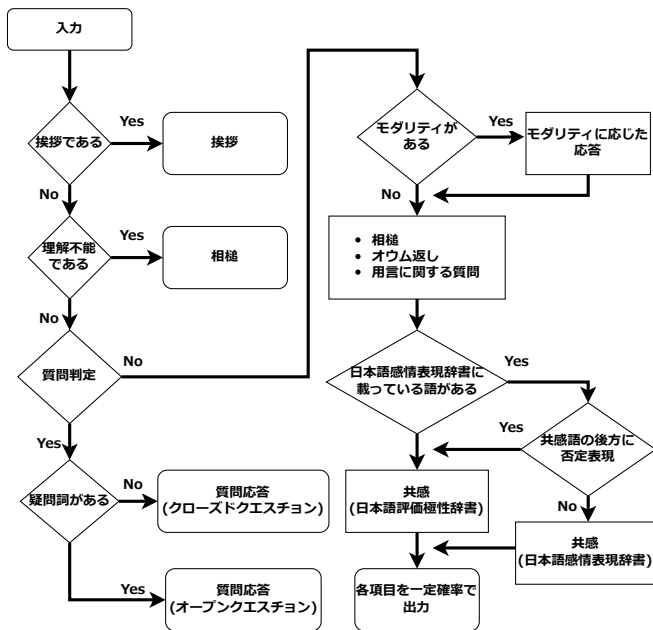


図 3 傾聴対話システムのフローチャート

題は最近あった楽しい/悲しい出来事ということを明かし、かつ話題リストを実験中も見られる位置に掲示した。

3.2 実験結果

実験参加者は大学生と大学院生合わせて 60 名だった。アンケート結果のうち、リッカート尺度を用いた項目をまとめたのが表 1 である。ほとんどの項目にて肯定的な評価を得られた。まず評価が良かった項目として、Q7, Q11, Q12, Q13, Q19 があげられる。Q11 と Q12 と Q13 の評価の平均は 6 に近く、十分に良い評価である

といえる。他には、Q10, Q16 も良い評価を得られた。ミラーリングはラポール形成に効果的であり、姿勢ミラーリングがこれらに影響を与えた可能性もある。また、傾聴対話システムの共感で、日本語感情表現辞書を用いた手法が Q16 において効果を発揮した可能性もある。Q7 の評価の平均は 5 を超えており、自然な動きができていたといえる。Q19 の評価の平均は 6 と十分に良く、マルチモーダル対話システムにおける SR ディスプレイの有用性が示された。次に評価があまり良くなかった項目として、Q1, Q4, Q5, Q9, Q14, Q18, Q21 があげられる。Q1 は対話ログを見ると、不自然な質問やユーザー発話の誤認識などがあり、これらは質問生成システムや音声認識が原因だと考えられる。Q4, Q5, Q9, Q18, Q21 は、話が遮られたや、反応が遅いなどの参加者からのコメントがあり、相槌やターンテイキングに問題があると考えられる。また、句末音調分析システムの Julius の処理に時間がかかっており、反応の遅さにつながっていると考えられる。システムの応答正解率は 59.6%、姿勢ミラーリングに気づいた参加者は 38.3% であった。

3.3 対話例

実験での対話例の一部を表 2 に示す。User はユーザー発話、Sys はシステム応答をあらわす。

4 むすび

姿勢ミラーリング効果を伴う 3D 視マルチモーダル傾聴対話システムを開発し、60 名の実験参加者から評価を

表 2 実験での対話例

話者	内容
Sys	こんにちは！お昼食べたら眠くなってきました…私はリンと います。あなたの名前を教えてくださいませんか？
User	NAME です
Sys	NAME さんですね？よろしくお願ひします！NAME さんの 最近あった楽しい出来事、または悲しい出来事について教えて いただけませんか？
User	最近あった楽しい出来事は焼きそばを上手く作れるように なったことです
Sys	なるほど、それは尊敬しますね。
User	もともと家を作るときにピシャピシャに行ってしまうとべ ちゃべちゃであんまり美味しくなかったんだけど
User	ちょっと2 覚えてうまく作れるようになりました
User	あなたの焼きそばがすごいですね
User	一人暮らししても大丈夫かなって
Sys	なるほど、大丈夫かなってなのですね。
User	焼きそば食べたことありますか
Sys	はい、焼きそば食べたことありますよ
User	美味しいよね
Sys	ふむふむ、美味しいのですね。それは幸せですね。何で美味し いのですか？
User	簡単に作れるし美味しいのはね
Sys	へえ〜、簡単に作れるし美味しいのなのですね。それは幸せで すね。何で簡単に作れるし美味しいのなのですか？
User	簡単に
Sys	簡単になのですね。良かったですね！誰に簡単に作れるし美味 しいのなのですか？
User	そうだねあとは最近カット野菜
Sys	はい、カット野菜です。
User	すぐ調理できる野菜っていうのが売ってるのでそれを使うと すごく簡単に作れるんだよ
User	可愛いね
Sys	へえ〜、何に可愛いのですか？
User	えりさんが可愛いなって思いました
Sys	どうして可愛いなって思いましたか？あ、そろそろ時間です ね。お話ししてくださってありがとうございます！

得た。実験参加者のアンケート結果では多くの評価項目にて肯定的な評価が得られた。特に話の聞き方と共感の示し方の評価が高く、アバターを親身だと感じた参加者が多いことがわかった。過半数の参加者はミラーリングされていることに気づかなかった。また、SR ディスプレイがアバターの印象に良い影響を与えたかというアンケートでは、7 段階で平均 6 という評価であり、SR ディスプレイがマルチモーダル対話システムに適していることが示された。話の聞き方と共感の示し方のユーザー評価が高い要因を分析したところ、アバターのこまめな反応や表情変化、姿勢ミラーリングのほか、SR ディスプレイによる立体的な表現が要因である可能性を示した。

関連発表

水谷林太郎, 鈴木寿, “姿勢ミラーリング効果を活用する 3D 視マルチモーダル傾聴対話システムの開発”,

第 97 回 人工知能学会 言語・音声理解と対話処理研究会, 2023/3/8 発表予定。

参考文献

- [1] 井上昂治, 山本賢太, 中村静, 高梨克也, 河原達也ほか. アンドロイド ERICA の傾聴対話システムにおける WOZ との比較評価. 人工知能学会研究会資料, SIG-SLUD-C002-21, 2020.
- [2] 下岡和也, 徳久良子, 吉村貴克, 星野博之, 渡部生聖. 音声対話ロボットのための傾聴システムの開発. 自然言語処理, Vol. 24, No. 1, pp. 3–47, 2017.
- [3] 青柳宏亮. 心理臨床場面でのノンバーバル・スキルに関する実験的検討—カウンセラーのミラーリングが共感の認知に与える影響について—. カウンセリング研究, Vol. 46, No. 2, pp. 83–90, 2013.
- [4] 吉田裕, 神田智子. 姿勢一致エージェントの開発と対話におけるラポール形成. HAI シンポジウム, 2D-5, 2012.
- [5] 香曾我部多門, 山田節夫, 西野哲朗ほか. 相づちのリアルタイム性の分析と自動生成アルゴリズムの設計. 研究報告数理モデル化と問題解決 (MPS), Vol. 2017, No. 24, pp. 1–6, 2017.
- [6] 河原大輔, 黒橋禎夫. 格フレーム辞書の漸次的自動構築. 自然言語処理, Vol. 12, No. 2, pp. 109–131, 2005.
- [7] 古川智雅, 吉野幸一郎, 須藤克仁, 中村哲. 曖昧性を持ったユーザ発話に対する格フレームを用いた聞き返し発話候補の生成. 言語処理学会 第 24 回年次大会 発表論文集, pp. 905–908, 2018.
- [8] Francis Bond, Timothy Baldwin, Richard Fothergill, and Kiyotaka Uchimoto. Japanese semcor: A sense-tagged corpus of japanese. In *Proceedings of the 6th global WordNet conference (GWC 2012)*, pp. 56–63, 2012.
- [9] Akinobu Lee, Tatsuya Kawahara, and Kiyohiro Shikano. Julius—an open source real-time large vocabulary recognition engine. 2001.
- [10] 波多野博顕, 石井カルロス寿憲. 日本語自然対話に現れる質問発話の句末音調. 音声研究, Vol. 21, No. 1, pp. 1–11, 2017.
- [11] Marianne La France and William Ickes. Posture mirroring and interactional involvement: Sex and sex typing effects. *Journal of Nonverbal behavior*, Vol. 5, pp. 139–154, 1981.