

博士論文

日常的な継続利用意欲の向上を目的とした TV 雑談ロボットのインタラクションデザイン

西村 祥吾

2021年3月17日

奈良先端科学技術大学院大学
情報科学研究科

本論文は奈良先端科学技術大学院大学情報科学研究科に
博士(工学) 授与の要件として提出した博士論文である。

西村 祥吾

審査委員：

加藤 博一 教授	(主指導教員)
小笠原 司 教授	(副指導教員)
萩田 紀博 教授	(大阪芸術大学)
神原 誠之 准教授	(副指導教員)

日常的な継続利用意欲の向上を目的とした TV 雑談ロボットのインタラクションデザイン*

西村 祥吾

内容梗概

現実世界において人々の日常的な発話が不足している。この問題を解決するために人に代わってコミュニケーションロボット（以下、対話ロボット）が話し相手となることで、人の日常的なコミュニケーションを促進する取り組みが注目されている。日常生活において自由なテーマについて雑談を行うロボットに関する既存技術では、ロボットが対話破綻を起こしてしまうことなどによってユーザがロボットに対して悪印象を与える恐れがある。このような状況ではユーザが日常的にロボットを使い続けたいと思う意欲（以下、継続利用意欲）を保つことが困難である。そのため、ユーザが継続的に日常生活の中で対話ロボットとコミュニケーションを取るために、ユーザの“継続利用意欲”が高く維持するための技術的なアプローチを確立する必要がある。本研究の目的は、継続利用意欲を促進するためロボットの振る舞いを決定するインタラクションデザインを提案し、その妥当性を示すことである。本研究では対話ロボットとして、テレビを見ている人と共に対話を行う TV 雑談ロボットに着目する。このロボットは TV という特定のテーマに関する発話を TV 番組に対して投稿される SNS 上のコメントを用いて発話を行うため、対話内容が文脈に合い且つ新規性の高い対話を実現することを目的として設計された。短期的な TV 雑談ロボットの利用を通して、ユーザがロボットに対し継続利用意欲を高めるための手法を提案する。TV 雑談ロボットに対し実装する機能として、人と人のインタラクションに関する心理学的知見を参考に、自然な振る舞いと共感を示す振る舞いの 2 つについて調査する。TV 雑談

*奈良先端科学技術大学院大学 情報科学研究科 博士論文, 2021 年 3 月 17 日.

ロボットの自然な振る舞いを実現する機能として、応答を示す発話、トピックに同期する発話の2つを調査する。このうち応答を示す発話では、TV雑談ロボットに対しユーザの発話相槌や復唱など素早く応答する手法を組み合わせるための手法について調査する。トピックに同期する発話は、TV雑談ロボットがユーザと雑談する際の話題に対応したコメントを発話するための手法について調査する。

TV雑談ロボットがユーザに対し共感を示す振る舞いを実現する機能として、嗜好を反映する発話、情動共有、韻律模倣の3つを調査する。嗜好を反映する発話では、ロボットに発話させるSNSコメントを選択する際にユーザ個人の趣向を反映する手法について調査する。情動共有では、興奮や笑い等の雰囲気にあった振る舞いをロボットが行うことによって、ユーザと感情を共有する手法について調査する。韻律模倣では、人の話し方や振る舞いといった非言語・パラ言語情報を模倣して発話を行うことで、ユーザに同調する手法について調査する。

各提案手法がユーザの継続利用意欲に与える影響を検証するために、それぞれの手法が適用されたロボットについて、主観および客観的指標による評価実験を行う。実験の結果から、TV雑談ロボットにおいてユーザの継続利用意欲を向上させるためには、ロボットの自然な振る舞いとロボットがユーザに対し共感を示す振る舞いの実現を目指した設計が効果的であり、ユーザの継続利用意欲を基準にTV雑談ロボットのインタラクションをデザインすることの妥当性が示された。

キーワード

対話ロボット, 雑談, テレビ視聴, 継続利用意欲, ヒューマンロボットインタラクション

Interaction Design for TV Chat Robot to Boost a Daily-Use Motivation*

Shogo Nishimura

Abstract

The lack of opportunities for daily conversation is one of a big social issue. In order to solve this problem, efforts to promote daily communication of people are attracting attention by having communication robots talk to each other instead of people. According to the conventional technology related to communication robots that chat about free themes in daily life, there is a risk that the user will give a bad impression to the robot due to a dialog failure of the robots. In such a situation, it is difficult for the user to maintain the motivation to use the robot on a daily basis. Therefore, in order for the user to continuously communicate with the dialog robot in daily life, it is necessary to establish a technical approach that keeps users highly motivated to continue using dialog robots.

In this research, I focus on a TV chat robot that has a dialog with a person watching TV. This robot was designed for the purpose of realizing a highly novel dialog in which the dialog content fits the context. TV chat robot can utter on a specific theme of TV using comments on SNS posted for TV programs. I propose methods for users to motivate to use the TV chat robots through the use it in short-term interaction. As functions to be implemented for TV chat robots, I investigate two behaviors, natural behavior and behavior that shows empathy with a user, with reference to psychological knowledge about human-to-human interaction. I show the investigation about two functions that realize the

*Doctoral Dissertation, Graduate School of Information Science, Nara Institute of Science and Technology, March 17, 2021.

natural behavior of TV chat robots: utterances that show a response to a user and utterances that synchronize with a topic. And I also show the investigation about three functions that realize the behavior that shows the empathy with a user: utterances that reflect user's preference, emotional sharing, and vocal synchrony. To verify the effect of each proposed method on the user's motivation to use the TV chat robot, evaluation experiments are be conducted using subjective and objective indicators for the robot to which each method is applied. The results show that in order to boost the user's motivation to use the robot, it is effective to design the robot with the aim of realizing the natural behavior of the robot and the behavior that the robot shows empathy with a user. It is also shown that the validity of designing the interaction design of the TV chat robot based on the user's motivation to continue using the robot.

Keywords:

dialog robot, chat, watching television, motivation to use, human-robot interaction

目次

1. 序論	1
2. 研究の位置付け	4
2.1 継続利用意欲を促進する対話ロボットのインタラクショndeザインに関する先行研究	4
2.2 TV 雑談ロボット	5
2.3 本研究と位置付け	6
3. 自然な振る舞いを実現する手法	11
3.1 応答を示す発話	12
3.1.1 4つの対話機能を複合した雑談システムの実装	12
3.1.2 雑談システムの評価	17
3.1.3 評価結果	20
3.1.4 考察	21
3.2 トピックに同期する発話	25
3.2.1 TV 雑談ロボットがユーザに与え得る不快要因の分類	25
3.2.2 タイムシフトによる遅延問題の軽減手法	26
3.2.3 タイムシフトを適用したロボットの評価	27
3.2.4 評価結果	29
3.2.5 考察	32
4. 共感を示す発話を実現する手法	34
4.1 嗜好を反映する発話	35
4.1.1 ソーシャルメディアコメントの評価手法	35
4.1.2 ソーシャルメディアコメントに対する評価データ収集実験	39
4.1.3 相撲における解析結果と考察	48
4.1.4 ニュースにおける解析結果と考察	53
4.2 情動共有	64
4.2.1 対話ロボットの情動共有に関する先行研究	65

4.2.2	雰囲気の推定とロボットの振る舞いを決定する手法	66
4.2.3	盛り上がりにおける情動共有の評価と考察	71
4.2.4	笑いにおける情動共有の評価と考察	76
4.2.5	盛り上がり雰囲気における情動共有の客観評価と考察	81
4.3	韻律模倣	90
4.3.1	非言語コミュニケーションにおける同調対話に関する先行 研究	90
4.3.2	韻律模倣の実装	90
4.3.3	評価方法	98
4.3.4	評価結果	104
4.3.5	考察	107
5.	考察と研究の限界	112
6.	結論	116
	参考文献	120
	付録	133
A.	研究実績	133
A.1	学術論文誌	133
A.2	査読付き国際会議・ワークショップ	133

目 次

1	身体性を伴った TV 雑談ロボットの使用時の様子	7
2	TV 雑談システムの概要図	8
3	スマートフォン上で動作する TV 雑談ロボット	9
4	研究の対象とする範囲の切り分け（黒影の範囲内は本誌では調査 しない要素を示す）	10
5	4つの対話機能を適用した TV 雑談ロボットの対話フロー	16
6	4つの対話機能の位置付け	17
7	提案する 4つの対話機能を適用した TV 雑談ロボットとの対話例	18
8	実験の外観図	20
9	実験で使用された TV 雑談ロボットの制御システム	21
10	ユーザ発話数の平均値と分散	22
11	質問における 5段階評価の平均値と分散	23
12	TV 雑談ロボットのリアルタイム利用により発生する 2つの問題	25
13	実験時の TV 雑談ロボットの実装	28
14	各 TV 番組視聴後のアンケートスコア	29
15	条件間のアンケートスコアの差	30
16	Dabelive のシステム概要	36
17	放送局の選択画面	37
18	設定画面	38
19	スマートフォン上の TV 雑談ロボットとの対話画面	39
20	ユーザ継続利用意欲モデルの作成フロー	41
21	ユーザ継続利用意欲モデルによる継続利用意欲継続利用意欲の推定	41
22	トピック抽出の模式図	43
23	実験で使用したユーザ継続利用意欲モデル	44
24	実験用に改良した Dabelive の対話画面	46
25	Russell の円環モデルにおける理由ボタンの配置	46
26	推定した継続利用意欲評価値の平均二乗誤差 (相撲), 1182 次元	48
27	興味度合い毎に推定した継続利用意欲評価値の平均二乗誤差 (相撲)	50

28	継続利用意欲評価値のヒストグラム データセット:相撲, 回帰手法:SVR(RBF カーネル)	51
29	各興味度合いの継続利用意欲評価値ヒストグラム データセット:相撲, 回帰手法:SVR(RBF カーネル), 誤差 0.5 未満のみをプロット	52
30	推定した継続利用意欲評価値の平均二乗誤差における相撲とニュースの比較	53
31	レベルごとの盛り上がり時のロボットの振る舞い	68
32	レベルごとの笑い時のロボットの振る舞い	69
33	アピランスの違いごとの実験時の様子	72
34	ロボットの盛り上がり時の表現に対する各質問における条件ごとのスコア	73
35	盛り上がり評価実験における各条件の組み合わせごとの比較 (左上: 情動共有なし条件・バーチャルロボットと情動共有なし条件・フィジカルロボット, 右上: 情動共有なし条件・バーチャルロボットと情動共有あり条件・バーチャルロボット, 左下: 情動共有なし条件・フィジカルロボットと情動共有あり条件・フィジカルロボット, 右下: 情動共有あり条件・バーチャルロボットと情動共有あり条件・フィジカルロボット)	74
36	ロボットの笑い時の表現に対する各質問における条件ごとのスコア	77
37	笑いにおける各条件の組み合わせごとの比較 (左上: 情動共有なし条件・バーチャルロボットと情動共有なし条件・フィジカルロボット, 右上: 情動共有なし条件・バーチャルロボットと情動共有あり条件・バーチャルロボット, 左下: 情動共有なし条件・フィジカルロボットと情動共有あり条件・フィジカルロボット, 右下: 情動共有あり条件・バーチャルロボットと情動共有あり条件・フィジカルロボット)	79
38	SNS コメント数の推移とモデルの出力例 (上図), 4段階に分類された TV 番組 (サッカー中継) 中の盛り上がりレベルの例 (下図)	82
39	TV 雑談ロボットとユーザ間のインタラクションの例	83

40	提案モデルにより推定された盛り上がり度(青線)と10秒間遅延させた ground truth の推移(赤線)	84
41	防音室内での実験の様子	86
42	アフェクトグリッドにより評価された被験者の覚醒度(左図)と感情価(右図)の平均値と標準偏差 (**, $p < 0.005$)	87
43	生理学的反応の平均値と標準偏差: スキンコンダクタンスレベル(左上図), 大頬骨筋(右上図), および皺眉筋(下図)の筋電位 (*, $p < 0.05$)	88
44	被験者がロボットに対して盛り上がっていると感じた度合い(左図)と、ロボットが共感してくれていると感じた度合いの平均値と標準偏差(右図) (***, $p < 0.001$)	89
45	各実験条件下での被験者の再選択率の平均値と標準偏差 (**, $p < 0.01$; *, $p < 0.05$)	89
46	推定された話速の値と VoiceText の話速に関するパラメータの関係	93
47	推定されたピッチの値と VoiceText のピッチに関するパラメータの関係(音声モデル: takeru)	94
48	推定されたピッチの値と VoiceText のピッチに関するパラメータの関係(音声モデル: hikari)	95
49	推定された音量の値と VoiceText の音量に関するパラメータの関係(音声モデル: takeru)	96
50	推定された音量の値と VoiceText の音量に関するパラメータの関係(音声モデル: hikari)	97
51	韻律同調システム評価のためのアンケートスコア	99
52	シナリオの例	100
53	アフェクトグリッド	101
54	シナリオの評価結果	102
55	シナリオの読み方に対する指示の表記例	103
56	Q1 から Q7 の平均および分散	106
57	Q7 のシナリオセットごとの結果	106

58	各生理指標における平均値および分散	107
59	男性における Q1 から Q6 の平均値と分散	107
60	女性における Q1 から Q6 の平均値と分散	108
61	提案した振る舞いが継続利用意欲を促進した理由	113
62	今後の課題として調査すべき範囲（黒影の範囲内）	114

表 目 次

1	実験の実装に用いられた各種ツール	24
2	実験から得られた数値的結果	24
3	実験から得られた数値的結果	31
4	ソーシャルメディアコメントの属性情報	40
5	使用する音声合成 API	40
6	収集したユーザの継続利用意欲評価が付加された発話文	47
7	視聴後アンケートの項目	47
8	各興味度合いでのデータセット詳細 (相撲)	49
9	発話文の選択例	56
10	雰囲気に応じたロボットの振舞いと要素	64
11	盛り上がりレベルと対応するロボットの発話時の韻律特徴	83
12	韻律同調システム評価のための客観データ	98
13	各質問に対する対応ありの数値的結果と効果量	105
14	各生理指標における対応ありの t 検定の数値的結果と効果量	105

1. 序論

年齢を問わず独居世帯が増加していく現代社会の中で現実世界における日常的な発話が不足している [1, 2, 3]. 現実世界において発話が不足することは幸福感の欠如やうつといった精神的問題の原因となることに止まらず、高齢者の場合は認知症発症のリスクが高まるなど、致命的な健康問題を引き起こす可能性がある [4]. 一方で新型コロナウイルスの感染拡大に伴い、世界的に Social Distancing (社会的距離拡大戦略) が求められ、人々の行動や企業活動にも制限が加わるような社会になった [5]. このような状況の中で、人に代わって自律的に行動するコミュニケーションロボット (以下、対話ロボット) が話し相手となることで、人の日常的なコミュニケーションを促進する取り組みが注目されている [6, 7, 8]. ここで、対話ロボットに対してユーザが日常的にコミュニケーションを継続するためには、そもそもユーザが対話ロボットを日常的に利用し続けたいと思う“継続利用意欲”が高く維持され続ける必要がある. この対話ロボットに対する継続利用意欲を促進するための対話ロボットの適切な振る舞いを決定する手法を提案し、対話ロボットデザインに関する体系的知識を構築するための一部となることを本研究の目的とする. この知識体系を確立することで、現実世界において日常的にコミュニケーションを行うユーザとロボットのペアを増加できることが期待される. このメリットとしてロボットからユーザに適切な働きかけを行うことができるようになり、ユーザの行動変容や自立支援、販売促進を実現し、今後の社会に対する新たな価値を生み出すと考えられる.

ロボットが長期的に使い続けられることで、ある対象への反復接触がその対象への好意度を高める単純接触効果 [9, 10] といった心理的作用により心理的に近い関係性を人とロボットの間で結ばれ、信頼関係構築に繋がる. 人とロボットの信頼関係構築に関する成功例としてソニーの犬型ロボット AIBO [11, 12, 13] が挙げられる. この例に見られるように、人とロボットのインタラクションが長期的に続くことで信頼関係は構築が可能であることが示されている. 日常的なユーザからの発話機会を創出するために、ユーザの日常的生活空間において話し相手となるように設計されたロボットの開発が行われてきた. 対話ロボットにはタスク指向型の対話を行うものと非タスク指向型のものとに二分される.

タスク指向型対話ロボットは、ロボットが請け負ったタスクを完遂するために行う対話である。例として Pepper を用いた店頭販売のロボットや、Google Home[14] や Amazon Echo[15] などの AI スピーカーを用いた情報検索が挙げられる。これらのロボットはタスクが決まっているため対話内容が限定される。そのため、非タスク指向型に比べて発話文の生成難易度は低い。非タスク指向型対話とは目的のない対話のことであり、主に雑談に該当する。話題が限定されないため話題に適した発話文の生成や、ユーザの発話に対して意味的に破綻しない返答の生成が課題となっている。Higashinaka ら [16] は、人とロボットの対話において、対話の破綻を防ぐためにロボットの不適切な発話を検出する課題に取り組んでいる。具体的には、不適切な発話の検出手法を確立するために、対話データセットの構築や発話文の評価法、対話破綻検出のワークショップの結果報告を述べている。このように、現在までに研究されている非タスク指向型対話ロボットの主な評価手法は、発話の意味的整合・文脈崩壊の防止に焦点が当てられているが、これを十分に満たす雑談対話手法は確立されていない。

本研究では短期的なインタラクションを通して非タスク指向型対話ロボットが日常的に話し相手としてユーザに認識され、魅力的な存在となるための複数の手法を提案する。多くの非タスク指向型対話ロボットは、ユーザから能動的に話しかけられることを前提に設計されている、あるいは事前に用意されたシナリオを元に発話を行うものが多い [17, 18, 19]。また、これら雑談型対話ロボットは応答文の意味的破綻こそ改善されてきたものの、ユーザの興味を引くような魅力的な対話文を生成し続けることが困難なため、日常的に言葉を交わし続けるような長期にわたる対話には対応が困難である。

本研究では日常生活で利用される対話ロボットの形の一つとして、TV を切り口として考える。我が国において TV 視聴という行為は多くの人々の日常生活中に浸透しており、これは独居の人々にも同様の傾向が見られる。この日常的なシーンの一つである TV 視聴時にユーザとインタラクションを行う TV 雑談ロボット [20] の利用に着目する。TV 雑談ロボットを用いるメリットとして、以下の 2 つが考えられる。1 つは TV を見ている際にロボットが話し相手となるため、ユーザが友達と TV を見ているような状況を作り出すことができ、自然な形でインタ

ラクションを行う環境を提供できる点である。もう1つのメリットは、システム構築の観点からTVという特定のテーマについてロボットの発話内容を決定することに特化できるため、フリーテーマの雑談と比較して対話内容が文脈に合い且つ新規性のある対話を実現できる可能性が高い点である。これらのメリットを持つTV雑談システムに対し、ユーザの日常的なロボットへの継続利用意欲を促進するための振る舞いを実現する5つのアプローチについて実装・評価を行なう。

本研究では2章にて継続利用意欲を促進する対話ロボットの先行研究について述べ、研究の位置付けを述べる。続く3章、4章では本研究にてユーザからの継続利用意欲を向上するために、TV雑談ロボットに適用するアプローチの提案とその評価について詳述する。5章で各アプローチに対する考察と研究の限界についてまとめ、6章で研究の結論と今後の展望を述べる。

2. 研究の位置付け

2.1 継続利用意欲を促進する対話ロボットのインタラクシオンデザインに関する先行研究

ロボットの利用が家庭で一般的になりつつある中で、消費者の目線からこれからのロボットに求められる要素としてより自然なインターフェースとより魅力的なインタラクシオンを提供することが必要であるとされる [21]. 社会的な行動を行なうソーシャルロボットの設計に関するガイドラインを提案する研究によると、ロボットの特性は形状、モダリティ、社会規範、自律性、および対話性によって構成される [22]. 宮下らは、ショッピングモールに案内ロボットを設置し、約一ヶ月ショッピングモールを訪れた人達と接した結果、実験終了時でも開始時と同様に、ショッピングモールを訪れる人達はロボットに親しみを抱いている事が明らかになった [23]. Sabelli は、介護施設に対話ロボットを設置し、4ヶ月間高齢者がロボットとどの様に接したかを観察した. この結果、ロボットに対してどの様に接したら良いか分からないでいる高齢者には、ロボットから挨拶や名前を呼ぶ等の行為を行うことで距離が縮まり、良い関係が築けることが明らかになった [24]. Heerink らは、高齢者が対話ロボットを利用する際に、対話ロボットを利用する楽しさと利用意思、利用意思と実際に利用する事の間にはどちらも相関がある事を示し、楽しさを感じる事が高齢者に対話ロボットを使用させる為には重要な要素になる事を明らかにした [25]. また、Miyazawa ら [26] は対話ロボットがユーザの日常的な利用意欲を向上するための要素として、対話相手の応答の自然さと発話内容の面白さが重要であると述べている. 以上からユーザが継続して対話を行いたいと考えるロボットを開発するには、“自然な応答”と“対話内容の面白さ”を付加し、より人間らしさを感じさせる仕組みを導入する必要があることがわかる.

このように、対話ロボットとの長期的な利用をユーザに促すためには、対話を通してロボットがいかにユーザと良好な関係（ラポール）を構築できるかが問題となる [7], [8]. この問題解決に向けたアプローチとして、インタラクシオンロボットが対人コミュニケーション戦略を取り入れ、人間性を高めることが効果的であると言われている [27, 28].

人とエージェントのインタラクションの継続性について、適応ギャップという観点から、ロボットの外見がどのように設計されるべきかについて述べられた研究がある [29]. なお本研究でエージェントはロボット同義の存在として扱う. 適応ギャップとは、ロボットの外観から人が想定するロボットの機能と、実際にロボットが持つ機能との間に存在するギャップである. 人とロボットが継続してインタラクションを行うには、ロボットの外見から人が想定する機能が実際の機能よりも低くなるように、設計を行う必要性が述べられた. 同様にこの適応ギャップにより、ロボットの機能によっては外見を人間や動物に近づける必要がない場合があることが示されている [30]. その一方、ユーザーの意思決定に対しロボットが説得を行うとき、身体性を伴ったロボットを適用した場合と、画面上に映るバーチャルなロボット適用した場合の差について調査が行われた [31]. この調査から、身体性を伴ったロボットが、バーチャルなロボットに比べ説得の効果が高いことが示された.

2.2 TV 雑談ロボット

私は研究の背景となる社会的問題として、人々のコミュニケーション不足があり、これを対話ロボットによって補う必要があると考えた.

本研究のゴールであり理想は、ユーザが日常生活シーンのどの場面においてもロボットと雑談を行うモチベーションが保たれることにある. この理想に到達するための段階として、人と雑談を行う可能性の高いと考えられる日常的な生活シーンを挙げる.

- 団らん時（食事，TV 視聴など）
- 作業時（料理，掃除，洗濯などの家事，身支度など）
- 外出時（車などモビリティを用いた，あるいは徒歩による移動，外食，買い物など出先でのイベント）

この中の特定の生活シーンにおいてのみ、ロボットとの雑談からロボットとのインタラクションを始めることで、ユーザのロボット利用に対する心理的障壁を緩

和すると考えた。ここから徐々に理想とする状態に近づけることでロボットの継続利用促進から愛着の醸成に繋げるという戦略の元、特定のシーンで人との関係を構築するロボットの提案が必要であると考えた。本研究では、この特定の生活シーン中から TV 視聴を選択した。TV 視聴を選択するメリットとして雑談のトピックが選択しやすいこと、視聴する番組が多くの場合ユーザの嗜好に合っており対話の中から嗜好を引き出す手間が省けること、トピックが番組内、番組間で更新されるためユーザを飽きさせない対話文を生成しやすいこと、3つが挙げられる。特に3番目の理由に関しては、すでに SNS コメントを使用したロボット発話に一定の効果が認められている [20]。生活中的対話ロボットとの雑談というタスクに着目した際、ロボットが共通の話題を提供することや、それによってユーザが楽しみを得られることなどが要求として挙げられる中で、それらの条件にマッチする手段として TV 雑談ロボットが適切であると判断した。

TV 雑談ロボットの利用時の様子を図 1 に示す。TV 雑談ロボットは TV 番組を共通の話題として、ユーザが視聴中の TV 番組に関連するコメントをロボットが能動的に発話するというコンセプトで設計された [20]。このコメントは TV 雑談サーバ [32] を通して一定の間隔でロボットに配布される。図 2 に示すように、TV 番組に対して SNS 上でコメントを書き込むユーザは我が国では多く存在する。TV 雑談サーバは各局に対して SNS ユーザが投稿するコメントをリアルタイムに収集することができる。TV 雑談ロボットは図 3 に示すように、身体性を伴うものに限らずスマートフォンやブラウザ上のキャラクターとしても動作し、株式会社 amirobotech[33] のサービスとして公開されている。ユーザは視聴している番組のチャンネルに合わせてロボットに局名を指定することで、その局に投稿されているコメントがロボットに送信され、発話を行うことができる。

2.3 本研究と位置付け

本研究では 2.2 にて述べたように対話ロボットとして TV 雑談ロボットを用いる。日常的な継続利用の先にあるものとして、私はユーザとロボットの間で愛着のような深い関係が構築されると考える。このようにして築き上げられた関係に基づき、ロボットからの能動的な働きかけによって、ユーザの行動変容といった

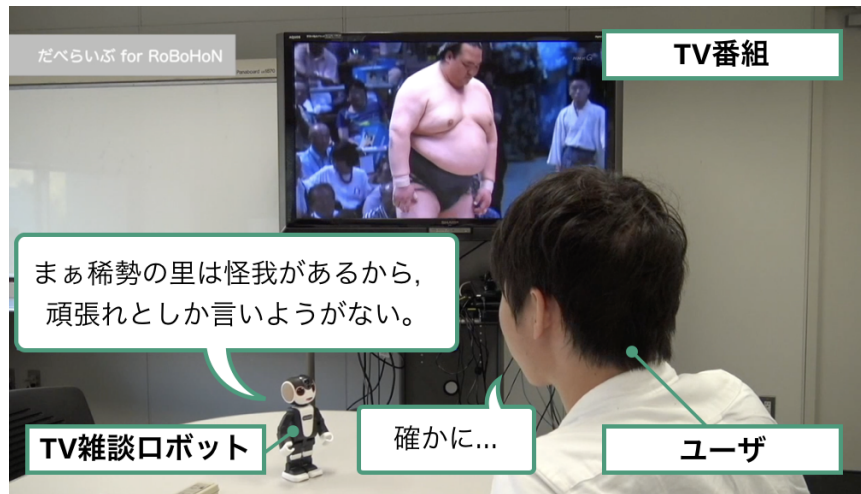


図 1 身体性を伴った TV 雑談ロボットの使用時の様子

健康促進の支援，QoLの向上等の促進を実現することが研究の最終的なゴールである．本研究の位置付けは，この最終目標に向けた前段階として，ユーザとの短期的な関係の中でどのようなロボットが求められるかを調査することである．つまり本研究の目的は，TV視聴という特定の生活シーンにおいて継続的に利用され続けるTV雑談ロボットとはどのようなロボットであるべきかを明らかにすることである．ここで研究を分類するための2つの軸について説明する．1つはインタラクションのどの時点でロボットが人に与える印象を調査するかを分類するために「ロボットの利用期間」を軸として用いる．これはロボットの利用期間が短期・長期のそれぞれについて，継続利用意欲の調査が必要であると考えたためである．短期間の調査では，ロボットに対する初見での印象から，ユーザが「ロボットをまた使いたい」と思う要素ごとの影響について分析を行う．長期間の調査では，ロボットとの体験の積み重ねと時間経過の中でユーザが「ロボットを使い続けたい」と思う要因の分析によって行われるべきであると考えた．もう1つの軸には，調査の対象とするロボットのインタラクションを分類するために「継続利用意欲を向上すると予想されるロボットの振る舞い」を選択した．この軸の中でさらに本研究ではロボットの「自然な振る舞い」，と「共感を示す振る舞い」をピックアップした．

「自然な振る舞い」は，2.1でも述べたように対話システムの自然な発話はユー

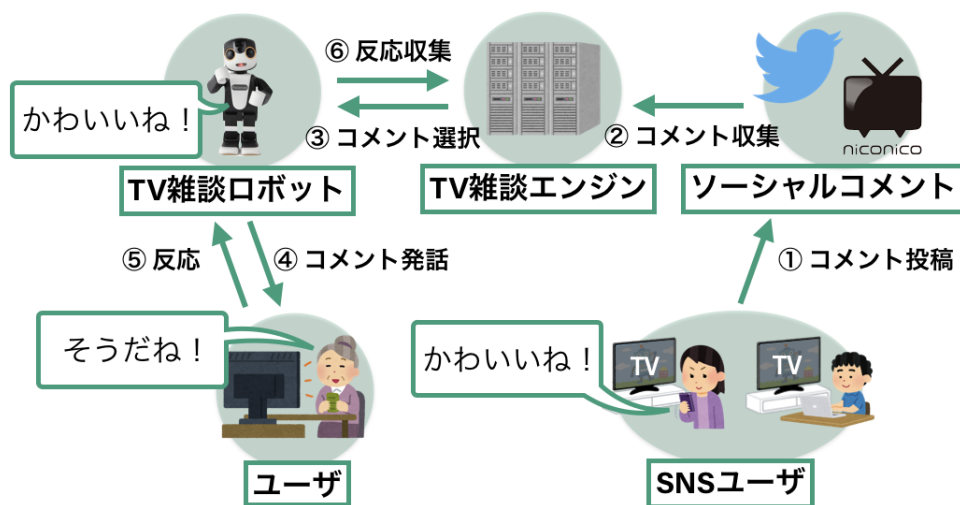


図 2 TV 雑談システムの概要図

ザの継続利用意欲を向上する要素であることが既存研究から示されていることからテーマとして選択することとした [26, 27, 28]. 「自然な振る舞い」の具体的な要素として現時点においてTV雑談ロボットに実装することが容易であり、実装による効果が大いに期待できる「3.1 応答を示す発話」と「3.2 トピックに同期する発話」を実現する手法を提案する. 一方「共感を示す振る舞い」においては、心理学的な知見から、人同士の関係において行動の同期などの共感を生む振る舞いは調和的な関係の構築やパートナーへの好感、共感的な行動高めることが明らかになっている [34, 35, 36]. この知見に基づき、ユーザーに対して共感を示す振る舞いを実装したロボットが継続利用意欲の向上に貢献することが期待される一方で、その効果が明らかになっていないことから、調査が必要であると考えた. 「共感を示す振る舞い」の具体的な要素としては、「4.1 嗜好を反映する発話」、「4.2 情動共有」、「4.3 韻律模倣」の3つを本研究で提案する要素として選択した. これらの振る舞いを選択した理由として、「自然な振る舞い」と「共感を示す振る舞い」を実現するための対象として、TV雑談ロボットが適切であると判断したためである. 本研究で提案する上記の5つの振る舞いの要素を短期的なインタラクションを通して調査することを本研究の目的とし、2軸による研究の分類を図4行った. 5つの機能が適用されたTV雑談ロボットについて、それぞれの機能がユーザーに



図 3 スマートフォン上で動作する TV 雑談ロボット

与える印象を個々に評価し、継続利用意欲の向上にどのような影響を与えるかについて調査する。調査の前提として、発話とジェスチャ変化させるマルチモーダルな TV 雑談ロボットに対し、継続利用意欲を高めると考えられる振る舞いについて実装を行う。ここでモーダルごとの重なりが示す部分の重要性についてではなく、振る舞いの要素ごとの影響を評価を行う。これは提案する各手法が主に心理学的根拠に基づいて設計されているため、提案する要素ごとの影響を調査するという心理学的研究の方針に従ったためである。

評価実験は、ロボットとの長期間に渡るインタラクションの観察を通してではなく、短期間のインタラクションを通して、アンケートを用いて主観的な継続利用意欲の測定を行う。ロボットの日常利用を想定した際、長期的な人とロボットとの関係を調査することは不可欠とされているが、長期的に日常生活中でロボットを動作させる試みが行われた調査数は少ない [37, 38, 39]。一方、人とロボットの短期的な関係性の調査からも、ロボットの要素行動ごとの影響について主観・客観的評価による分析を通して、ロボットの持つ第一印象が人に与える影響や、初期の過程の関係性構築において、ロボットに不足している機能や改良点の特定に効果的である [40]。本研究では、ロボットとユーザ間の親密で深い関係の構築、

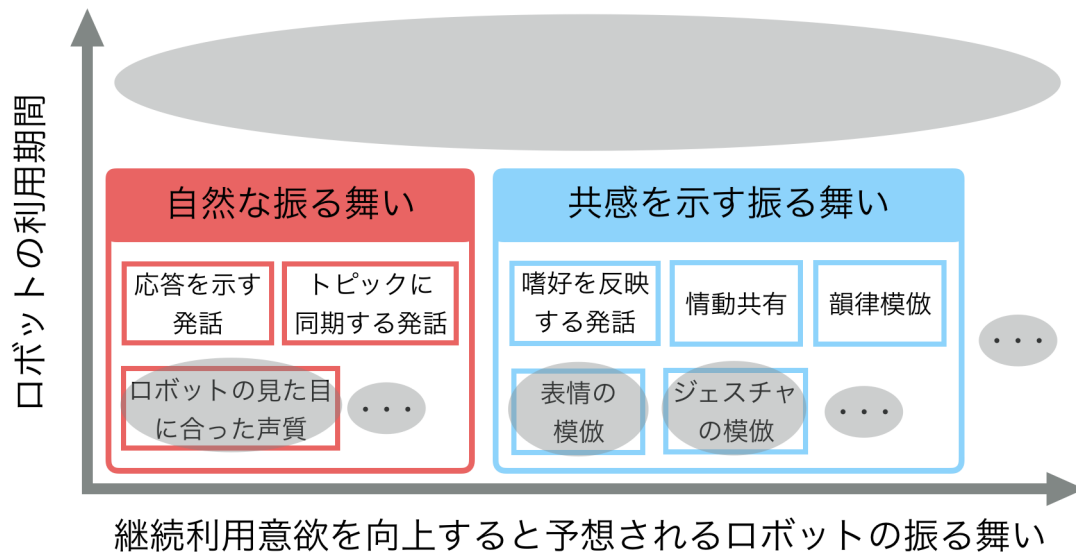


図 4 研究の対象とする範囲の切り分け（黒影の範囲内は本誌では調査しない要素を示す）

あるいはユーザのロボットに対する愛着の醸成を最終的な目標とするが、本誌ではその目標に向け人とロボットの関係が始まる初期段階に焦点を当てる。従ってロボットが初見時にユーザに対し、また使いたいという印象を与えることが可能かを、提案する機能ごとに評価するため、評価実験は短期間のロボットとのインタラクションを通して行う。

3. 自然な振る舞いを実現する手法

本章では、2.3で述べたように、対話システムの自然な発話はユーザの継続利用意欲を向上する要素という知見を基に、TV雑談ロボットがユーザに対して自然な振る舞いを実現するための手法を提案する。「自然な振る舞い」を実現するための要素としてTV雑談ロボットに対し実装が現実的であり、最も効果的であると考えられる以下の2つの手法を提案し、継続利用意欲への影響について短期的なインタラクションを通して評価を行う。

- 応答を示す発話 3.1
- トピックに同期する発話 3.2

応答を示す発話では、ロボットがSNSコメントを発話するだけでは雑談ができている、という実感をユーザが感じることができないという問題に対処するための手法を提案する。

トピックに同期する発話では、ロボットの発話を、ユーザと共有するトピックに同期させた際の印象評価を行う。具体的にはロボットの発話内容が視聴する番組のトピックと一致しない原因を排除し、またロボットの発話やTV音声が重複することでユーザに不快感を与え、継続利用を妨げる要因を同時に排除するため手法の提案する。

続く3.1, 3.2では提案する各手法について対応する先行研究、実装の詳細と評価結果、手法ごとの考察を述べる。

3.1 応答を示す発話

3.1.1 4つの対話機能を複合した雑談システムの実装

Miyazawa ら [26] は対話ロボットがユーザの日常的な利用意欲を向上するための要素として、対話相手の応答の自然さと発話内容の面白さが重要であることを明らかにした。ここで既存の対話ロボットの自然な応答速度と、対話内容の面白さに着目し、既存技術について紹介する。Matsubara ら [41] はロボットの頷き動作を制御する事でユーザがロボットに話しやすくし、ユーザの愚痴を傾聴するロボットを開発した。このロボットは、頷き動作を用いながらユーザの話を傾聴するもので、自然な応答速度で応答できている。しかし、このロボットは発話行動を行わないため発話内容の面白さはない。NTT docomo[18] は、発話文のテンプレートや話題辞書を用いた自動応答文生成と音声認識技術を用いた雑談エージェントであるしゃべってコンシェルを開発した。これは、キャラクターエージェントがユーザの発話内容を理解し、最適な応答を行う事ができる。しかし、自動応答文生成の機能を用いて生成された発話文は、テンプレート文を用いている事から発話文が単調なものになっている。また、応答生成を雑談エンジンにリクエストし、ロボットが返答を行うまでに一定の時間を必要とするため、自然な応答の速さとは言えない。

上で述べた2つの内容を統合したロボットとして、Kobayashi ら [42] が開発した傾聴ボランティアに代わって高齢者の話を傾聴し、話し相手となるロボットが挙げられる。このロボットは、機械応答を用いて会話をするもので、発話文のテンプレートや話題辞書を用いた自動応答文生成の機能、そして相槌と復唱機能を用いて話題提示と傾聴動作を行う。これによりユーザとの対話を実現し、自然な応答を可能にしている。しかし、自動応答文生成の機能を用いて生成された発話文は、テンプレート文を用いている事から発話文が単調なものになっている。以上のように、機械応答つまり相槌、復唱、自動応答文生成機能だけでは面白い発話文を生成することは困難であり、人と継続して雑談が可能なロボットは開発できないことが分かった。

自動応答文生成の技術を用いず、面白い発話文を生成する研究も行われている。ロボットの応答が単調となる問題を克服し、ロボットの発声の新規性を向上させ

るために、web から収集したテキストコーパスを使用した応答生成に関する研究が行われている [43, 44, 45, 46]. Mavridis らは [43, 44] 人とロボットのインタラクションを維持するために、ロボットとの対話履歴または Facebook を通じて共有された記事と友人情報を使用することで長期的な相互作用効果を示した. Yoshida らは [45] ユーザが知らない情報の連続して提供することで、ユーザーにとって対話により楽しくなることを示唆した. 新規性の高い情報を提供するための仕組みとして複数の言語リソースを使用したユーモアを備えた対話システムが提案された, このシステムでは、コメディや駄洒落で使用されるテクニックに基づいたユーモアによって面白い表現が行われていた. これらの既存のシステムでは対話ロボットの発話文の生成に対し、常に新しい情報をマニュアルで追加しなければならないことや、場に応じた適切な発話を行うことが求められた場合十分な応答困難であることなどのデメリットが挙げられる.

2.2で説明したように、本研究ではソーシャルネットワークサービス (以下 SNS) を仲介し、面白い対話文を提供するロボットを開発してきた [20]. このロボットは、ユーザが視聴しているテレビ番組の内容に関して実況している SNS 上のコメント (以下ソーシャルコメント) をロボットの発話情報源にする事で人が話す様なユーザの興味を引く対話文を生成する事ができる. これにより、常に新規性の高い情報をリアルタイムにロボットの発話リソースとして追加でき、かつ TV 視聴という状況に限定することで場に適した発話を可能とする. 一方で SNS 上のコメントのみを用いて人とインタラクションを行うことには限界がある. 例として、ユーザの発話文に対する応答は SNS 上で行われ、SNS においてユーザの発話文に対しコメントを書き込むまで応答を返す事が出来ず、自然な応答ができていないことが挙げられる.

本節で提案する機能は、SNS コメントを発話するだけでは雑談ができていない、という実感をユーザが感じる事ができないという問題を解決し、ユーザの継続利用意欲の向上を目的として提案された. 発話の自然さの向上に貢献する要素として、発話の仕方や視線、ジェスチャ、応答などが考えられるが、今回は要素を絞り、ユーザの発話に対して応答する機能を追加する方法を提案する. つまり、TV 雑談ロボットに対し、発話内容の面白さを残しつつ、自然な応答速度による

応答機能を組み合わせることで発話の自然さを高め、ユーザからのネガティブな印象を排除することを試みる。

既存研究において、システムの反応時間は短いものほどユーザに好まれる傾向があるとされ [47, 48, 49]、反応時間の速さがユーザに与える印象が既に調査されている。いわゆる 2 秒ルールと呼ばれる知見からは、0 秒よりも 1 秒、2 秒よりも 1 秒の方が好まれることがわかっている [50, 51]。これらの根拠をもとに、ユーザの問いかけや発話に対しては 2 秒以内で応答を示すことを優先するシステムの設計を行なった。

提案するシステムでは、Kobayashi ら [42] が応答性を得る為に用いた相槌、復唱、自動応答文生成、そして Takahashi ら [20] が提案する面白い発話内容を生成する為に用いたソーシャルコメントの機能を複合する。これら 4 つの対話機能の組み合わせることで応答性が良く発話内容が面白いロボットの実現が期待できる。図 20 は各対話機能を用いて TV 雑談ロボットの対話フローを示す。以下にて 4 つの対話機能について説明する。

- 相槌

相槌はユーザの発話終了時に優先的に実行され、実行によってロボットは“はい”、“うん”などと発話する。相槌はユーザの発話に対し即座に反応を返す機能であるため、応答速度は自然である。しかしロボットがユーザに対して話を聴いているという情報だけ伝えるので、発話内容の質は低い。相槌によってユーザは自身の発話をロボットが聞いてくれているように感じることが導く。ユーザの音声区間認識には Julius [52] を用いる。ユーザの発話から音声区間から一定長のポーズを検出した際に相槌を返す [53]。

- 復唱

ユーザ発話内容が Julius によってテキスト化され、そのテキストがキーワード抽出用の API [54] に送信される。復唱はこれにより取得されるキーワードを含んだ簡単な返答 (キーワード + “だね”) をする機能である。ユーザ発話にキーワードが存在しない場合、ロボットは“そうだね”と発話する。ロボットの発話内容は簡単なものだが応答する速さは自然である。人間の発話内容に関連したキーワードを復唱するため会話が伝わっている事

はユーザは認識できるが、発展的な対話には至らない。復唱によってユーザはロボットが自分の発言した内容を正しく認識したと感じさせることを導く。復唱によるロボットの発話は相槌の直後に実行される。

- 自動応答文生成

応答文の生成には NTT DOCOMO が提供する雑談エンジン [18] を利用して生成される対話文を用いて、ロボットに発話させる機能である。この機能による発話は SNS コメントが利用できない場合にのみ実行される。この機能によって違和感無く対話を続けられる発話文の生成が可能となり、発話内容の質は高い。しかしユーザへの対話文生成を雑談エンジンにリクエストし、ロボットが返答を行うまでに一定の時間を必要とするため、応答は自然な速さとは言えない。

- ソーシャルメディアコメント

ソーシャルメディアコメント（以下、SNS コメント）は SNS 上で人がテレビ番組について投稿するコメントを用いてロボットが発話を行う機能である。リソースとしてここでは SNS の一つである Twitter[55] を用いる。人間がリアルタイムに投稿した TV 番組に対するコメントを用いるため、発話内容の魅力（面白さ）は 4 つの機能の中で最も高い。しかし SNS 上にコメントが書き込まれない限り対話文を生成する事は出来ず、またユーザの発話を受けたタイミングで必ずしも反応を返せるとは限らないため、応答は自然な速さとは言えない。

図 6 に示されるように、ロボットの 4 つの対話機能はそれぞれ応答速度と面白さに関して互いにトレードオフの関係にある。それぞれの長所を活かし、且つ短所を補いあう事で自然な対話が可能なロボットができると考えられる。各対話機能の組み合わせにより、他の機能の欠点を補うことで応答性と新規性が高く面白い発話をユーザに提供し、ロボットとの継続的な対話に対するユーザのモチベーションが向上できるかについての実験について以下で説明する。

図 7 では TV で相撲を見るユーザと、提案する 4 つの対話機能によって発話する TV 雑談ロボットが雑談した際の例を示す。タイムスタンプはコメント No.1

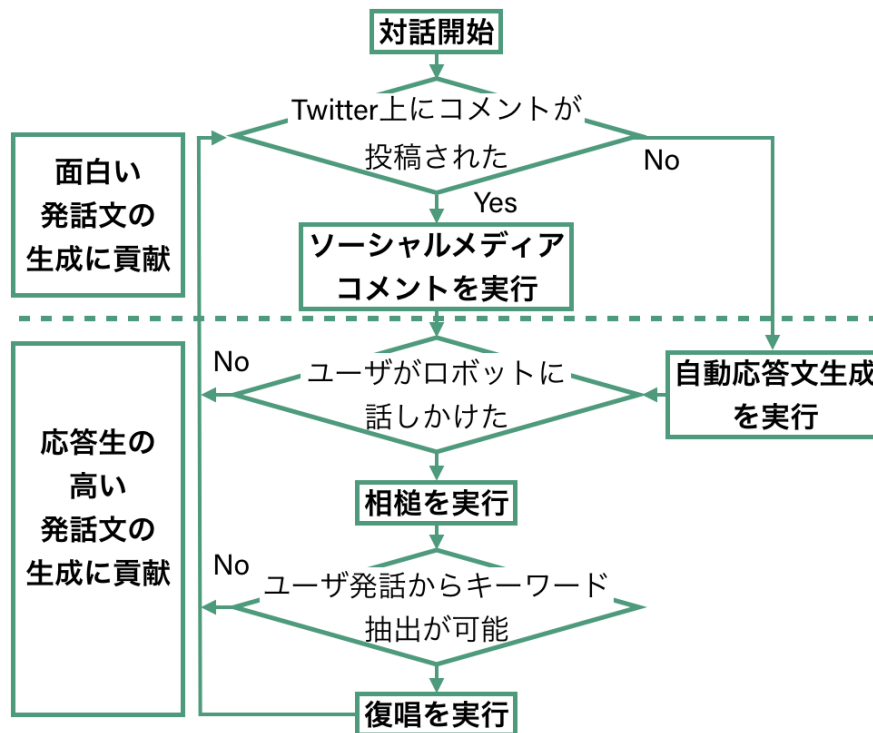


図 5 4つの対話機能を適用したTV雑談ロボットの対話フロー

のコメントのタイムスタンプを“0:00”とし、相対的な時間をそれぞれのコメントに付けたものである。この対話を行った時、ロボットはtwitterから「今日は美人が多いなあ」というSNSコメントを取得し、発話文として用いた。この発話文に対してユーザは、ロボットが相槌の取り組みではなく観客の方を観ていた事に意外性を感じ「観客の方観てるのかよ」とツッコミを入れた。このユーザの発話に対し、ロボットはユーザの発話が終了した時に相槌機能による「はい」というコメントを返す。続いて復唱機能を用いて「一般観客か」というユーザの発話文に含まれている「観客」というキーワードを用いた発話文をユーザに返している。この様に、ロボットはSNSコメントによりユーザの興味を引き発話意欲を向上させ、その後も自動応答文生成を用いる事で応答を速くする事に成功している。同様に、コメント No.5 のSNSコメント「ひいっ、でけえ!」というコメントに対し、ユーザは同意を覚え「たしかにデカイよねえ」と返答した。この後、先

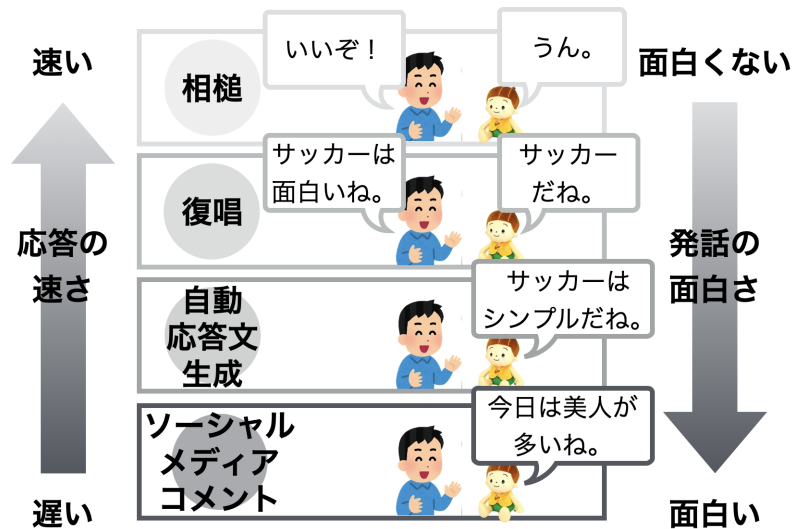


図 6 4つの対話機能の位置付け

ほどと同様にロボットはユーザの発話が終了した時に相槌機能による「うん」というコメントを返す。しかし、先ほどとは違いこの後には復唱機能を用いた対話文は生成されなかった。これはユーザの発話文にキーワードとなる語句が存在しなかった事が原因であると考えられる。この後、通常ならばSNSコメントを取得し、ユーザにロボットが話しかけるのだが、取得できるSNSコメントが存在していなかった。ここで、ロボットは自動応答文生成機能を用いて「スポーツはわかりやすいですね」という発話文を生成し、ユーザに話しかけた。この時ユーザは、今までスポーツの一種である相撲を観ていた のでこの対話文に不自然さを感じる事無く、そのまま対話を継続させ「相撲はわかりやすいね」と発話した。この様に、ロボットはSNSコメントを取得できない場合でも自動応答文生成機能を用いて、ユーザとの対話を継続させる事が可能である。

3.1.2 雑談システムの評価

本項ではTV雑談ロボットが提案された対話機能を適用して際のユーザの印象を評価する。評価対象として4つの対話機能を適用した場合と、SNSコメントのみを行う2条件のロボットを比較する。これらの条件間でユーザがロボットに対

TV番組（相撲）視聴時の状況	対話内容	コメント番号	タイムスタンプ
 <p>▲ TV雑談ロボットが美人な観客に着目し発話した場面。ユーザは相撲の取り組みではなく、観客に注目していたロボットに驚きコメントした。</p>	ソーシャルメディアコメント 今日は美人が多いなあ	1	0:00
	ユーザ 観客の方観てるのかよ	2	0:02
	相槌 うん	3	0:06
	復唱 一般観客か	4	0:09
	ソーシャルメディアコメント ひいっ、でけえ！	5	0:12
	ユーザ たしかにデカイよね	6	0:14
	相槌 うん	7	0:17
<p>Twitter上にコメントが投稿されなかったため、自動応答文生成が実行され、TV雑談ロボットが発話した場面。</p> <p>ユーザはロボットに対して返答する形でコメントした。</p>	自動応答文生成 スポーツは わかりやすいですね	8	0:22
	ユーザ 相撲はわかりやすいね	9	0:25

図 7 提案する 4 つの対話機能を適用した TV 雑談ロボットとの対話例

して、ユーザが思わず発話をしてしまう度合いとして発話頻度の差と、ロボットに対して感じる自然さ、および継続利用意欲の差を調査する。4 人の実験参加者は、テレビ番組を見ながら 2 条件の TV 雑談ロボットと自由に話すように指示された。ここで、先行既存から、SNS コメントを発話する TV 雑談ロボットが継続利用意欲の向上に対して一定の効果を持つことが計 6 名の被験者による実験にて示された [20]。また、本手法で実装した相槌・復唱・自動応答文生成といった機械応答も要素ごとに先行研究にて検討された機能であり [56, 57, 41, 42]、一定の効果を持つことが事前に示されていた。よって、本研究において 4 名の被験者で検証を行うことは妥当であると判断した。

発話頻度には、1分あたりの参加者の平均発話数を用いる。ここではロボットからの働きかけによる発話頻度の変化から、TV雑談ロボットの魅力を調査するために、ロボットの発話によって引き起こされたユーザーの発話のみがカウントされた。したがって、参加者がテレビ番組でのコメンテーターの発話への応答など、ロボットの発話に起因しないものを自発的に発声した場合、参加者の発話はカウントされない。ロボットの継続利用意欲を評価するための指標としてアンケートを用いる。各テレビ視聴セッションの後、次の2つの質問に回答するよう参加者に導入が行われた。質問1:「ロボットの対話は自然でしたか?」、質問2:「ロボットと再び話したいですか?」

質問1はロボットの発話タイミングの自然さを評価するために、質問2はロボットがユーザの継続利用意欲を評価するために用いた。各質問は5段階のリッカート尺度で評価される(1:全くそう思わない, 5:とてもそう思う)。実験参加者は事前に録画された15分間の動画を、各条件のロボット共に視聴する。この実験で用いるTV番組が実際に放送されている間に取得したSNSコメントが事前に収集され、これを用いてロボットはSNSコメントを行う。ここでSNSコメントが利用できない時間帯に、自動応答文生成機能を代替機能として用いる事は可能なかを検証するため、4つの対話機能を用いて対話をおこなう条件のロボットは動画視聴の開始後1分間はSNSコメントを話し、次の1分間は自動応答文生成を、その次にSNSコメント...、というように、1分ごとにSNSコメントと自動応答文生成機能を入れ替えた。適用する条件の順序による影響をなくすために、各条件を適用する順番にはカウンターバランスをとった。実験時の様子を図8に、実験時に用いたソフトウェアおよびツールについて表1に示す。実験中はビデオカメラで実験風景を撮影するが、撮影以外は被験者がなるべく自然体でロボットとテレビを視聴できる環境を用意した。実験室内にはテレビ番組を視聴するためのテレビと、机上に被験者の対話相手となる対話ロボットとして、うなずきかぼちゃん[58]を配置した。図9に示すようにロボットの機構はRaspberry Pi Model B+(CPU:ARM1176JZF-S single-core メモリ:512MB)によって制御され、他のソフトウェアはラップトップPCで実行される。ロボットは発話を行う際、腕が上下に振れるように両肩部分にサーボモーターが内蔵された。この動作はロボットの



図 8 実験の外観図

発話意図を事前に参加者に知らせることを目的に実装された。音声認識で 사용되는音響モデルと言語モデルは、Julius パッケージに含まれている標準的な話者に依存しないモデルを採用した。

3.1.3 評価結果

図 10 および表 2 は各条件においてロボットによって誘発された実験参加者の発話数を示す。アスタリスクは対応のある t 検定における有意差を示す (**, $p < 0.001$.)。発話数は 4 つの対話機能を組み合わせた条件において、SNS コメントのみを用いた場合と比較して有意に高い値となった ($t(28) = 1.9 \times 10^{-4}$, $p < 0.001$.)。SNS コメントのみを用いて発話するロボットがユーザーの発言を促すことがこれまでの研究で明らかとなっていたが [20], 今回の結果は SNS コメント以外の複数の対話機能を組み合わせることでユーザの発話がさらに引き出されることを示した。

図 11 は主観的指標として用いた 2 つの質問に対するスコアを, 表 2 は数値的な結果を示す。質問 1: 「ロボットの対話は自然でしたか?」において, 4 つの対話機

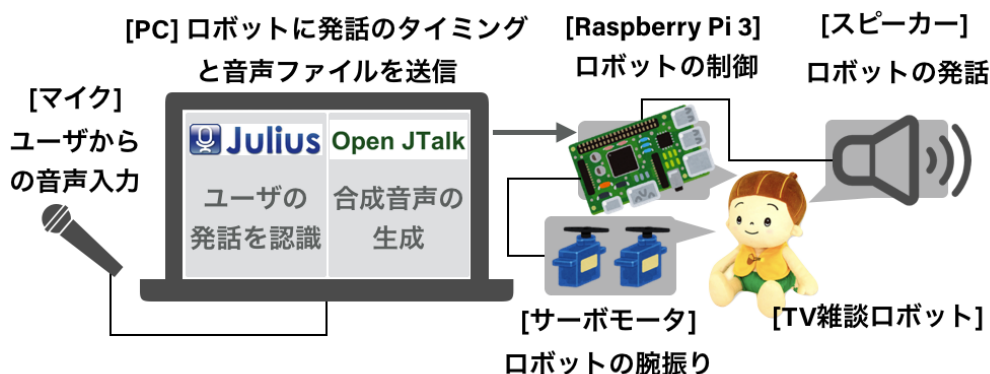


図 9 実験で使用された TV 雑談ロボットの制御システム

能を組み合わせた条件が有意に高いスコアを示す結果となった ($t(6) = 0.4 \times 10^{-4}$, $p < 0.001$). 質問2:「ロボットと再び話したいですか?」においても同様に, 4つの対話機能を組み合わせた条件が SNS コメントのみの条件と比較して有意に高いスコアを示す結果となった ($t(6) = 0.001$, $p < 0.05$).

被験者からの自由記述から, SNS コメントのみの条件で発話を行うロボットに対する意見を以下に示す。「話す内容は面白いが応答がない」「騒音を感じる」「ロボットが一方向的に話すだけで不自然に感じた」。一方, 4つの対話機能すべてが使用されている条件に対しては,「きちんと応答してくれているので,話していたいと思う」,「反応してくれるので,もう少しはなそうかなという気持ちになった」といった意見が得られた。また, 4人の実験参加者の内, 2人はロボットが複数の対話機能を用いて被験者に話しかけていたことに気付かず, 残りの2人は気付いていた。複合された対話機能によるロボットの応答に気づいた参加者からは「いきなり番組の内容と関係のない見当違いな話題を話してきたため, 複数の対話機能に気づいた」という意見が得られた。

3.1.4 考察

質問2の結果と, 4つの対話機能が適用された条件に対する実験参加者からのフィードバックから, ユーザの発話に関連した応答を行うことで, ロボットへの

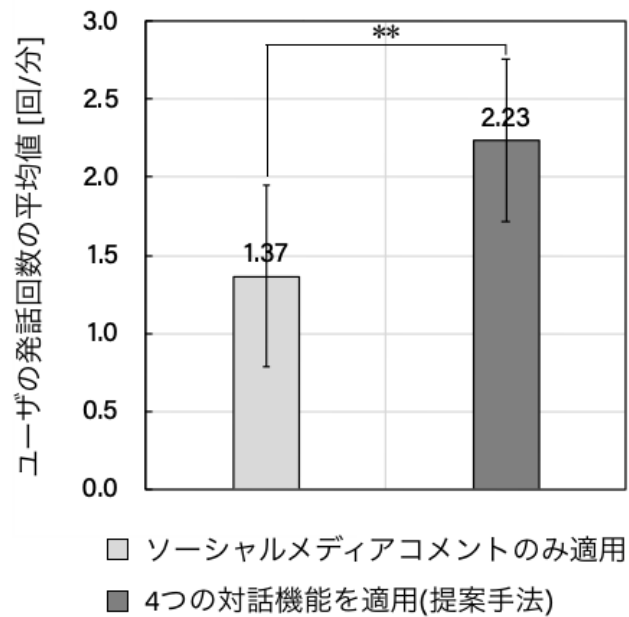


図 10 ユーザ発話数の平均値と分散

継続利用意欲が高まることが示された。ただし、ロボットがテレビ番組の内容とは無関係のトピックについて発話したというコメントから、4つの対話機能の内いずれかが対話破綻 [16] につながる可能性があることが示された。提案する4つの対話機能を用いたとしても、この対話破綻を防ぐことは難しく、ユーザーの継続利用意欲を低下させる可能性が残ることが考えられる [60]。したがって生成された発話文がTV番組の内容と確実に一致することが今後の課題として挙げられる。

また、本実験は事前に録画されたTV番組の映像と、番組放映時に収集されたSNSコメントが使用されており、リアルタイムで動くシステムとしての評価は行っていない。リアルタイムで動作するシステムの構築を検討する場合、この実験の結果から以下の2つの問題が発生すると考えられる。1つは、対話機能のSNSコメントを用いる上で発生する遅延の問題である。SNSユーザーがソーシャルメディアに書き込んだコメントに基づいてロボットの合成音声を生成するのにかかる時間に応じて遅延が発生する。この遅延のため、テレビ番組の内容が、ロボットのSNSコメントの発話が利用可能になるまでに別のトピックに移行する可能性

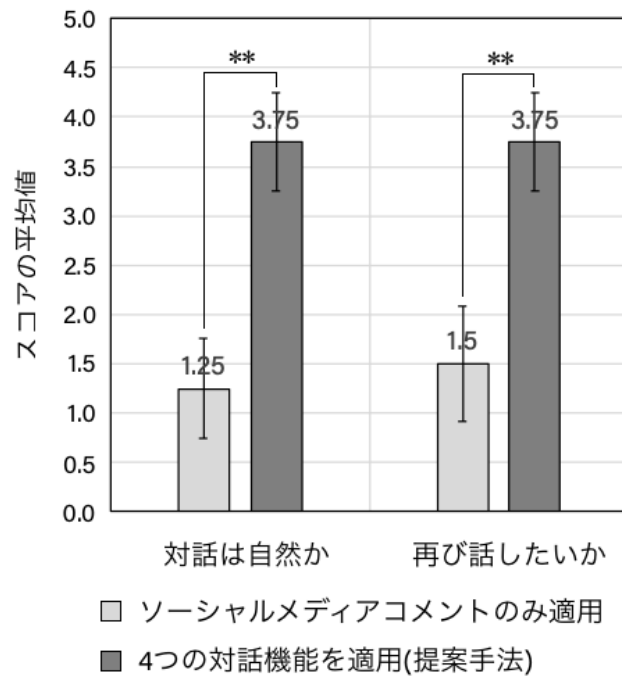


図 11 質問における 5 段階評価の平均値と分散

があり、ロボットが発声する内容との間に不一致が発生すると考えられる。したがって、SNS コメント機能でも対話破綻が発生する可能性がある。

リアルタイムで動作するシステムの構築を検討する場合に考えられるもう 1 つの問題は、TV の音声とロボット音声の重複が発生することである。TV 雑談ロボットのコメントが「騒音を感じる」というユーザの意見からも、ユーザが TV 番組とロボット音声の両方の声を聞くことが困難となり、不快な印象を与えることが考えられる。本実験から得られた問題である、TV 雑談ロボットがユーザに不快な印象を与えることを回避するための発話方法の提案を続く 3.2 にて述べる。

表 1 実験の実装に用いられた各種ツール

項目	リソース
対話ロボット	うなずきかぼちゃん [58]
ロボット制御ユニット	Raspberry Pi Model B+ (Memory: 512 MB)
SNS リソース	Twitter
キーワード抽出 API	Yahoo! JAPAN, Text analyzer
雑談エンジン	NTT DOCOMO, 雑談対話
音声認識ツール	Julius 4.4.1
音声合成ツール	VoiceText[59]

表 2 実験から得られた数値的結果

		ソーシャルメディアコメントのみ	4つの対話機能
1分あたりの発話数	M	1.37	2.33
	SD	0.58	0.52
	p	< 0.001	
対話は自然か	M	1.25	3.75
	SD	0.50	0.50
	p	< 0.001	
再び話したいか	M	1.50	3.75
	SD	0.58	0.50
	p	< 0.001	

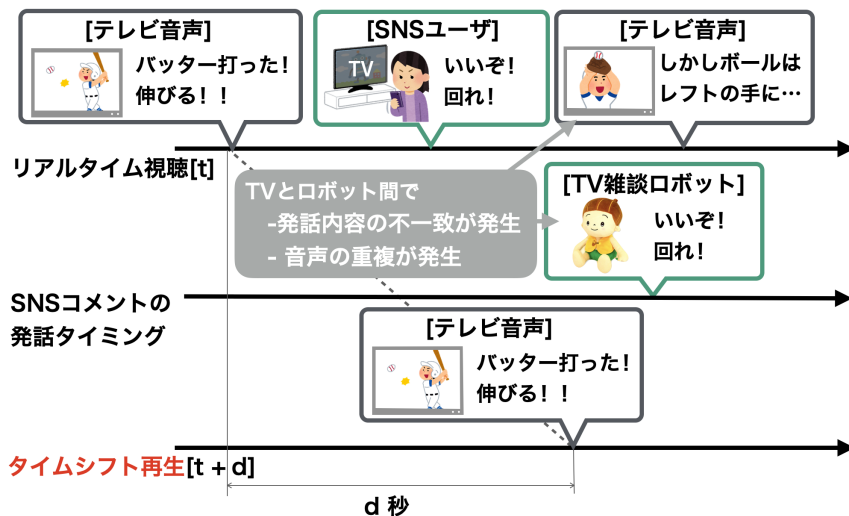


図 12 TV 雑談ロボットのリアルタイム利用により発生する 2 つの問題

3.2 トピックに同期する発話

3.2.1 TV 雑談ロボットがユーザに与え得る不快要因の分類

3.1.4にて述べたように、TV 雑談ロボットのリアルタイムでの運用を考えた際、SNS コメントの取得から TV 雑談ロボットの発話までにかかる時間によって発生する遅延の問題とテレビの音声とロボット音声の重複の 2 つの問題が挙げられた。図 12 に発生することが考えられる問題について示す。これらの問題によってロボットはユーザに不快感を与え、継続利用意欲を低下させる可能性が考えられる。

本節ではこれらの問題を解決することにより、ユーザの対話継続意欲を向上させることが可能か検証することを目的とする。この問題を解決するためには TV 雑談ロボットがテレビのトピックに同期した SNS コメントを発話でき、テレビとの音声重複を避けるよう、ロボットの発話タイミングを調整するための処理時間が必要となる。この処理時間を確保し、SNS との連携によるロボットの発話の遅延や、テレビとロボットの音声の重複を解決し、ユーザの TV 雑談ロボットとの対話継続意欲の向上を目指す。

3.2.2 タイムシフトによる遅延問題の軽減手法

ロボットの発話内容とテレビ番組のトピックの時間的同期のための処理時間を確保するために、テレビの録画機能の一つである“タイムシフト再生”を利用する。ここで、タイムシフトとは一般的なDVD/BDレコーダにも備わる機能の一つであり、テレビ番組を録画しながら再生できる技術である。12に示すように、タイムシフトを適用しない場合では、テレビ番組のシーンに対して、コメントの収集・選択の処理時間により、ロボットの発話する内容がテレビの内容より遅延する。これは、SNSコメントは絶対時間（テレビ番組放送時にテレビ番組を視聴していた人々がSNSにコメントを書き込んだ時間）を基準に生成され、ロボットの発話はこの絶対時間に基づいて行われるためである。また、テレビ番組のシーンに対するコメントがSNS上に書き込まれるまでには、テレビ番組のシーンが放送されてからユーザがテキスト入力する作業が発生するため、遅延が発生することが理由として挙げられる。そこで、テレビ番組にタイムシフトを適用し、図12に示すように実時間 $t[\text{sec}]$ から $t+d[\text{sec}]$ の遅延を発生させることでテレビ番組の内容とロボットの発話内容の時間的同期を行う。ここで時間 $d[\text{sec}]$ についてはトピックの同期がおおよその場面において見られる値を用いるものとする。

タイムシフト再生によって生じた遅延を利用してロボットの発話をトピックに同期させることにより、ユーザがロボットの発話とテレビの音声の両方を簡単に聞くことができ、ロボットの発話の意味を理解する可能性が高まることが期待される。本説ではこのタイムシフトを使ってTV雑談ロボットの発話タイミングの調整を行うことで、SNSコメントの遅延により発生することが考えられる2つの問題を同時に解決でき、ユーザのTV雑談ロボットとの継続対話意欲を促進することができるという仮説を検証する。さらに、ロボットがユーザに与える印象は、タイミングが調整されたテレビ番組の特性に依存することが考えられる。例えば、テレビ番組からの音声が必要な情報源として提供される番組（以下、音声指向型コンテンツ）では、タイムシフトによるトピックに同期した発話によりユーザは映像重視の番組（以下、映像指向型コンテンツ）に比べよりロボットに対し好印象を抱く傾向があると考えられる。以上から、タイムシフトによるロボットの発話タイミングの調整がある場合とない場合を比較して以下の仮説の検証を行なう。

H_1 タイムシフトが適用された場合、そうでない場合と比較してユーザはロボットの発話内容とテレビ番組の内容とが意味的により一致していると感じる.

H_2 タイムシフトが適用された場合、そうでない場合と比較してユーザはロボットの発話を邪魔に感じない.

H_3 タイムシフトが適用された場合、そうでない場合と比較してユーザはよりロボットに対し継続利用意欲を抱く.

H_4 映像指向型コンテンツと比較して、ユーザは音声指向のコンテンツにおいてタイムシフトによるロボットのパフォーマンスにより大きな違いを感じる.

3.2.3 タイムシフトを適用したロボットの評価

トピックに同期する発話を行う TV 雑談ロボットの有効性を検証するために、12 人の実験参加者に 3 つの異なるタイプのテレビ番組を見ながらロボットを使用するように指示した。12 人の日本人参加者（女性：2、男性：10、23-25 歳 [M = 23.69、SD = 0.72]）は、各番組を見た後に以下の 3 つのアンケートに回答する。

Q_1 ロボットの発話内容とテレビ番組の内容が同期していると感じましたか？

Q_2 ロボットの発話音声を邪魔に感じましたか？

Q_3 ロボットと継続して対話したいと思いましたか？

これらの質問によって各仮説を検証する。アンケートには 5 段階のリッカート尺度で評価してもらう（1:全くそう思わない、5:非常にそう思う）。アンケートスコアを対応ありの t 検定によって評価を行なう。

番組には事前に録画した動画を用いる。動画の長さは約 5 分で、実験参加者は事前に記録されたスポーツ（相撲）、バラエティ（コント）、ニュースの 3 種類のテレビ番組を視聴する。スポーツ番組には、力士同士が互いに競い合うダイナミックなシーンを含んでおり、映像指向型コンテンツに属するものとする。バラエティ番組は、複数のコメディアンがコントを行なうシーンが含まれる。これは映像指向および音声指向型の間接的なコンテンツとして位置付ける。ニュース番

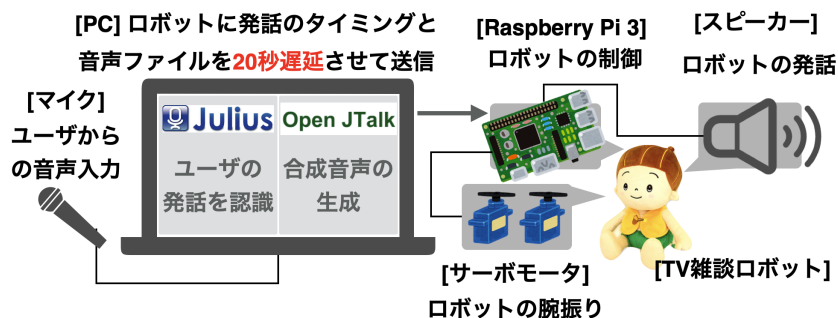


図 13 実験時の TV 雑談ロボットの実装

組は、いくつかのトピックが取り上げられ、順に紹介され流内容であり、音声指向型コンテンツに属するものとする。各番組においてロボット対話を行なった際のアンケートスコアを比較して、仮説 H_4 を検証する。

実験で用いる TV 雑談システムの概要を図 13 に示す。TV 雑談ロボットが被験者と対話を行うために、Julius を用いて発話区間検出と音声認識を行う。ロボットの外装として“うなずきかぼちゃん(ピップ RT 株式会社製)” [58] を使用する。音声認識には Julius 4.1.3 [52] を、ロボットの音声合成には HMM テキスト音声合成システムである Open JTalk[61] を用いた。SNS コメントのリソースには twitter[55] を用いる。テレビ雑談ロボットには、3 種類の対話機能（相槌、復唱、SNS コメント）を用いた。ロボットにはアクチュエータとして、両肩にサーボモータ、口元にスピーカを埋め込み、発話音声は内部スピーカから出力された時に発話音声に同期して両腕が上下するよう設計した。制御部には Raspberry Pi3 を使用し、これを介してロボットを制御する。電源部にはモバイルバッテリーを用い、制御部・電源部・駆動部がすべてロボット内部に挿入できるよう設計した。本実験ではタイムシフトが仮想的に適用される状況をシミュレートするために、ロボットが SNS コメントを発するタイミングは、テレビ番組の内容と SNS コメントの時間内の同期を実現するために 20 秒に設定された。さらに、テレビの音声とロボットの音声との重複を避けるため、ロボットの発話の調整はマニュアルで行なった。

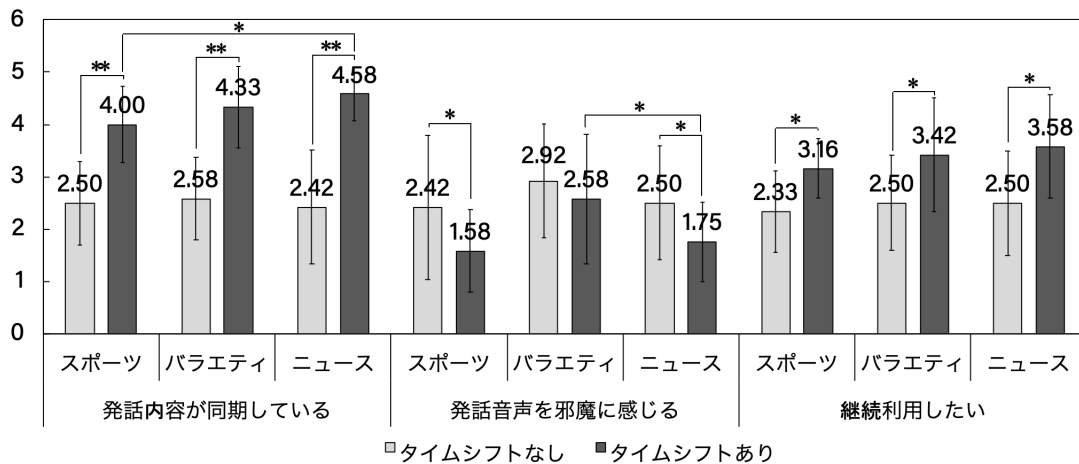


図 14 各 TV 番組視聴後のアンケートスコア

3.2.4 評価結果

図 14 にて Q_1 , Q_2 , Q_3 のそれぞれについて、タイムシフトを適用した条件と、適用しない条件におけるアンケートスコアを示す。また、表 3 にて数値的な結果をまとめる。アスタリスクは対応ありの t 検定における有意差を示す (**, $p < 0.001$; *, $p < 0.05$)。

これらの結果は番組の種類に関わらず、タイムシフトを適用した条件では、タイムシフトを適用しない条件と比較してロボットの発話がテレビの内容により一致することを示した。テレビの内容とロボットの発話内容の一致に関する Q_1 では、3 種類のテレビ番組すべての結果から、非タイムシフト条件よりもタイムシフトを適用した条件の方が平均値が有意に高いことが示された。

ロボットの発話を邪魔に感じるかを検証する Q_2 では、スポーツとニュースでタイムシフトあり・なし条件間における有意な差が観察された。ただし、バラエティ番組について、有意差は認められなかった。以上から、スポーツとニュース番組において、タイムシフトが適用されたロボットの発話は、タイムシフトが適用されない条件よりも実験参加者のテレビ視聴を妨げないことが示された。

ユーザのロボットに対する継続利用意欲に関連する Q_3 の結果から、3 種類のテレビ番組すべてについて、アンケートスコアがタイムシフトを適用しない条件よ

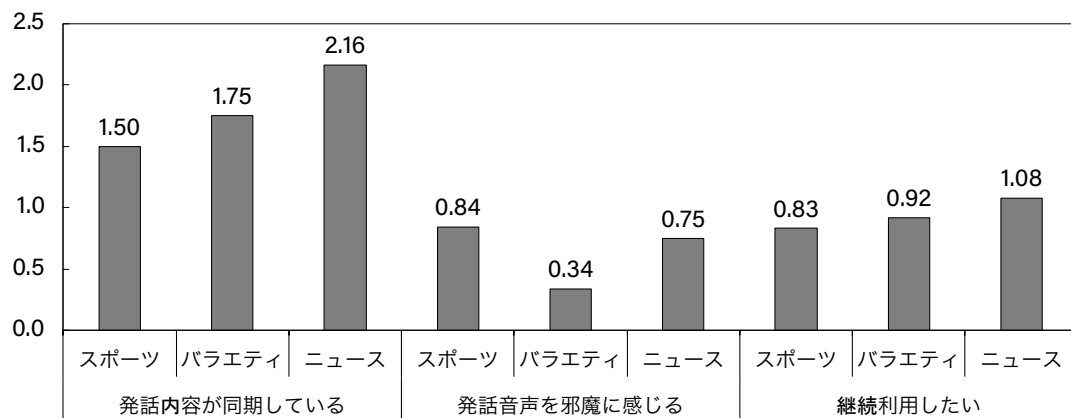


図 15 条件間のアンケートスコアの差

りもタイムシフトを適用した条件が有意に高いことが示された。したがって、タイムシフトによって発話タイミングが調整されたロボットが実験参加者のロボットの継続利用に対する意欲を向上することに貢献した。

以上をまとめると、 Q_1 および Q_3 の結果は H_1 および H_3 を支持し、スポーツおよびニュース番組において、 H_2 が一部支持される結果となった。ただし、バラエティ番組に関しては H_2 を支持しない結果となった。

図 15 は、各番組、各質問におけるタイムシフトを適用した条件とそうでない条件の平均値の差を絶対値で示す。 H_4 は TV 番組の音声指向性が高いほど、タイムシフトが適用された条件下でロボットがユーザーに与える印象が良くなるというものである。 Q_1 と Q_3 の結果は、この仮説を部分的に支持した。さらに、図 14 に示される Q_1 における結果から、タイムシフトが適用された条件下での映像指向型コンテンツと音声指向型コンテンツの間に有意差が認められた。一方 Q_2 においては、映像指向から音声指向コンテンツ間の平均値の差が他の質問と同様に順に大きくなることが想定されたが、想定と異なる結果となった。以上より、 H_4 は支持されない結果となった。

表3 実験から得られた数値的結果

	スポーツ		パラエティ		ニュース	
	タイム シフトなし	タイム シフトあり	タイム シフトなし	タイム シフトあり	タイム シフトなし	タイム シフトあり
Q ₁	M	4.00	2.58	4.33	2.42	4.58
	SD	0.74	0.79	0.78	1.08	0.52
	p	< 0.001		< 0.001		< 0.001
Q ₂	M	1.58	2.92	2.58	2.50	1.75
	SD	0.79	1.09	1.24	1.09	0.75
	p	0.034	0.474		0.021	
Q ₃	M	3.16	2.50	3.42	2.50	3.58
	SD	0.58	0.91	1.08	1.00	1.00
	p	0.005	0.034		0.005	

3.2.5 考察

結果は、想定される SNS コメントの遅延が原因で発生する問題のうち、テレビ番組の内容と TV 雑談ロボットの対話の間に意味的な不一致が実際に起こりうることを示した。これらは Q_1 におけるタイムシフトを適用しない条件の結果から、明らかである。一方、タイムシフトが適用されていない条件での Q_2 の平均値が中央値を超えないことから、テレビ音声とロボット音声の重複がユーザに与える印象は想定していたよりも発生していないことがわかった。

同時に、タイムシフトによる TV 雑談ロボットの発話タイミングの調節によって、ロボットがテレビ番組の内容と意味的に同期したコメントを発することを可能とし、ユーザのロボットに対する継続利用意欲を促進するための効果的なアプローチであることが示された。特に提案する方法には映像指向型コンテンツよりも、音声指向型コンテンツに対するユーザの不快な体験を減らすことに貢献することが示唆された。ただし、バラエティ番組の例では図 14 に示すように、音声間の重複に関して Q_2 は有意差を示さなかったことから、ロボット発話のタイミングの調整が音声重複の問題を改善する手法であるとは言えないことがわかった。このバラエティ番組においてのみ仮説と異なる結果を導いたことについて考察する。

まず、12 人の実験参加者という人数が不十分だった可能性がある。実際にテレビの音声とロボットの音声はどの程度重複するかについて調査を行なったところ、それぞれの 3 つのテレビ番組に対し 50% 以上の発話音声重複時間の減少がみられた。音声重複の回避はどのテレビ番組でも同等の値を示していたことから、音声重複の程度による差が今回の結果に影響していないと考えられる。今回、バラエティ番組は映像指向および音声指向型の両方の性質を持つ中間型コンテンツとした。バラエティ番組では、複数のコメディアンはコントを行なうものとなり音声と映像の重要性はシーンによって異なる。 Q_2 に関する参加者の意見によると、「テレビ番組が面白かったので少し静かに番組を見たいと思った」、「出演者のセリフとロボットの発話音声がかぶる場面がありセリフを聞き取れなかった」など、TV 雑談ロボットに対する否定的な意見が得られた。一方で「普段バラエティーは家族や友達と見ているので気にならなかった」、「ロボットの発話内容が面白く、楽しくテレビを観られたので邪魔には感じなかった」などの肯定的なフィー

ドバックも得られた。つまりバラエティ番組を音声指向型コンテンツと捉えた被験者はロボットの発話を邪魔に感じ、音声指向型コンテンツと捉えた被験者は邪魔に感じなかったことが考えられる。したがって、映像指向および音声指向型の両方の性質を持つ中間型コンテンツとして用いたバラエティ番組の位置付けが適切でなかった可能性がある。インタラクション中の音声重複の回避がロボットの印象向上に対し重要であるという知見 [62] から、今後は人のテレビ音声のみならずユーザの発話を検出した際に重複しないことも重要であるため、ロボットが能動的な発話を行う際は周囲の状況のセンシング結果を元に発話を行う必要がある。

4. 共感を示す発話を実現する手法

3ではTV雑談ロボットがユーザに与える自然な振る舞いを実現するための手法と、それを適用したロボットが継続利用意欲に及ぼす影響について調査した。本章ではユーザの感情や嗜好に基づき、共感を示す振る舞いを実現するための手法を提案する。2.3にて述べたように、人同士の関係において行動の同期などの共感を生む振る舞いは調和的な関係の構築やパートナーへの好感、共感的な行動高めることが明らかになっている。この知見に基づき、ユーザに共感を示す振る舞いを実装したロボットが継続利用意欲の向上に貢献するかを調査する。本章では以下の3つの手法について提案および短期的なインタラクションを通じた評価を行う。

- 嗜好を反映する発話 4.1
- 情動共有 4.2
- 韻律模倣 4.3

嗜好を反映する発話では、ユーザにとってTV雑談ロボットが面白いと感じる存在となるよう、ユーザ個人がSNSコメントに対し面白さのラベル付けを行うことで、嗜好の推定を行う手法について提案する。

情動共有では、SNSを用いて一般的な大衆の意見から推定される情動の変化に基づき、TV雑談ロボットがユーザと情動を共有するような振る舞いを適用するための手法について提案する。

韻律模倣では、非言語コミュニケーションに関する同調対話の心理学および工学的知見に基づき、対話者の韻律について同調するロボットを実現するための手法について提案する。ここで“同調”とは、対話者の非言語・パラ言語情報に同調するロボットの機能と定義する。同調対話ではSNSの分析は行わず、ユーザの発話音声を直性センシングして韻律情報を取得し、それに合わせてロボットの韻律のみを変化させる。

続く4.1, 4.2, 4.3では上記で述べた各手法について対応する先行研究、実装の詳細と評価結果、手法ごとの考察を述べる。

4.1 嗜好を反映する発話

3.1にて述べたように、TV雑談ロボットはテレビ視聴時という限られた環境下ではあるが、ユーザの継続利用意欲を向上させる可能性を示した。しかし、発話するソーシャルメディアがユーザの継続利用意欲に与える影響は考慮されておらず、新しいコメントが無作為に選ばれる。ソーシャルメディアコメントには、ユーザの継続利用意欲を向上させるコメントだけでなく、継続利用意欲に影響のないコメントや、逆に継続利用意欲を低下させるコメントも含まれる。例えば、「あー」「まあ」「えー」などの感嘆詞のみの文や、テレビ番組の出演者への誹謗中傷といったコメントである。このように様々な種類の文が混在しているソーシャルメディアコメントの中から無作為に選択して発話すると、ユーザの継続利用意欲を意図せず低下させてしまう問題が発生する。したがって本節では、よりユーザの継続利用意欲を向上させるために、継続利用意欲を基準とした発話文の選択技術について説明する。発話文の選択を行うにあたり、ユーザにとって魅力的だった発話内容を特定する必要がある。まず、TV雑談ロボットが発話した内容がユーザの継続利用意欲に与える影響を、データセットとして収集する。そして、収集した継続利用意欲に関するデータセットから、ユーザの継続利用意欲を推定して発話文を評価する手法を提案する。

4.1.1 ソーシャルメディアコメントの評価手法

ロボットの発話した内容がユーザの継続利用意欲に与える影響を収集するための要件を2つ挙げる。1つは、ロボットの発話についてユーザ自身が感じた継続利用意欲の変化を入力できるインターフェースを作成すること、もう1つは、より多くのユーザに対話してもらうために対話システムを一般公開することである。ユーザ自身が感じた継続利用意欲の変化を入力できるインターフェースとして図3に示す、スマートフォン上で動作するTVアプリケーション”Dabelive(だべらいぶ)”を利用する。Dabeliveは、iOSとAndroid上で動作するスマートフォンアプリケーションとして株式会社 amirobo tech[33]が開発し、App Store[63](iOS)とGoogle Play Store[64](Android)、およびweb上で無料で一般公開されている。

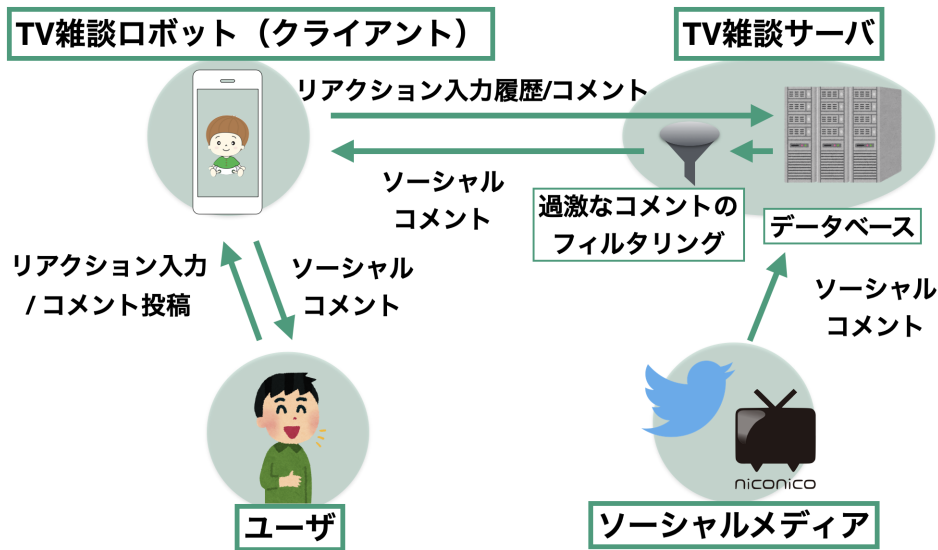


図 16 Dabelive のシステム概要

Dabelive のシステム概要を図 16 に示す。Dabelive は主に、クライアントであるスマートフォンアプリケーションと、SNS 上のコメントを収集しクライアントへ配信する TV 雑談サーバからなる。

ユーザは自身が視聴しているテレビ番組を能動的にシステムに提示する必要がある。この処理を少ない負荷で行ってもらうため、余計なログイン等の処理を省き図 17 に示すようにユーザに視聴している放送局を選択を促す。アプリの画面中に表示されるチャンネル一覧から放送局名を選択することで、ユーザが現在視聴しているテレビ番組がシステムに共有される。なお、表示される放送局名は、図 18 に示す設定画面の「お住いの地域」にて選択した地域で視聴可能な放送局名が表示されるが、ソーシャルメディアコメントは放送局単位で収集しているため、全国放送とローカル放送の区別がつかない。そのため、ローカル放送の時間帯にソーシャルメディアコメントを配信してしまうと、視聴地域によってはテレビ番組の内容と対応していないソーシャルメディアコメントがクライアントへ送信されてしまう。これに対処するため、Dabelive の TV 雑談サーバは全国放送の時間帯にのみソーシャルメディアコメントを配信する仕様にした。

放送局を選択して初めて TV 雑談サーバからソーシャルメディアコメントが一

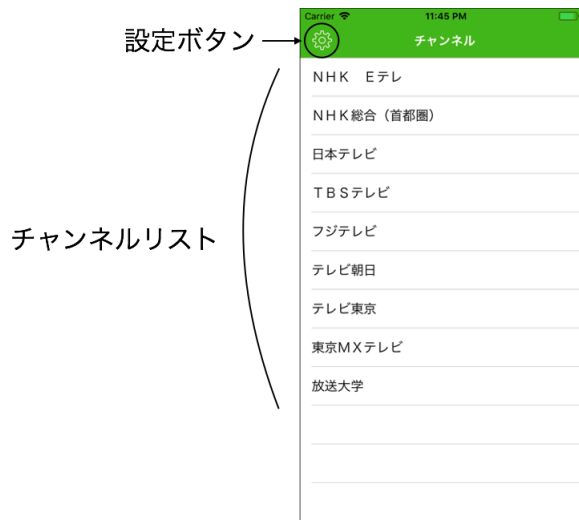


図 17 放送局の選択画面

定間隔で配布され、画面中央に表示される TV 雑談ロボットがそれを読み上げる (図 19)。TV 雑談ロボットが発話した内容は画面上部の吹き出し内に表示される。

図 19 に示すように、スマートフォン上にある TV 雑談ロボットの対話画面には、ロボットが発話したソーシャルメディアコメントの内容に対し、ユーザの評価を反映できるインターフェースを用意している。ユーザは、スマートフォン上のロボットの発言に同意や共感といったポジティブな印象を感じたときには「いいね」、否定や反論の発言を持つときは「うーん」の評価ボタンを押す。「いいね」の評価ボタンを押した際、TV 雑談ロボットは「いいね」、「うんうん」、「だよね」のうちいずれかランダムに発話する。「うーん」の評価ボタンを押した際は「うーん」、「そうかな?」、「えー」、「そっかー」のうちいずれかを発話する。

ソーシャルメディアコメントには他者を誹謗中傷する内容のものや、公序良俗に反する語句が含まれている過激なコメントが存在する。このような過激なコメントがユーザに不快感を与える可能性がある。よって、Dabelive はロボットが過激なコメントを発話しないようにするソーシャルメディアコメントのフィルタリング機能を持つ。フィルタリング機能は、図 18 の設定画面から有効/無効を切り替えることができる。ソーシャルメディアコメントはニコニコ実況 [65] と Twitter [55]



図 18 設定画面

から最新のコメントをリアルタイムで収集される。1つのソーシャルメディアコメントに対して表4に示す5つの属性情報をデータベースに記録する。

クライアントはTV雑談サーバへユーザが選択した放送局のソーシャルメディアコメント要求を送信する。TVはクライアントからソーシャルメディアコメント要求を受信すると、クライアントへソーシャルメディアコメントを一定間隔で送信する。送信するソーシャルメディアコメントは送信時点で最新のコメントを選択する。ソーシャルメディアコメント要求時に過激なコメントのフィルタリング要請が含まれていた場合、TV雑談の禁句データベース内に記録されている禁句を含むソーシャルメディアコメントは送信しない。

クライアントでは受信したソーシャルメディアコメントのテキストを元に音声

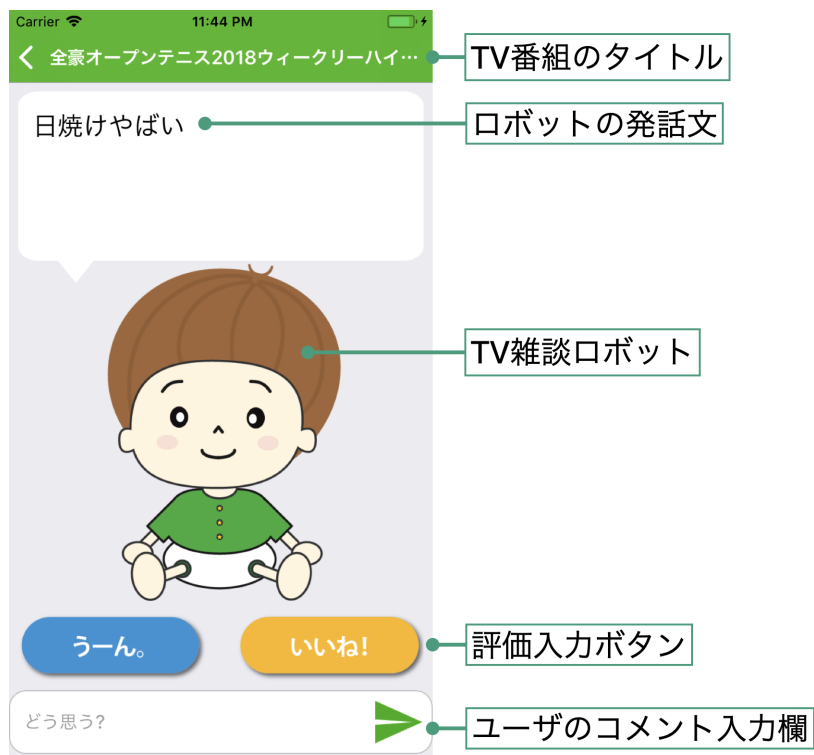


図 19 スマートフォン上の TV 雑談ロボットとの対話画面

合成処理を施して発話する。使用する音声合成 API は表 5 に示すスマートフォン OS 標準のものとする。

Dabelive ではロボットの発言に対して、ユーザは評価ボタンとコメント入力によりリアクションを行うことができる。これらのインターフェースによって入力された情報を、ソーシャルメディアコメントに対するユーザの継続利用意欲評価データとする。

4.1.2 ソーシャルメディアコメントに対する評価データ収集実験

本項ではユーザの継続利用意欲を基準としたロボットの発話文を自動的に評価するために、発話文がユーザに与える継続利用意欲の影響のモデル化を行う。このモデルをユーザ継続利用意欲モデルと定義し、以下の特徴量を用いた重回帰分

表 4 ソーシャルメディアコメントの属性情報

属性名	説明
station	コメントを書き込んだソーシャルメディアのユーザが視聴していた放送局
timestamp	ソーシャルメディアコメントが書き込まれた時刻
user	ソーシャルメディアコメントを書き込んだユーザを特定する文字列
comment	ソーシャルメディアコメントの本文
source	取得元のソーシャルメディア名

表 5 使用する音声合成 API

OS	API
iOS	AVSpeechSynthesizer[66]
Android	TextToSpeech[67]

析で解く。

ユーザ継続利用意欲モデルの作成フローを図 20 に示す。図 21 のように，作成したユーザ継続利用意欲モデルに，新規の発話文の特徴量を入力することで，入力した発話文の継続利用意欲評価値が推定・出力される。

モデル作成には，以下の特徴量を用いる。

- 説明変数 1: 発話文
- 説明変数 2: 発話文を発話したときのテレビ番組のトピック
- 説明変数 3: ユーザ属性
- 目的変数: 発話文の継続利用意欲評価値

各特徴量の詳細について説明する。発話文はテレビ番組のトピックに関連した内容のため，似たテレビ番組のトピックのときには似た発話文が存在しやすいと予想される。テレビ番組の内容に深く関連している発話文が多いため，発話文単体では意味の分からない内容が多い。したがって，発話文とテレビ番組トピック

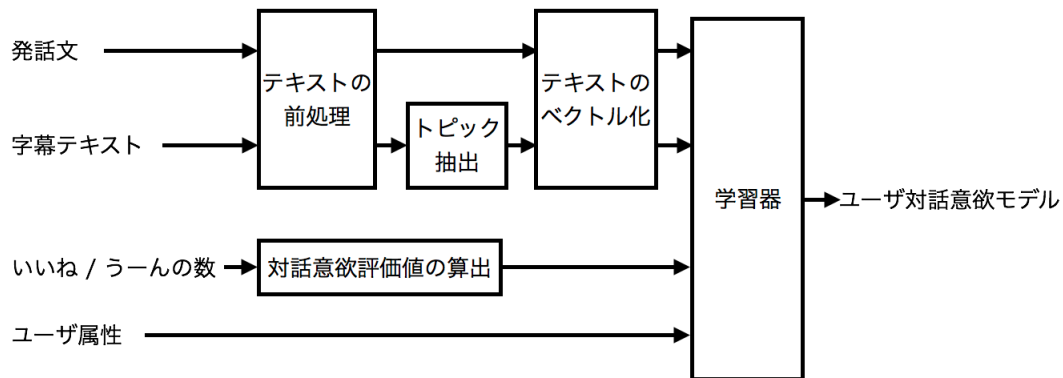


図 20 ユーザ継続利用意欲モデルの作成フロー

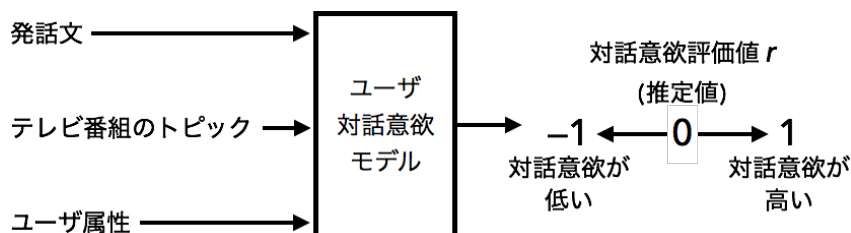


図 21 ユーザ継続利用意欲モデルによる継続利用意欲の推定

には、共起関係が成り立つと仮定し、発話があったときに放送していたテレビ番組のトピックと、発話文のペアを一つの特徴としてモデルに入力する。なお、テレビ番組の内容はテレビ番組の字幕テキストから後述するトピック抽出処理を経て抽出する。

ユーザ属性については、発話文が与えるユーザの継続利用意欲への影響はユーザの属性によって異なると考えられる。例えば、野球はユーザによって応援するチームが異なる。そのため、ユーザが応援しているチームを肯定するようなコメントは継続利用意欲が高まり、逆にユーザが応援するチームを否定するコメントは継続利用意欲が低くなる可能性がある。

最後に発話文の継続利用意欲評価値について、その定義から説明する。発話文の継続利用意欲評価値とは、発話文がロボットに発話されたときにユーザの継続利用意欲へ与える影響を定量化したものである。継続利用意欲評価値は、「継続利

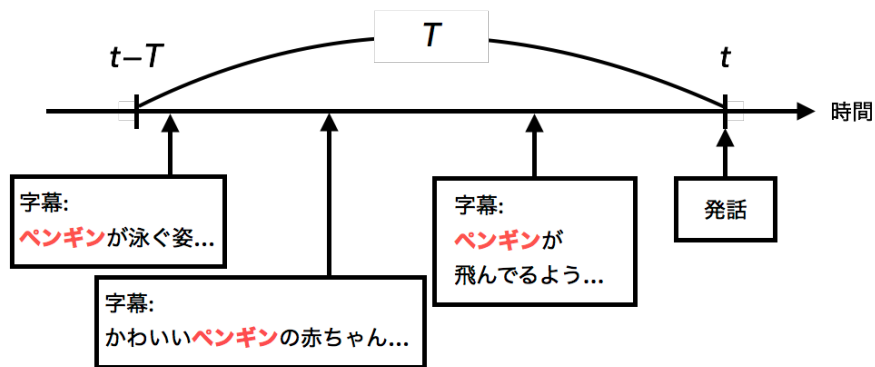
用意欲が向上した (good 評価) / 低下した (bad 評価)」の 2 値によって算出する。ある発話文の good 評価数を g , bad 評価数を b , 発話文の配信数を c とした時の、発話文の継続利用意欲評価値 r は式 1 と定義する。継続利用意欲評価値 r は 1 に近づくほど、より継続利用意欲を向上させる発話文である。一方で r が -1 に近づくほど、より継続利用意欲を低下させる発話文であることを示す。この継続利用意欲評価値 r は発話文が持つ継続利用意欲への影響力を示すため、継続利用意欲評価値 r をユーザ継続利用意欲モデルの目的変数とする。

$$r = \frac{g - b}{c} \quad (-1 \leq r \leq 1) \quad (1)$$

図 20 に示したモデル作成フローについて、処理の詳細を説明する。テキストの前処理として、発話文や字幕テキストに含まれる語は、不要なスペースや、半角/全角が揃っていない同じ単語、記号等が含まれているため、前処理ではテキストに含まれるカタカナ文字はすべて全角、アルファベットや数字は半角に変換する。アルファベットはすべて小文字で統一し、スペースと記号はすべて削除する。

本研究で発話文として用いているソーシャルメディアコメント特有の特徴として、笑いの度合いを表すネット用語の「w」が文章中に含まれることが多く、「w」の連結数が発話文毎に異なる。表記の揺れを統一するために、本手法では 2 文字以上の「w」の連続をすべて 1 文字の「w」に変換する。この「w」変換の副作用として、「www (World Wide Web)」を表現できなくなる。しかし、ソーシャルメディアコメントにおいて「www (World Wide Web)」の使用頻度は、「w (笑いの度合い)」に比べてごくわずかであるため、より使用頻度の高い「w (笑いの度合い)」を優先する。

次にテレビ番組のトピックを抽出する処理について説明する。テレビ番組の字幕テキストから、ある時刻に放送されていたテレビ番組の内容が取得できる。一つの字幕テキストは 5~10 秒ほどの、ごく局所的な内容しか表現しておらず、人間の目からもテレビ番組のトピックを推測し難いものである。そこで、本手法では有安ら [68] のトピック抽出手法を参考に、分単位の粒度でテレビ番組のトピックを取得し、発話文と関係付ける。トピック抽出の模式図を図 22 に示す。ある発話文に対応するトピックは、発話文が発話された時刻 t から $t - T$ までの期間 T の字幕テキスト群に含まれる名詞から決定する。ここで、テキスト群に含まれる



名詞	出現頻度
ペンギン	3
赤ちゃん	1
姿	1
⋮	⋮

上位 N 個を
トピックを表す
単語群とする

図 22 トピック抽出の模式図

出現頻度の高い名詞が、そのテキスト群のトピックであると仮定する。したがって、期間 T の字幕テキスト群に含まれる名詞のうち、出現頻度が高いものから N 個をテレビ番組のトピックとする。

以上で述べた発話文の評価法を用いて、テレビ視聴型雑談ロボットの実際の発話文に適用したときの、継続利用意欲評価値の推定精度を明らかにするための実験を行う。

実験の方法について説明する。ユーザ継続利用意欲モデルを作成するため、Dabelive を用いて発話文に対するユーザの継続利用意欲評価を収集し、ユーザ継続利用意欲モデルを作成するためのデータセットとする。これにより、テレビ番組の内容に関する発話文と、各発話文に対してユーザの継続利用意欲評価 (good/bad 評価) を収集する。次に、収集したデータセットを用いてユーザの継続利用意欲モデルを作成する。本実験では図 23 に示すように、ユーザ継続利用意欲モデルに用いるユーザ属性は省略する。

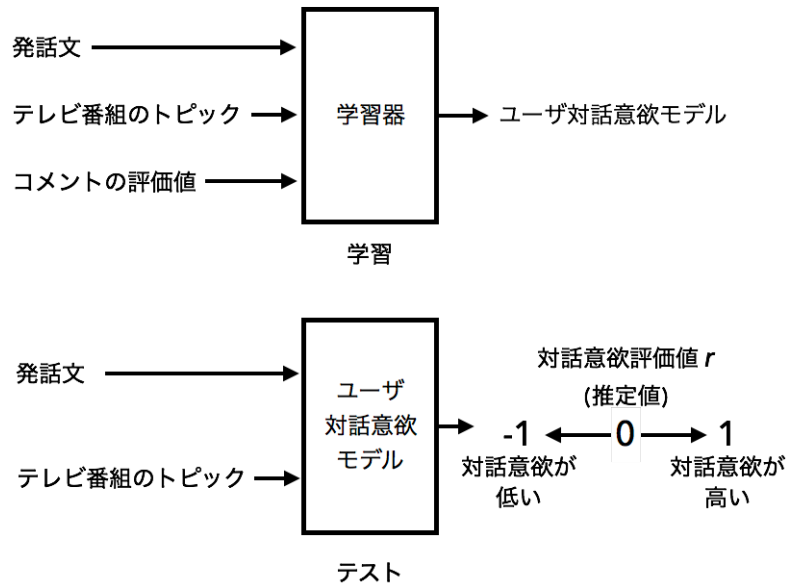


図 23 実験で使ったユーザ継続利用意欲モデル

ユーザ継続利用意欲モデルの作成に用いるパラメータについては、トピック抽出の期間を $T = 5[\text{min}]$ 、トピックに含める名詞は出現頻度の上位 $N = 10$ とした。日本語の形態素解析には、形態素解析エンジン MeCab[69] を用い、MeCab用のシステム辞書として Web の単語に強い mecab-ipadic-NEologd[70] を使用する。モデルが出力した推定結果は K-fold 法を用いたクロスバリデーションで評価する。本実験では $k = 5$ とする。

次に、発話文を収集するための TV 雑談ロボットの利用シーンとなるテレビ番組の選択について説明する。テレビ番組は相撲とニュースの 2 種類のジャンルを選択し、それぞれのジャンルでデータセットを構築する。発話文の内容はテレビのトピックに強く依存するため、まずは似たトピックの番組として日毎に番組のトピックが大きく変化しない大相撲を相撲の番組として選択した。相撲のデータセットは、ユーザ継続利用意欲モデルの分析に用いる。一方、放送されているテレビ番組の多くは、日々トピックが移り変わるものも多い。そのため、もう一つのデータセットとして、実際に想定される TV 雑談環境に近いニュース番組を選択した。ニュースのトピックは時事に影響されるため、日々トピックが移り変わ

る番組である。また、相撲とニュースの2種類は、テレビ番組の内容を理解する情報源の違いがユーザの継続利用意欲に影響を与えると仮定し設定した。彼らの実験では、相撲は視覚から得られる情報で内容を理解するテレビ番組、ニュースは主に音声から得られる情報で内容を理解する番組として選択していた。

次に、発話文に対するユーザの継続利用意欲評価の収集方法について説明する。ユーザの継続利用意欲評価を収集するために、被験者に対して Dabelive 内の TV 雑談ロボットと共にテレビ番組を視聴してもらった。テレビ視聴時、ロボットは放送しているテレビ番組の内容に関するコメントを一定間隔で発話する。被験者はロボットの発話を受けて、自身の継続利用意欲の変化を評価ボタンで評価する。評価の仕方については被験者に、「キャラクターともっと話したい場合は「いいね」を、キャラクターと話したくない場合は「うーん」を押してください」と説明した。「いいね」を good 評価、「うーん」を bad 評価として扱う。

発話文が継続利用意欲に与える影響には個人差があると考えられ、被験者間で評価ボタンを押す基準も異なると予想される。よって、図 24 に示すように、Dabelive に評価ボタンを押した理由を表す理由ボタンを配置した。理由ボタンの選定は Russell の円環モデル [71] 上に図 25 のように配置されるように選択した。Russell の円環モデルとは、人間の感情を 2 次元のグラフで表したものである。縦軸が脳の活性度、横軸が快・不快を示す。

被験者には Dabelive に初めて触れる人が含まれるため、Dabelive に慣れさせる必要がある。よって、視聴を開始してから最初の 1 時間は練習時間としてデータセットには含めないものとする。収集したデータセットの詳細を表 6 に示す。表 6 の発話文数は、練習時間に発話されたものを差し引いた数である。

収集したデータセットを用いて Bag of words モデルの辞書を作成する際、出現頻度が極端に少ない単語・極端に多い単語は文章を表す単語として適さない。よって、本実験においては、相撲のデータセットでは出現頻度が 3 回以下の単語、全体の 3 割以上の単語は除外、ニュースのデータセットでは出現頻度が 10 回以下の単語、全体の 3 割以上の単語は除外して辞書を作成した。数値は表 6 に示す 2 種類のデータセットのサンプル数を考慮し、経験的に決定した。

収集したデータセットを用いて Bag of words モデルの辞書を作成する際、出



図 24 実験用に改良した Dabelive の対話画面

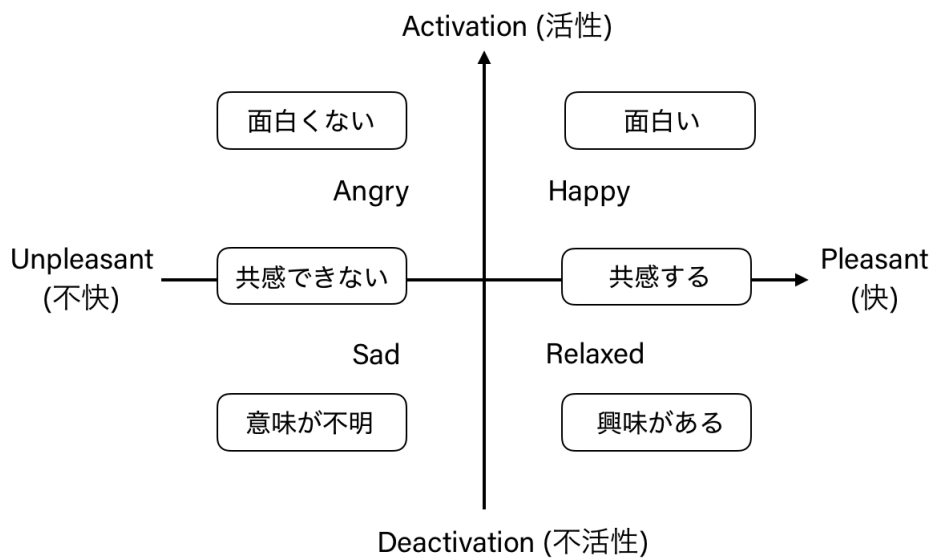


図 25 Russell の円環モデルにおける理由ボタンの配置

表 6 収集したユーザの継続利用意欲評価が付加された発話文

	相撲	ニュース
テレビ番組	大相撲十一月場所	ミヤネ屋 報道ステーション NEWS ZERO 新・情報 7days ニュースキャスター サンデーモーニング 真相報道バンキシャ！ Mr. サンデー
期間	2017/11/12 ~ 11/14 2017/11/16 ~ 11/17	2018/01/08 ~ 01/14 2018/01/22 ~ 01/28
平均視聴時間	10 時間/人	13 時間/人
被験者	20~40 代の男女 9 人	10~60 代の男女 38 人
発話文数	2,683 件	13,846 件
発話回数	57,810 回	143,533 回
評価件数	23,064 件	51,749 件

表 7 視聴後アンケートの項目

項目	回答方式
年齢	数値
性別	選択
番組自体に興味がありましたか？	1~3 の 3 段階から選択
どのような基準でコメントを評価しましたか？	自由記述

現頻度が極端に少ない単語・極端に多い単語は文章を表す単語として適さない。よって、本実験においては、相撲のデータセットでは出現頻度が 3 回以下の単語、全体の 3 割以上の単語は除外、ニュースのデータセットでは出現頻度が 10 回以下の単語、全体の 3 割以上の単語は除外して辞書を作成した。数値は表 6 に示す 2 種類のデータセットのサンプル数を考慮し、経験的に決定した。

テレビ番組視聴後には番組ごとに視聴後アンケートを実施した。アンケートでは、テレビ番組への興味の度合いと、具体的な評価した理由の収集を目的とする。アンケート項目を表 7 に示す。

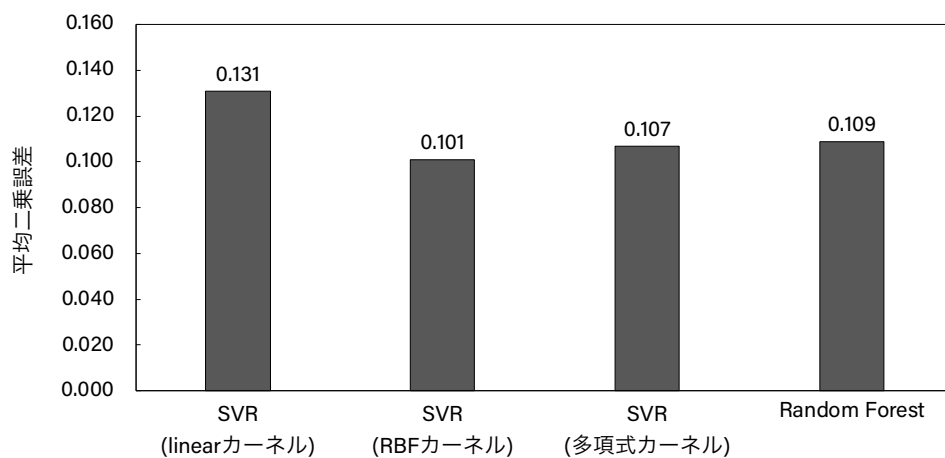


図 26 推定した継続利用意欲評価値の平均二乗誤差 (相撲), 1182 次元

4.1.3 相撲における解析結果と考察

テレビ番組が相撲のときの発話文をユーザ継続利用意欲モデルに入力して、継続利用意欲評価値を推定する。解析手法は、サポートベクタ回帰 (SVR) と Random Forest を採用した。SVR のカーネルは linear・RBF・多項式を用いた。Random Forest の決定木の本数は 1000 とした。 $k = 5$ の k -fold クロスバリデーションにて算出した 5 つの平均二乗誤差の平均を図 26 に示す。図 26 のグラフの縦軸は平均二乗誤差を示しており、この値が小さいほど誤差が少なく推定した継続利用意欲評価値が正確であることを表す。グラフは、学習データに含まれない未知のデータをテストデータとして入力し、継続利用意欲評価値を推定した結果である。次元数は 1182 次元である。境界値の処理として、推定値 r が $r < -1$ の場合は $r = -1$ 、 $1 < r$ の場合は $r = 1$ に丸める。また、入力するデータは正規化済みである。

図 26 より、テスト時に最も推定値の平均二乗誤差が少ないものは、非線形回帰である SVR の RBF カーネルであるとわかる。一方、同じく非線形回帰である Random Forest の誤差は線形回帰である SVR の多項式カーネルより大きくなっており、多項式カーネルのほうが推定精度が高い結果となっている。しかし、図 26 の結果はハイパーパラメータのチューニングをしていないため、本実験での Random Forrest の決定木本数ではモデルのバリエーションを十分に低下できていな

表 8 各興味度合いでのデータセット詳細 (相撲)

	興味なし	どちらでもない	興味あり
1人当たりの総視聴時間	10時間	10時間	9時間
被験者数	5人	7人	3人
発話文数	2683件	2683件	2337件

い可能性がある。SVRにも同様のことがいえ、 C, ϵ, γ をグリッドサーチにて調整すれば、ある程度の推定精度の向上が予測される。

ユーザのテレビ番組への興味度合いは継続利用意欲に影響を与える可能性があるため、視聴後アンケートで回答された3段階の興味度合いでデータセットを分割し、興味度合い毎にユーザ継続利用意欲モデルで継続利用意欲評価値を推定した。各興味度合いの詳細を表8に示す。テレビ番組への興味度合いは1時間毎に収集しているため、同一被験者が異なる興味度合いを回答している場合がある。図27に興味度合い毎の平均二乗誤差を示す。すべての解析手法において、図27に示す興味度合い毎の平均二乗誤差は全体のものより大きく、平均で0.119増加している。特に、興味ありの平均二乗誤差が全体の2倍以上になっており、推定精度が低下した。この推定精度低下の原因は、表8に示すように、興味度合い毎のグループ分けにより、1グループあたりの被験者数が少なくなったためだと考えられる。最も被験者数が少ない興味ありグループでは、継続利用意欲評価値の計算式に当てはめると、1人の評価が異なることで値は0.3変化する。そのため、1サンプルの正解値の変化が大きくなり、推定の難易度が高まったと考えられる。

本実験で作成したユーザ継続利用意欲モデルが出力した継続利用意欲評価値は、図28のように分布している。図28は相撲のデータセットにおける継続利用意欲評価値の正解値と、SVR(RBFカーネル)を用いたときの推定値のヒストグラムである。縦軸が頻度、横軸が継続利用意欲評価値となっており、横軸の階級は0.1区切りである。ヒストグラムには、図26に示したk-foldクロスバリデーション($k=5$)の結果から、全5試行のテスト時データを用いた。

図28からわかるように、ユーザ継続利用意欲モデルで推定した継続利用意欲評価値は、0付近のものが多く、境界値(-1や1)付近のものは非常に少ない。そ

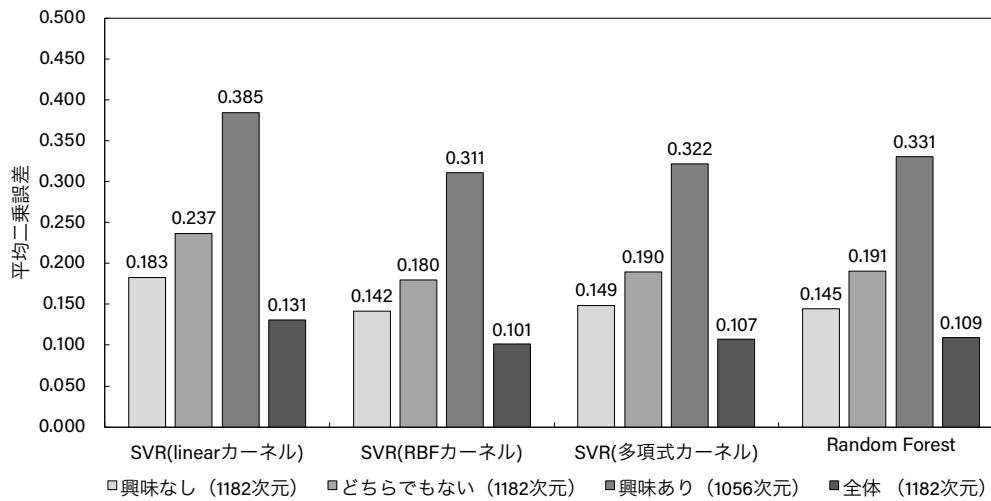


図 27 興味度合い毎に推定した継続利用意欲評価値の平均二乗誤差 (相撲)

のため、正解値が境界値付近の発話文に関しては、推定値の誤差が大きい結果となった。この原因としては、境界値付近のサンプル数の少なさが考えられる。図 28 における継続利用意欲評価値 r の正解値は、全 2683 サンプル中 $r \leq -0.8$ のものが 4 サンプル、 $0.8 \leq r$ のものは 120 サンプルしか存在しなかった。継続利用意欲評価値の分布は図 28 のグラフから正規分布に従っているため、境界値付近のサンプルが統計的に希少であることは正しい。以上のことから、継続利用意欲評価値が境界値付近になる発話文はサンプル数が少ない傾向にあるため、推定難易度が高いといえる。そして、継続利用意欲評価値が境界値付近になる発話文は、ほとんどのユーザの継続利用意欲を大きく向上させる、あるいは大きく低下させる発話文である。つまり、万人の継続利用意欲を向上あるいは低下させる発話文は、非常に希少であり推定難易度も高いといえる。

被験者には視聴後アンケートにて、評価基準についての質問項目について回答を得た。結果、被験者が好む発話文と嫌う発話文の傾向が示唆された。

被験者が好む発話文は、相撲の場合、共に応援してくれる内容、共感、面白い、番組の登場人物に関する補足情報などであった。これは宮澤ら [26] の主張する「発話文の新規性」に該当している。逆に被験者が嫌う発話文は、アスキーアートや「w (笑いの度合い)」のみの文、テレビトピックと関係のない発話、意味の分か

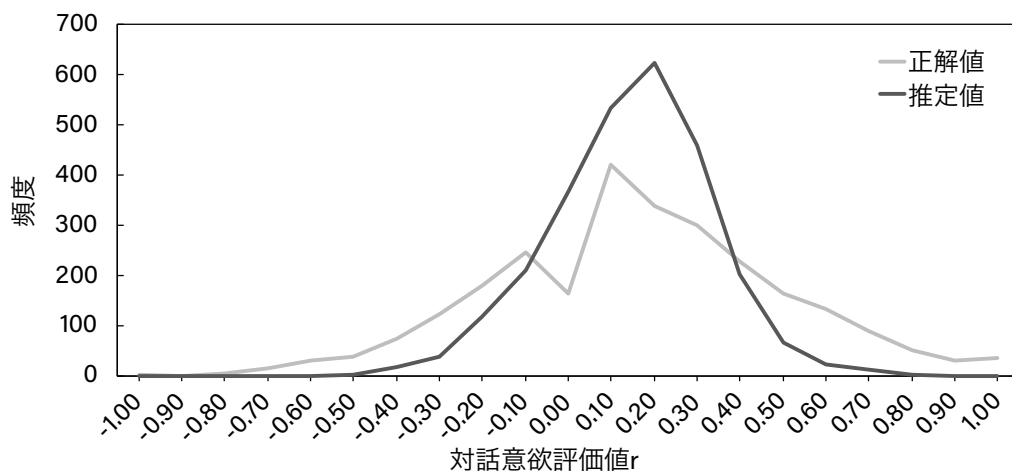


図 28 継続利用意欲評価値のヒストグラム データセット:相撲, 回帰手法:SVR(RBFカーネル)

らない内容であると分かった。特に意味の分からない内容に関しては、発話文がソーシャルメディアコメントであるため、ネット用語が多く含まれることが影響していた。そのため、ソーシャルメディアコメントを発話文として用いる場合は、発話前に文を加工して、口語で用いる言葉遣いや聞き手に伝わる単語に変換する必要があると考えられる。

被験者の中には、登場人物を弄る発話が実験当初は不快であったが、次第にロボットがその人物に好意を抱いているからこそ弄っていると判明したため、継続利用意欲が向上した者も存在した。また、実験の中でネット用語を次第に理解し、ロボットの発話の意味がわかるようになってきたため、good 評価が増えた被験者もいた。これらの事例から、テレビ視聴型雑談ロボットとの対話には第1段階としてロボットとの対話に慣れるための期間が必要であると推察される。慣れ期間にユーザの継続利用意欲を低下させ、ユーザが対話を放棄してしまうとロボットが継続利用されなくなるため、慣れ期間はロボットが超えなければならない一つの壁であるといえる。

次にテレビ番組への興味度合いの違いによる継続利用意欲への影響について考察する。相撲のデータセットに対してアンケート回答から発話文をテレビ番組へ

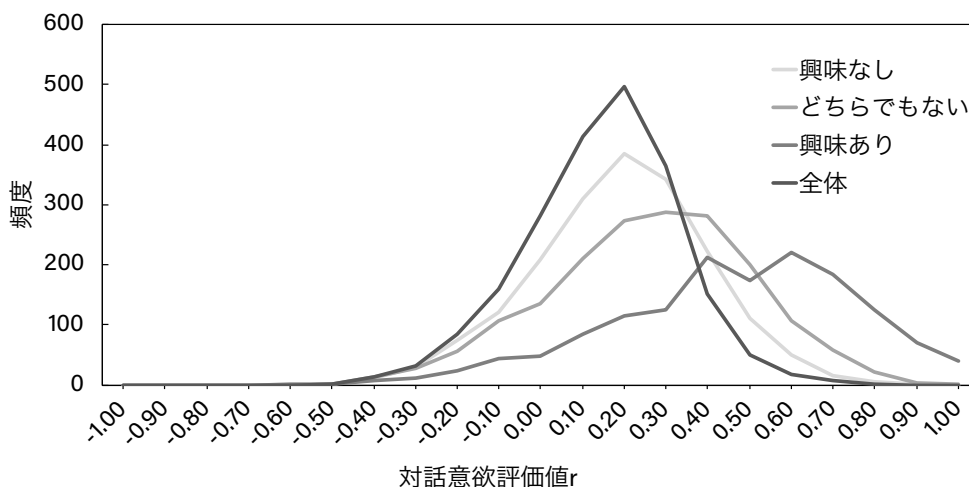


図 29 各興味度合いの継続利用意欲評価値ヒストグラム データセット:相撲, 回帰手法:SVR(RBF カーネル), 誤差 0.5 未満のみをプロット

の興味度合いで 3 分割し, 継続利用意欲評価値を推定した. その結果, 全体に比べて各興味度合いのグループでは推定値の誤差が大きくなった. 本項では, その他に興味度合いのグループ分けが被験者間の継続利用意欲評価に与えた影響を考察する. 各興味度合いの継続利用意欲評価値のヒストグラムを図 29 に示す. 図 29 には, SVR (RBF カーネル) で推定した継続利用意欲評価値のうち, 正解値との誤差が 0.5 未満のもののみをプロットしている. 図 29 より, 全体と比較して各興味度合いにおける継続利用意欲評価値の分布は, 1 の境界値側に遷移している. 特に興味度合いが大きくなるほど, より継続利用意欲の向上を示す 1 の境界値側に遷移している. これは, テレビ番組への興味度合いが高いほど, 推定した継続利用意欲評価値が高いことを示している. このことから, 興味度合い毎の推定結果は被験者数の減少により誤差が増加したものの, 本来推定すべき継続利用意欲を大きく向上させる発話文をより多く推定できたといえる.

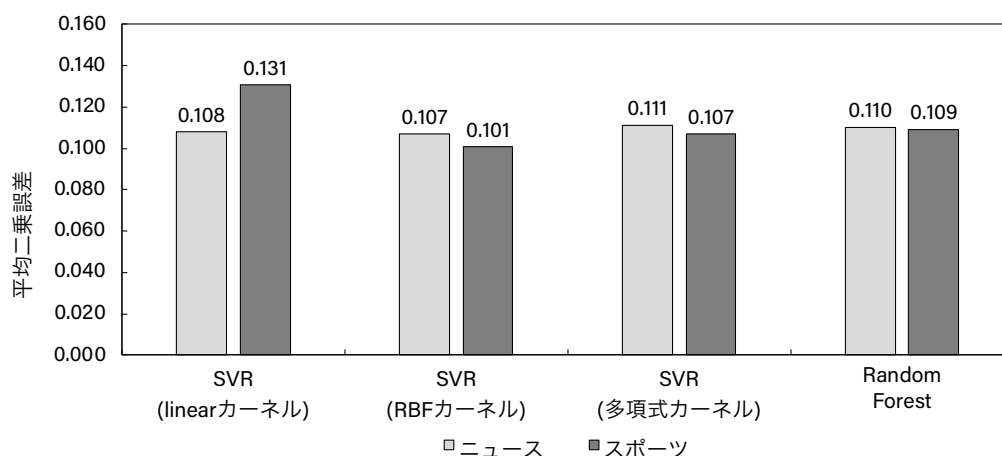


図 30 推定した継続利用意欲評価値の平均二乗誤差における相撲とニュースの比較

4.1.4 ニュースにおける解析結果と考察

テレビ番組がニュースのときの発話文をユーザ継続利用意欲モデルに入力して、継続利用意欲評価値を推定する。解析手法は、相撲と同様にサポートベクタ回帰 (SVR) と Random Forest を採用した。 $k = 5$ の k -fold クロスバリデーションにて算出した5つの平均二乗誤差の平均を図30に示す。次元数は1224次元である。境界値の処理も同様に、推定値 r が $r < -1$ の場合は $r = -1$ 、 $1 < r$ の場合は $r = 1$ に丸める。また、入力するデータは正規化済みである。

図30から、ニュースのデータセットにおいて最も平均二乗誤差が小さい手法は、SVRのlinearカーネルとRBFカーネルであることがわかる。相撲とニュースの誤差を比較すると、linearカーネルの場合のみニュースのほうが誤差が0.023少ない。その他の手法では、相撲とニュースの間に大きな差は見られない。結果から、トピックが大きく変化するニュースのジャンルにおいても、時間ごとのトピックの変化が少ない相撲と同等程度の継続利用意欲評価値の推定精度を得られたといえる。

ニュース番組の解析の結果を踏まえ、相撲と同等の継続利用意欲評価値の推定精度が得られた原因について考察する。

まず、ニュースのデータセットを用いて推定した継続利用意欲評価値の精度が相撲の場合に比べて低くなった理由について考察する。実験設計として、相撲はニュースに比べてテレビトピックの移り変わりが小さいものとして選択した。特に本実験で採用した相撲番組である相撲は、ひとつの場所で2週間続けて試合があり、登場する力士も変わらない。よって、データセット内で似たテレビトピックの似た発話文が出現しやすいと予想した。それに対しニュースのトピックは時事に依存し、1日で複数のニュースが報道される。そのため、相撲よりはデータセット内で似たトピックの似た発話文が出現しにくく、相撲よりニュースのほうが継続利用意欲評価値の推定難易度が高いと予想した。

一方、ニュースの特徴として、ニュースで扱われるトピックは1番組だけでなく複数のニュース番組で報道される。大きな事件のトピックであれば数日に渡って報道される。よって、実験設計の段階では複数のニュース番組に関する発話文を収集すれば、トピックの変化が激しいニュースでも類似する特徴を持った発話文を十分に収集できると予想していた。この予想に基づき、被験者数を相撲より多く設定し、ニュースのデータセットでは複数の番組が採用された。結果的に、ニュースにおけるトピックの変化による発話文の変化は、データ数の多さで補った形となり、相撲と同程度の推定精度を得られたと考えられる。

次に、各ユーザが抱いたDabeliveへの印象について述べる。ニュースのデータセットは、10～60代の幅広い年齢層の男女の被験者に参加していただいた。彼らのアンケート回答から推察されるユーザのDabeliveへの印象について考察する。

Dabeliveはソーシャルメディアコメントを発話文とするため、ネット用語や過激な発言が多い傾向にある。そのため、高齢になるほどDabeliveのロボットが発話した内容は受容されにくいと予想していた。ところが、実験後のアンケートの結果からは、「レポーターを軽く野次ったコメントは笑えました(50代女性)」、「しばらく笑いが止まりませんでした(50代女性)」など、好意的なコメントが多く見られた。一方、やはり発話に含まれる用語の意味が分からないという理由でbad評価をつけたと回答した被験者は高齢になるほど増えていく傾向も見られた。ある60代の男性は「おちょくったコメントが多すぎるのに驚いた」とコメントしており、テレビ番組の登場人物を弄るような発話が受け入れられない人も存在し

た。ソーシャルメディアコメントの利用はユーザの奇をてらう発話を可能にする
が、ソーシャルメディアに慣れ親しんでいない年齢層の人ほど、上記のような感
想を抱きやすいと考えられる。

全体的にテレビの内容と一致していない発話や、下品な内容は bad 評価となる
傾向にあり、これは全年齢に当てはまる傾向であった。漢字の読み間違いも継続
利用意欲を低下させることがアンケートより明らかになった。上記で述べたネッ
トの受容度合いによる影響はあるものの、継続利用意欲の評価の仕方は概ね全年
齢で一致していた。

次に、Bag of words モデルを用いた際の問題点について述べる。ユーザ継続利
用意欲モデルに用いた Bag of words モデルの問題点を議論する。実験で用いた
1 件あたりの発話文に含まれる単語数は、2~10 個ほどと字幕テキストに比べて
極端に少ない傾向にある。そのため、テキスト中に含まれる単語の頻度を特徴量
とする Bag of words モデルでは、一つの発話文に現れる単語の頻度はたかだか 1
つとなってしまう。更に、類似したテレビトピックの発話文であっても、デー
タセット全体で数回しか出現しない単語も多い。本実験では Bag of words の辞書
を作成する際に、出現頻度が極端に少ない単語を辞書に含まない処理をしている
ため、稀な単語のみで構成される発話文をベクトルで表現できていない場合があ
る。一般的に Bag of words が用いられる文書分類等の場面では、統計的に多く出
現するテキストをベクトル表現できていればよい。しかし、本研究で対象とする
ユーザの継続利用意欲を向上させる発話文は奇抜な内容や珍しい文章が含まれる
ため、稀な単語を正確にベクトル表現する必要がある。

Bag of words には単語の持つ意味をベクトル表現に反映できない特性がある。
そのため、同じ意味だがテキスト表現上は違う単語は、別の単語として辞書に含
まれてしまう。収集したデータセットでは、「稀勢の里」「キセ」など、人物の名
前で多く見られた。別の単語として扱われると、「稀勢の里勝った」と「キセ勝っ
た」のベクトルが大きく異なってしまう。

これらの問題点は継続利用意欲評価値の推定精度低下に起因していると考えら
れる。解決策としては、Word2Vec[72] という単語の意味を表現できるベクトル表
現手法の採用が挙げられる。Word2Vec は単語の意味の違いを距離としてベクト

ル空間で表現するため、似た使われ方をする単語がベクトル空間上で近い距離に配置される。よって、稀な単語であっても似た使われ方の単語からベクトル表現可能である。

最後に、推定した継続利用意欲評価値に基づく発話文の選択を行なった。本研究で提案したユーザ継続利用意欲モデルを用いて、実際の発話文を選択した例を示し、選択された発話文の傾向を考察する。表6に示したニュースのデータセットのうち、2018/01/08～2018/01/27までを学習データとしてモデルを作成した。そして、2018/01/28のデータについて継続利用意欲評価値を推定し、1分間で最も継続利用意欲評価値が高い発話文を選択した。従来手法としては、1分間の発話文のうちランダムに一つ選択した。各手法で選択された発話文を表9に示す。

表9より、ランダムに選択した場合は下品な発話文や、攻撃的な発話文が多く見られる一方、継続利用意欲評価値で選択した発話文は、ランダムな場合に比べて下品・攻撃的な発話文が少なくなった。また、「..ふさjk」、「あし」、「ふm」など文章として成立していない発話文も継続利用意欲評価値に基づく場合では選択されていない。継続利用意欲評価値の高い発話文は、テレビ番組の内容に合致しており、的を得た発話文がより多く選択されている。よって、本研究で提案するユーザ継続利用意欲モデルを用いて選択した発話文は、より継続利用意欲が高まる発話文が選択される傾向にあるといえる。しかし、1分間の選択期間に発話文が少ない場合は、ランダムに選択するときと同じ発話文が選択され、継続利用意欲が低下するような発話文になってしまう。そのため適切な選択期間の設定が必要となり、表9で示したような発話文をユーザに実際に発話した際の影響について、改めて調査する必要もある。

表 9: 発話文の選択例

時刻	ランダムな選択	継続利用意欲評価値による選択 (提案手法)
18:00:00	ハゲの非	資産隠しを頑張ったか
18:02:00	ちゃんと前見て謝れ	読んでる

次ページに続く

前ページからの続き

時刻	ランダムな選択 (従来手法)	継続利用意欲評価値 による選択 (提案手法)
18:04:00	大事にならなけりゃその まま逃げてたんだろ	大事にならなけりゃその まま逃げてたんだろ
18:05:00	やっぱ着物保管してるは 嘘なのかな	資産を海外へ移動するの に一生懸命
18:06:00	絶対金持ってるわな	絶対金持ってるわな
18:07:00	タイミングが悪い	結局経営者の器では無か った
18:09:00	詐欺師が詐欺認めるかよ	アカン w
18:10:00	無能神奈川県警で追い込 めるか	言い訳を考えるのに一生 懸命
18:12:00	だれかあたしの xxxx 気持 ちよくしてくれへん w	隠したから出てきたんだ ろ
18:13:00	仮想通貨のニュース自粛 だってよ wcm 打ち切るっ て cc 以外がいいだしてる らしい	いい写真やな
18:15:00	ビットコイン終了	テッちゃんかわいそう
18:16:00	思ってもないことを言う な	兄さんはこうなることを 知っていた
18:17:00	ありえない	昭恵「私も行きます」
18:18:00	1週間雪溶けないんだが	1週間雪溶けないんだが
18:18:00	マスコミ w	降らなくても寒いから
18:20:00	うん謝罪なんか要らない よね	実体化しない仮想通貨の ままあ盗まれたから大ご とにならん

次ページに続く

前ページからの続き

時刻	ランダムな選択 (従来手法)	継続利用意欲評価値 による選択 (提案手法)
18:22:00	かけえな	外国人力士もいいもんだ
18:22:00	全員首で	全員首で
18:24:00	コレは偶然ですか	こいつすぐ態度豹変して 信用できん w
18:25:00	春日野部屋のリークした の誰だろうね	被害者本人が声上げたの かと思ったがちがうのか
18:26:00	cm 明け、また最初から見 せられるのかな	さっきから cm しか見てな い
18:28:00	八角 4 人 w	保身に必死
18:30:00	くる … ぞ ..	協会から出ていけよ
18:30:00	いいお	これは寒いだろうな
18:31:00	怖い怖い w	僕たちマスコミ
18:33:00	クスリやってんだろ w	怖い w
18:34:00	長い	勝ったな
18:36:00	ちょっとお酒入ってるから 明るい	11 人いれば余裕
18:37:00	他の離反票を確保してる なら出すだろうけど	他の離反票を確保してる なら出すだろうけど
18:38:00	流石に他人の出る出ない の言及はその部屋の力士 に迷惑かかるし	バンキシャおめでと w
18:41:00	女性をかばって	こええ
18:42:00	芸人みたい	ひええ
18:43:00	論文かけるでー	がんばれ

次ページに続く

前ページからの続き

時刻	ランダムな選択 (従来手法)	継続利用意欲評価値 による選択 (提案手法)
18:44:00	ノーマーク言ってもそば だろ w	これは無理だ w
18:46:00	3千年って何でわかるの	落石こわすぎ
18:48:00	日本に安全地帯などない わ	ほんと信用できんな地震 学者の類は
18:49:00	地球のコアを停止させた ら地震なくなるよ	3千年前とかって言い訳し てるだけやろ
18:50:00	亜門は違いが分かるから	ファミチキ派
18:52:00	言えよ	量子力学の時代に未来予 知とか w
18:52:00	斎藤って増毛したのに禿 げだよな	唐揚げはカロリー高い
18:53:00	今年はすごいな ..	今年はすごいな ..
18:54:00	海岸は海岸を開拓しろよ	海岸は海岸を開拓しろよ
22:00:00	パワーおめ	アッーいいっすね
22:01:00	春日野は理事選出るんだ ってな w	もはや協会いらないな
22:03:00	親方がなんで謝るの	親方がなんで謝るの
22:04:00	暴行が日常	なんで運転したんだ
22:05:00	なんや序ノ口やろこいつ	貴乃花はこれがあったか ら、相撲協会に報告しな かったのか
22:06:00	マスゴミ	マスゴミ
22:08:00	ホテルがあるんだここに	くだらんコメントすんな

次ページに続く

前ページからの続き

時刻	ランダムな選択 (従来手法)	継続利用意欲評価値 による選択 (提案手法)
22:09:00	..おっさんはよ答えんかい	主な視聴者のじじばば引きつけようと思ったら相撲しかねえんだろ
22:11:00	親戚って言われたら信じる	貴ノ岩って誰もう忘れたわ
22:11:00	でた、でくの棒 w	でた、でくの棒 w
22:12:00	今立行司って不在なのか	フジテレビが不祥事だよ
22:14:00	見送りじゃないか反貴乃花の願望	寺尾も出馬しないかな
22:16:00	bba	これが協会のダメなところだな
22:17:00	尾木は反貴乃花派	マスコミにいるいるのか
22:19:00	日本人ダメだな	邪魔
22:20:00	..君もおっちゃんやろ	またテレビに出まくってダメになります
22:21:00	..ふさ jk	八代「相撲うぜえ
22:23:00	みえ	励みになったんやな
22:25:00	でれ	国別はじくソフトあるだろ
22:26:00	ベンツ	もうこいつが横綱でいいよ
22:27:00	すごいニダねえ	真央ちゃんいなくなってbsしかいねーなあ..
22:29:00	ええな	凍らせて削るしか
22:30:00	こけし界のヒロイン	大化けするか

次ページに続く

前ページからの続き

時刻	ランダムな選択 (従来手法)	継続利用意欲評価値 による選択 (提案手法)
22:31:00	あし	無理
22:31:00	スピードスケートの選手 みたい	スピードスケートの選手 みたい
22:33:00	その心はって聞けよ	その心はって聞けよ
22:33:00	噛み下ろした方がいい	髪型大事
22:35:00	よいやさ	成長期ってすごい
22:37:00	ベッキーは会見で嘘つい たからな	小室はかわいそう
22:38:00	こいつを張り込もう	元から勝ってないだろ w
22:40:00	公人だけは許すはあととは どうでもいい	週刊誌として間違ってる ない、テレビが叩きすぎ
22:40:00	そもそもしゅうかんしか わかない	文春がネタにして勝って る人がいるんだから何か 問題でもあるの
22:41:00	テレビ局がずっとやって るせいだろ	これよりコインチェック のほうがやばいのに
22:43:00	資産隠してるだろ	読んでるだけ
22:44:00	籠池に代わるおもちゃに なるかなこの社長は	わろてるで
22:44:00	詐欺ではない	コインチェックは nem 流 出だから対して話題にな らない。
22:46:00	ダイエット w	失礼な
22:48:00	nhk が自慢するのは 8k	出川コインやった

次ページに続く

前ページからの続き

時刻	ランダムな選択 (従来手法)	継続利用意欲評価値 による選択 (提案手法)
22:49:00	あいからずのいフジだな w	はーすげえな
22:50:00	ウジ「やったぜ」	木の根っこが見えた、はいはいすごいね
22:52:00	止まっているのか	こわ1
22:53:00	だれだよ	テレビ局の映像よりきれいじゃん
22:54:00	やったぜ	やばいよやばいよ
22:56:00	俺も gopro 買おうかな	さっき見たけど
22:58:00	こわ	動くのかよすごいな
22:59:00	こんなん当たったら	こんなん当たったら
22:59:00	運パラ上げとこ	いるはずの無い女性の声...
23:00:00	ゴンドラから落ちた人	多分振り返ったら連れて逝かれるタイプの奴だわ
23:01:00	4kって地上波放送ないんでしょう	無理だべ
23:03:00	ふ m	すごい映像
23:04:00	九死に一生を得たら笑っちゃうよ	ひるおび感がすごい
23:06:00	あほかもうなれんわ w	富士山噴火したらまじやばい
23:06:00	1 文無しになる可能性のほうが高い	寒いからいや

次ページに続く

前ページからの続き

時刻	ランダムな選択 (従来手法)	継続利用意欲評価値 による選択 (提案手法)
23:07:00	もう戻ってこないよ	返金するとは言ったがいつとは言っていない…つまり…10年後20年後もありうるということ

以上

今後は ground truth となるコメントへの評価値の入力数 (51,749 件の正解データ) や番組の種類 (ニュース・相撲) をさらに増やした上で評価を行うことが必要である。また、個人の趣向を反映するために必要なモデルを構築するためには膨大な量のデータが必要であるため、フィルタリングされたコメントとランダムに抽出されるコメントによってユーザの印象がどのように変化するかについて評価するには至っていない。従って、今回提案する手法によってはユーザの利用初期の嗜好の傾向を推定することは困難である。そこで新たに追加する手法そのものの検討として、TV 雑談アプリの初回使用時にユーザの嗜好を絞り込むための質問を設定することや、協調フィルタリングにより他のユーザの嗜好を用いた方法についての効果を検証する必要がある。また TV 雑談サーバにて SNS コメント評価の収集を今後も続け、フィルタにフィードバックすることで、個人特化するフィルタの精度を高め、TV 雑談ロボットの発話するコメントに対しユーザの印象がどのように変化するかを調査する必要がある。また収集された評価データの分析から、個人特化するためのフィルタを検証するだけでなく、性差による SNS コメントの好みの違い、年代別の好みの傾向などが明らかになると考えられる。このように収集された SNS コメントと、それに対する評価を不特定多数のユーザに行なってもらうことで、新たな機能の検討と言った可能性につながることを期待できる。

4.2 情動共有

本節で提案する情動共有とは対話ロボットとユーザが共有する空間においてユーザが喚起するであろう感情を、対話ロボットがリアルタイムに推定し、ロボットがユーザに対し共感を示す振る舞いを実行する機能と定義する。人が持つ感情には怒り、悲しみ、恐怖、興奮など様々あるが、TV視聴時にはそれらを喚起する雰囲気が存在すると考えられる。雰囲気には例えば、スポーツ中継などの視聴時に発生すると考えられるネガティブおよびポジティブな感情をともに含んだ盛り上がり（高活性）やバラエティ番組などで見られる笑いが上げられる。各雰囲気を構成すると考えられる要素とそれに関連して決定される振舞いとその要素を表10に示す。これらの要素を基に、ロボットの振舞いを大きく発話およびモーションに分類した。雰囲気推定するためのリソースとして、ユーザ情報を取得する方法が挙げられるが、今回はユーザ自身の持つ情報以外の情報として、TV雑談システムにて収集されるSNSコメントという一般大衆の意見を用いる。ユーザ以外の情報から雰囲気を推定しロボットの振舞いを決定することで、ユーザに対するセンシングされているという心理的負担やセンサを装着する身体的負担を軽減する。雰囲気の具体的な推定方法と、その雰囲気に応じてTV雑談ロボットの振る舞いをどのように変化させるかについて以下で詳述し、情動共有が適用されたロボットがユーザに与える印象の評価を行う。

表 10 雰囲気に応じたロボットの振舞いと要素

要素	振舞い		
	ロボットの発話・モーション		
	いつ	何を	どのように
適用区間	○		
レベル		○	○
発話文		○	
韻律情報			○
感情			○

4.2.1 対話ロボットの情動共有に関する先行研究

ロボットとのテレビ視聴時の臨場感をより高めるためにどのような要素が効果的であるか、実装の方針の決定に関連するいくつかの研究を紹介する。まず複数台のロボットによる対話が人に与える印象について調査された研究について述べる。飯尾らは複数の対話ロボットとともにユーザが対話を行う環境において、ロボットが人に与える対話感への影響を調査した [73]。被験者は、1台、または3台のロボットと対話を行った後、それぞれの状況でロボットの対話に対する印象を評価した。その結果、3台のロボットとの対話において「対話が成立した」と見なされ、「ロボットの発話した内容について考えが深まった」という印象を与えることができ、1台のロボットとの会話よりも肯定的に評価された。

次に人と対話ロボットが感情を共有する事により得られる効果について調査された研究について述べる。松本らは対話ロボットとともに映画を視聴する際、ロボットに対する主観的な印象を評価する実験を行った [74]。実験には映像から換気される感情に応じて、ロボットは身体動作や韻律を動的に変えて発話を行うことができるロボットが使用された。この結果、被験者は独りで映画を視聴する条件と比較して、ロボットとともに映画を見る条件においてより感情が増幅されることが示唆された。つまりロボットの演出する感情と、映像によって引き起こされた人間の感情が一致するとき、結果として共感が生じており、これにより人間の感情が増幅する可能性がある。この共感によってユーザはロボットに対してだけでなく、他の現象に対しても感じる印象を改善する可能性がある。

最後に SNS の解析に関する研究について述べる。一般的に、TV 番組が視聴者を惹きつけると、番組に関する SNS 上のコメント数が急増することがしばしば発生する。この現象を利用し、一部の研究者らは SNS へ投稿されたコメント数から TV 番組の感情的な活性度を推定した [75, 76, 77]。Shamma らは TV で放送された大統領選の討論のシーンをトピック単位に分割し、各トピックの活性度を推定し、コメント数の平均と標準偏差から活性のピーク点を検出する方法を提案した [78]。ピーク点は Twitter [55] に投稿されたコメント数の推移に対してニュートン法 [79] を用いて検出された。中澤ら [80] は、Twitter のコメント機能である Tweet 数から各シーンの主要人物とイベントをラベル付けすることで、TV 番組

中の重要シーンを検出する手法を提案した。これらの手法はTV番組の活性度のピーク点を検出するものであり、活性度の強度や持続時間がどの程度のものかを判定することができない。また、未知のピーク点を判断するためには、コメント数の情報が必要であるため、ピーク点をリアルタイムで検出することができない。

4.2.2 雰囲気 の推定とロボットの振る舞いを決定する手法

情動共有をTV雑談ロボットにあたり、盛り上がりおよび笑いの2つの条件を扱う。ロボットの振る舞いとして、主に発話・モーションのタイミングおよびレベルを決定する方法について述べる。それらの要素を基にロボットの振る舞いを設計する方法について述べる。今回はTV雑談ロボットユーザが放送中のTV番組に対する感情の変化が、SNS上にリアルタイムにコメントを投稿する不特定多数のSNSユーザの感情変化と一致すると仮定して、情動共有適用時のロボットの表現を決定する。使用するSNSリソースには、SNSユーザのコメントをリアルタイムに収集するという点で安定性のあるニコニコ実況を用いる。ロボットが盛り上がり表現を行う場合、SNS上に投稿されるコメント数が一斉に増加するタイミングで盛り上がり区間の判定を行う。

本研究では、4.2.1にて述べた中澤らの手法[80]を参考に、TV番組放送時にニコニコ実況に書き込まれたコメント数を5秒ごとに取得し、以下の式2、3により盛り上がり区間の判定を行う。ここで μ 、 ρ は、判定区間から15分前までの盛り上がりでないデータ群およびコメント数の増加量が負であったデータ群の平均と標準偏差である。

$$Threshold = \mu + 2\rho \quad (2)$$

$$C_i > Threshold: Exciting \quad (3)$$

取得したコメント数が設定した閾値以上の場合、その区間を盛り上がり区間とし、区間内にロボットに盛り上がり表現を行わせるものとする。一方ロボットが笑いの表現を行う場合は、SNS上に投稿される特定のコメントに注目する。特定のコメントとして、笑いを表現するキーワード（今回は'w','W','w','W','笑','草','芝','わろた','ワロタ','わら','吹いた','フィタ','藁'が含まれるコメントと

した) 数を, 盛り上がり表現と同様に5秒ごとに取得し, 以下の式 4, 5 により笑い区間の判定を行う. ここで μ , ρ は, 判定区間から15分前までの笑いを表すキーワードを含まないデータ群およびコメント数の増加量が負であったデータ群の平均値および標準偏差である.

$$Threshold = \mu + \rho \quad (4)$$

$$L_i > Threshold: \textit{Laughing} \quad (5)$$

ここで述べた手法により盛り上がりと判定された区間においても, 盛り上がりの程度に差があると考えられる. そのため, 盛り上がりの程度を盛り上がりレベルと定義し, レベルを判定する手法についてでる. これまでに Twitter において5秒間に取得された Tweet 数と基準となる閾値との差から, 盛り上がり区間における盛り上がりレベルの判定を行ってきた [81]. この手法を用いて, 以下の式 6, 7, 8, 9 より盛り上がり区間におけるレベル判定を行う. ここでニコニコ実況において5秒間に書き込まれたコメント数 $Threshold$ は式 2, 3 により求めた閾値である.

$$C_i < Threshold: \textit{Excitinglevel} = 0 \quad (6)$$

$$Threshold < C_i < 1.3 \times Threshold: \textit{Excitinglevel} = 1 \quad (7)$$

$$1.3 \times Threshold < C_i < 1.6 \times Threshold: \textit{Excitinglevel} = 2 \quad (8)$$

$$1.6 \times Threshold < C_i: \textit{Excitinglevel} = 3 \quad (9)$$

同様に笑いにもレベルが存在すると仮定し, レベル 0 から 3 の 4 段階に設定した. 以下の式 10, 11, 12, 13 より笑い区間におけるレベル判定を行う. ここではニコニコ実況において5秒間に書き込まれたコメント数, $Threshold$ は 4, 5 の手法により求めた閾値である.

$$L_i < Threshold: \textit{LaughingLevel} = 0 \quad (10)$$

$$Threshold < L_i < 0.33 \times Threshold: \textit{LaughingLevel} = 1 \quad (11)$$

$$0.33 \times Threshold < L_i < 0.66 \times Threshold: \textit{LaughingLevel} = 2 \quad (12)$$

$$0.66 \times Threshold < L_i: \textit{LaughingLevel} = 3 \quad (13)$$



図 31 レベルごとの盛り上がり時のロボットの振る舞い

以上から得られた情動判定手法をもとに盛り上がりを表現する手法について述べる。情動共有に基づく振舞いを行うロボットの Appearance は、ともに人型であり 2次元のバーチャルエージェントと身体性を伴ったロボットとを用いる。図 31 に示すようなバーチャルエージェントには TV 雑談ロボットをスマートフォンアプリケーションに実装したアプリケーション “Dabelive” (株式会社 amirobo tech[33]) をもとに構成した。身体性を持つロボットには pip 製のうなずきかぼちゃん [58] を用い、肩関節部にモータを挿入することで腕を大きく振る動きができるよう改造した。



図 32 レベルごとの笑い時のロボットの振る舞い

ロボットの発話内容を決定するために、リアルタイムに放送されている番組に対する SNS 上のコメントを収集可能な TV 雑談サーバを用いる。TV 雑談サーバが収取する SNS コメントのソースとしては、高いリアルタイム性が安定して確保できるニコニコ実況を主に用いる。収集された各局の番組に対するコメントから、過激なコメントを除外したコメント群からランダムに選択した1つのコメントを7秒おきにクライアント（TV 雑談ロボット）に送信し、音声合成によってコメントを発話させる。本研究では、ニコニコ実況からコメントを取得する際に各種区間とレベルを判定し、その結果をコメントサーバに記録する。コメントサーバは

区間とレベルに応じて、クライアントに送信するコメントを情動に合わせて変更する。盛り上がりの例では、盛り上がり区間中複数のロボットに対して、たたみかけるように連続して発話を行わせることで活性度の高いムードを演出する。クライアントに対し発話コメントをレベルによって以下のように変更する。

Exciting Level 0 : 7秒につき、1体のロボットが発話1コメントを発話。

Exciting Level 1 : 7秒につき、2体のロボットがそれぞれ1コメントずつ発話。

Exciting Level 2 : 7秒につき、3体のロボットがそれぞれ1コメントずつ発話。

Exciting Level 3 : 7秒につき、5体のロボットがそれぞれ1コメントずつ発話。

バーチャル環境におけるロボットの場合は、図 31 に示すようにレベルに応じて発話するロボットの数を増やす。身体性を伴うロボットにおいても同様にロボットの台数を増やすことで対応する。

一方、笑い区間中のロボットの振舞いでは、通常の7秒間に1回発話を行うロボットに加え、笑い声を発話するロボットの数と笑い声をレベルに応じて変更していく。笑い声は事前に録音された人の笑い声を、主観的に大笑い・中笑い・小笑いに分類した。レベルごとに以下の変更を行う。

Laughing Level 0 : 7秒につき、1体のロボットが発話1コメントを発話。

Laughing Level 1 : 7秒につき、1体のロボットが1コメントを発話し、別の1体が小笑いを発声。

Laughing Level 2 : 7秒につき、1体のロボットが1コメントを発話し、別の2体がそれぞれ中笑、小笑いを発声。

Laughing Level 3 : 7秒につき、1体のロボットが1コメントを発話し、別の4体がそれぞれ別の大笑いを発声。

バーチャル環境でのロボットの振舞いは図 31, 32 に示す通りである。笑い声の発声タイミングは、本章で述べた通りであるが、複数のロボットが笑い声を発生する場合は各ロボットの発生タイミングを2ミリ秒ずつずらすことで、笑い声が重なりエコー掛かることがないようにした。

4.2.3 盛り上がりにおける情動共有の評価と考察

TV番組の盛り上がり場面においてバーチャル・フィジカルの両条件においてロボットを用いて盛り上がりの表現を行い、ユーザのロボットに対する印象の差を比較について詳述する。被験者にはアピランスの異なるロボットと共にTV番組を視聴してもらい、各番組の視聴後にアンケートに回答してもらおう。

実験構成について、映像コンテンツは盛り上がるタイミングが明確であり、かつ盛り上がりの種類に個人差が少ないサッカー中継とした。TV番組には2018年10月16日に放送されたサッカーキリン杯2018の日本対ウルグアイ戦を用いた。番組放送時にTV雑談エンジンによって収集されたSNSコメントを用いて、3章にて述べた手法により盛り上がり区間と盛り上がりレベルを検出し、盛り上がりと判定された場面の中の1つを実験映像コンテンツとした。

本実験で用いたフィジカルなTV雑談ロボットには、うなづきかぼちゃんを使用する。情動共有時は5台のロボットが識別可能なように、音声合成にはVoiceText Web API[59]を用いて異なる音声モデルを設定し、見た目での識別も考慮し異なる色の服装にした。本実験では実験時の再現性を保つため、時系列に並べたコメントを事前に作成し、7秒間隔でTV雑談ロボットがコメントを読み上げるようデザインした。コメントは、本実験で使用する映像コンテンツが放送されていた間にTV雑談エンジンにて収集されたコメントを用いる。11名の被験者(20~30代男女)にフィジカル・バーチャルの2条件および盛り上がりの情動共有あり・なしの計4条件のロボットと共にTV視聴を行ってもらった際の印象をアンケートによって評価する。TV番組の映像を事前に1度視聴してもらい、映像内容を把握してもらおう。その後、

- 情動共有なし条件・バーチャルロボット
- 情動共有あり条件・バーチャルロボット
- 情動共有なし条件・フィジカルロボット
- 情動共有あり条件・フィジカルロボット



図 33 アピアランスの違いごとの実験時の様子

の各条件で計 4 回視聴してもらおう。図 33 に実験時の様子を示す。視聴順序による影響を考慮し、被験者ごとに各条件にて視聴する順序を変更し、カウンターバランスをとった。視聴後は逐次以下 5 項目に対して 7 段階のリッカード尺度による評価値を付けてもらおう (1:全くそう思わない, 7:非常にそう思う)。

- Q₁ ロボットは盛り上がっているように感じましたか？
- Q₂ ロボットは親しみやすかったですか？
- Q₃ ロボットの盛り上がり違和感はなかったですか？
- Q₄ ロボットを継続的に利用したいと感じましたか？
- Q₅ ロボットの発話内容に共感できましたか？

主観評価について、各項目の平均スコア、標準偏差を図 34 に示す。各項目に対し、以下の 4 つの組に分けて t 検定を行なった。その結果を図 35 に示す。ここで、* : $p < 0.05$, † : $0.05 < p < 0.1$ とする。

- Pair 1: 情動共有なし条件・バーチャルロボットと情動共有なし条件・フィジカルロボット
- Pair 2: 情動共有なし条件・バーチャルロボットと情動共有あり条件・バーチャルロボット

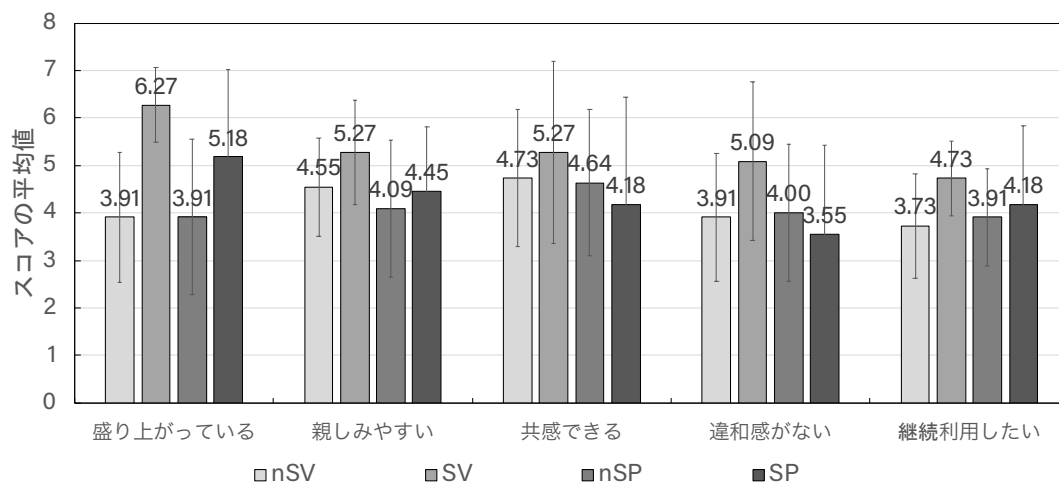


図 34 ロボットの盛り上がり時の表現に対する各質問における条件ごとのスコア

Pair 3: 情動共有なし条件・フィジカルロボットと情動共有あり条件・フィジカルロボット

Pair 4: 情動共有あり条件・バーチャルロボットと情動共有あり条件・フィジカルロボット

Pair 1 の組み合わせに関して、全ての項目で有意差は認められなかった。

Pair 2 に関して、「ロボットは盛り上がっていたか」の項目について有意差が認められた。他の項目に関して有意差は認められないものの情動共有あり条件の方が好印象であった。

Pair 3 に関して、「ロボットは盛り上がっていたか」について有意傾向は認められたものの、Pair 2 の比較ほど差は見られなかった。

Pair 4 の比較に関して、「ロボットは盛り上がっていたか」、「ロボットの盛り上がり違和感はないか」、「ロボットの発話内容に共感できたか」の3項目で有意差が認められた。残りの項目に関して、複数体のロボットの方が好印象であった。

以上の結果から、本実験における盛り上がり共有の場面では、情動共有が適用されていないロボットを用いるよりも、適用されたロボットを用いた方がユーザーにより盛り上がっている印象を与えることがわかった。主観評価における自由記

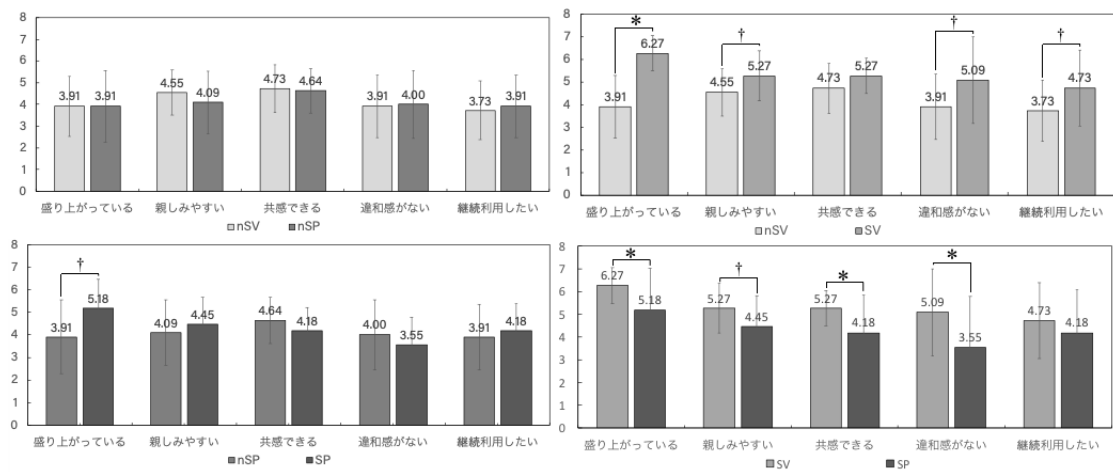


図 35 盛り上がり評価実験における各条件の組み合わせごとの比較 (左上: 情動共有なし条件・バーチャルロボットと情動共有なし条件・フィジカルロボット, 右上: 情動共有なし条件・バーチャルロボットと情動共有あり条件・バーチャルロボット, 左下: 情動共有なし条件・フィジカルロボットと情動共有あり条件・フィジカルロボット, 右下: 情動共有あり条件・バーチャルロボットと情動共有あり条件・フィジカルロボット)

述をまとめると、バーチャル・フィジカルロボットの両方の場合において、情動共有適用時には「興奮するようなシーンで畳み掛けるように発話していたから」、「ワイワイしている感じあった」とあり、1体の場合では「淡々と話していたから」、「盛り上がりそうなところでも一定の周期で発言しているから」とある。つまり、ロボットのアピランスに関わらず、盛り上がりの情動共有で表現した発話者数と発話頻度の増加がユーザへの盛り上がり印象の向上の要因になっていると考えられる。

ロボットのアピランスの比較に関して、情動共有なしの場合は、バーチャルなロボットに対し「字幕があると「?」とか「!」「w」などを確認できるので、ただ音声を聞くだけの時と比べて印象が変わった」、「文字で表示される点はよかった」といったポジティブな意見と、「あまり表情が変わらなかったり、物理的に人形がないのがやや不満だった」といったネガティブな意見が見られた。これらの結果から、バーチャル条件においても表情の変化には物足りなさを感じる被験者が

見られ、表情の変化などノンバーバルな情報を提示することが盛り上がりの雰囲気の中でユーザに求められる重要な要素であることがわかった。また、バーチャル・フィジカル両条件で発話する内容は同じため、字幕機能をフィジカルロボットに適用することがアピアランスに関わらず共通して重要な要素であることがわかった。

ロボットに対しては「一緒にテレビを見てくれる感じがしたから」、「膝に乗せて一緒にテレビを見ていたい」といったポジティブな意見と、「字幕がないため、何を話しているのかあまりわからないところがあった」といったネガティブ意見があった。これらの意見から、バーチャルなロボットは発話に合わせた字幕表示が高評価を受け、実体がないことに低評価を受けている。また、フィジカルなロボットは実体があることに高評価を受け、字幕がないことに低評価を受けている。これらの点は相反するものであり、評価が分かれたために結果として、条件間のスコアに有意差が認められなかったと考えられる。

次に、情動共有適用時のロボットのアピアランスの比較に関する考察を述べる。「盛り上がっているように感じたか」、「親しみやすかったか」の両項目において、アピアランスの違いに関わらず「盛り上がっている感が出ていた」というコメントがある一方で、バーチャルロボットでは「それぞれのキャラがピコピコ動いていたため」のようなスマートフォン上でロボットの数が変化することに関するコメントがあり、ロボットでは「擬音が違和感」、「返事や相槌がなかったところがすこし違和感を感じる」のような発話や動作に関するコメントがあった。これらのことから、バーチャルなロボットが盛り上がり時に追加で出現することが盛り上がり時とそうでない時の視覚差を生み、ユーザが好印象を持つ要因になっていると考えられる。またフィジカルなロボットは、バーチャル条件では感じられなかった発話の違和感や、返事や相槌といった動作に対する違和感が生じていたことから、実体を持つことでユーザがロボットに対する期待感を増幅させ、これにより違和感が生じ、スコアが低くなる要因となったと考えられる。

4.2.4 笑いにおける情動共有の評価と考察

TV番組の誘い笑いを喚起する場面においてバーチャル・フィジカルの両環境においてロボットを用いて笑い表現を行い、ユーザのロボットに対する印象の差を比較について詳述する。4章での評価方法と同様、被験者にはアピランスの異なるロボットと共にTV番組を視聴してもらい、各条件での番組視聴後にアンケートに回答してもらう。

映像コンテンツは比較的万人に対して明確に笑いを誘う場面が多く見られた、2018年12月30日に放送されたバラエティ番組“アメトーーク!”中の“アメトーーク大賞2018ビジュアル映像部門”の一部を用いる。番組放送時にTV雑談エンジンによって収集されたSNSコメントを用いて、3章にて述べた手法により笑い区間とレベルを検出し、笑い区間と判定された場面が含まれたシーンを実験映像コンテンツとした。なお今回は映像コンテンツの時間とロボットが発話を行うタイミングが被験者によって同期したものとなるよう、図33に示すように実験映像コンテンツと各条件のロボットが並んだ状態の映像を視聴することでロボットの評価を行う。

4.2.3と同様、実験で用いたフィジカルなTV雑談ロボットには、うなづきかぼちゃん使用する。情動共有時は5台のロボットが識別可能なように、音声合成にはVoiceText Web APIを用いて異なる音声モデルを設定し、見た目での識別も考慮し異なる色の服装にした。また、バーチャルなロボットにおいて、ロボットが笑い声を発生する際、ロボットの表情のみ図32に示すような笑顔に変化した状態となる。ロボットの発話文には、時系列に並べたコメントを事前に作成し、7秒間隔でTV雑談ロボットがコメントを読み上げるようデザインした。コメントは、本実験で使用する映像コンテンツが放送されていた間にTV雑談エンジンにて収集されたコメントを用いる。

10名の被験者(20~40代男女)にフィジカル・バーチャルの2条件および笑いの情動共有あり・なしの計4条件のロボットと共にTV視聴を行ってもらった際の印象をアンケートによって評価する。被験者はロボットの評価を行う前に、実験に使用する映像を1度視聴し映像内容を把握する。その後以下の各条件で計4回視聴してもらう。

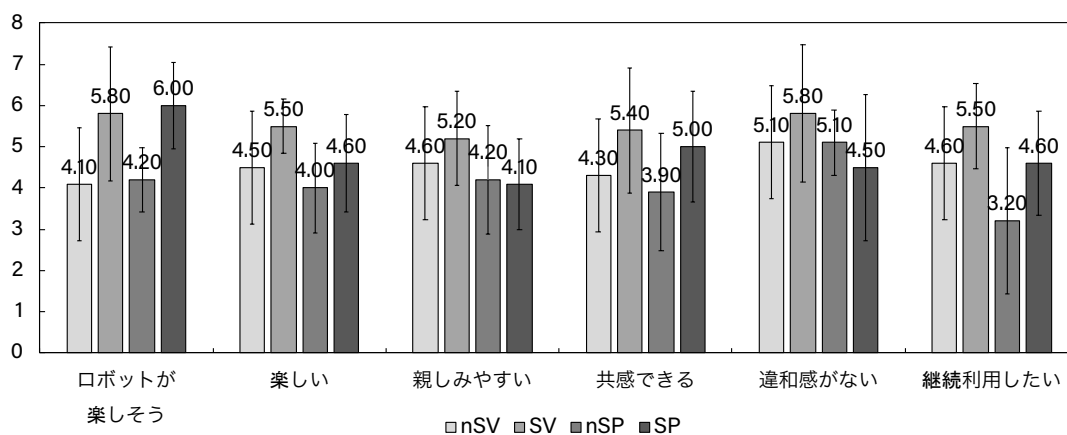


図 36 ロボットの笑い時の表現に対する各質問における条件ごとのスコア

- 情動共有なし条件・バーチャルロボット
- 情動共有あり条件・バーチャルロボット
- 情動共有なし条件・フィジカルロボット
- 情動共有あり条件・フィジカルロボット

視聴順序による影響を考慮し、被験者ごとに各条件にて視聴する順序を変更し、カウンターバランスをとった。視聴後以下6項目に対して7段階のリッカード尺度による評価値を付けてもらう(1:全くそう思わない, 7:非常にそう思う)。

Q_i ロボットは番組を楽しんでいるように感じましたか？

Q_{ii} ロボットと番組を視聴して楽しいと感じましたか？

Q_{iii} ロボットは親しみやすかったですか？

Q_{iv} ロボットと同じ感情を共有しているように感じましたか？

Q_v ロボットの言動に違和感はなかったですか？

Q_{vi} ロボットを継続的に利用したいと感じましたか？

主観評価について、各項目の平均スコア、標準偏差を図 36 に示す。各項目に対し以下の組み合わせの4つの組に分けてt検定を行なった。その結果を図 37 に示す。ここで、*: $p < 0.05$ (有意差), †: $0.05 < p < 0.1$ (有意傾向)とする。

Pair I: 情動共有なし条件・バーチャルロボットと情動共有なし条件・フィジカルロボット

Pair II: 情動共有なし条件・バーチャルロボットと情動共有あり条件・バーチャルロボット

Pair III: 情動共有なし条件・フィジカルロボットと情動共有あり条件・フィジカルロボット

Pair IV: 情動共有あり条件・バーチャルロボットと情動共有あり条件・フィジカルロボット

Pair Iの組み合わせに関して、ロボットとの共感について有意傾向が見られたが、有意差は得られなかった。一方で継続利用意欲の促進に関して有意差が認められ、バーチャルな見た目のロボットがフィジカルなそれに比べ高いスコアを得ることが示された。

Pair IIに関して、「ロボットが楽しんでいる」の項目について有意差が認められた。他の項目に関して有意差は認められないものの情動共有あり条件の方が一貫して好印象であった。

Pair IIIに関して、「ロボットが楽しんでいる」、「共感している」、「継続利用意欲」の項目について有意差が認められた。しかしフィジカルなロボットを用いる場合において、情動共有の有の場合に一貫して有意となる傾向は見られなかった。

Pair IVの比較に関して、「ユーザが楽しいと感じる」、「ロボットは親しみやすい」、「ロボットの言動に違和感がない」の3項目で有意差が認められた。情動共有を適用した条件においてロボットはバーチャルな見た目の方がよりユーザにとって好印象であることがわかった。

また、4条件のロボットのうち、同じようなTV番組を視聴する際にどの条件のロボットを最も使いたいかという質問を被験者に行ったところ、10名中7名が

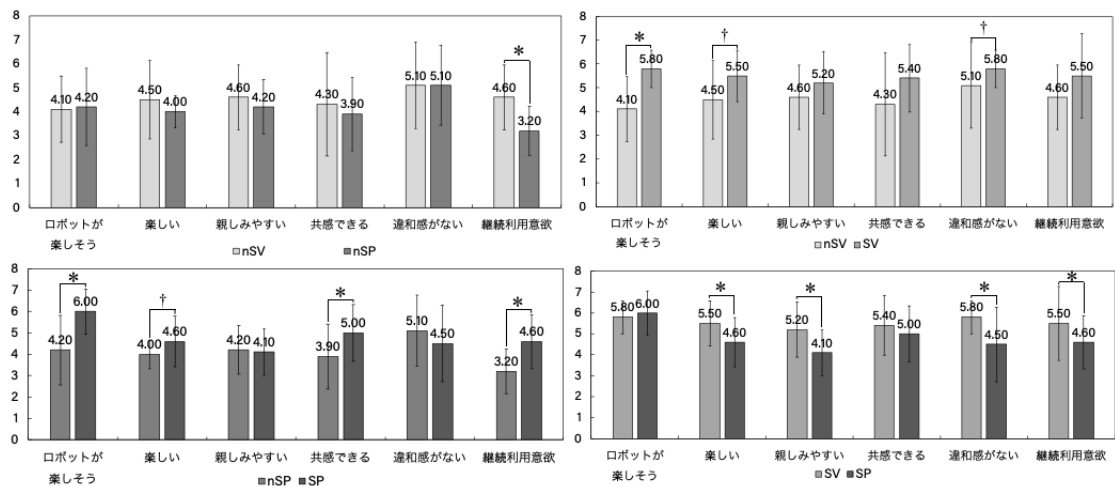


図 37 笑いにおける各条件の組み合わせごとの比較 (左上：情動共有なし条件・バーチャルロボットと情動共有なし条件・フィジカルロボット，右上：情動共有なし条件・バーチャルロボットと情動共有あり条件・バーチャルロボット，左下：情動共有なし条件・フィジカルロボットと情動共有あり条件・フィジカルロボット，右下：情動共有あり条件・バーチャルロボットと情動共有あり条件・フィジカルロボット)

情動共有あり条件かつ、バーチャルロボットを用いた場合の条件で利用したいという回答を得た。

本実験の結果から、情動共有において笑いを扱う場合、ロボットの評価はバーチャルなロボットであるほうが好ましいことが示された。この理由として、今回の実験では特にロボットの表情が変化すること、発話文章が表示されること、笑い区間外ではロボットが画面から消えること、などバーチャルロボットにおいてのみ適用される表現が含まれていたことから、このような結果に繋がったと考えられる。今回実装した笑い表現について、情動共有を適用した際のロボットについての被験者の自由記述からポジティブな意見として、「自分が面白いと感じる場面でロボットが笑っており、同じ感情を共有しているように感じた」、「ロボットの笑い声に誘われて笑ってしまうことがあった」、「笑い声が聞こえると楽しく感じる」などが得られた。

一方、ネガティブな意見として、「少し笑い声が多すぎる」、「(大笑) などの字幕

によってややしらけた」,「ロボット内で笑っていて取り残されてる感があった」,「笑い声がロボットの見た目とあっておらず,違和感を感じた」,といった意見が得られ,被験者によっては笑いの表現によって継続利用意欲が抑えられてしまう可能性が示された.

以上から,笑いの表現によってロボットに対し共感を感じるユーザが見られ,バラエティ番組のような笑いを喚起する内容のコンテンツと合わせて用いることで,ユーザとの親和性を高めることに貢献する可能性を示した.また,笑いの表現自体に違和感を感じるユーザも見られ,笑い声のバリエーションやロボットの見た目をユーザの好みに合わせるなどといったカスタマイズ性が根本的に必要であることが示された.

以上のように盛り上がり・笑いの雰囲気における情動共有をTV雑談ロボットに適用した際の印象を主観評価により比較した結果,盛り上がり・笑いの両条件において情動共有をロボットに適用した場合,非適用時に比べユーザが雰囲気に対して変化する情動にロボットが追従し,ユーザから共感を得られる可能性が示唆された.盛り上がり表現の評価においては,情動共有非適用時よりも,適用した際のロボットの方が,ユーザが受ける盛り上がりの印象を強めることがわかった.情動共有におけるロボットの視差がユーザに対する盛り上がり・笑いにおいて感情を増幅する可能性が示唆された.一方でロボットが実体を持つことにより,ユーザがロボットに対して持つ期待値が上昇する可能性が示唆された.情動共有を評価する際は,ロボットのアピランスや声質を事前にユーザの好みにあったものに変更可能な状態に設定する枠組みの必要性も示された.

研究の限界として,今回はTV番組に対するSNSコメントを用いてTV雑談ロボットをどのように発話させるべきかについて調査を行ってきたが,今回実験で検証したジャンル以外にも調査すべき要素が多く存在する.例えば,雰囲気に基づく情動共有では,対象とする雰囲気として盛り上がりと笑いを取り上げた.しかしTV番組視聴時において生じる雰囲気には今回調査した2種類以外にも悲しみや恐怖,怒り,感動,といった細かな雰囲気に分解され,それらに合わせたロボットの振る舞いについても網羅的に実装・調査する必要がある.また,TV視聴時の雰囲気は特定の状態を維持し続けるのではなく,常に遷移を繰り返すもの

と考えられ、これに対応するためのシステム構築についても同時に行っていく必要がある。

4.2.5 盛り上がり雰囲気における情動共有の客観評価と考察

本章では 4.2.2 章で述べた情動共有手法について、生理指標に基づく客観評価による評価実験について説明する。本実験は主観評価のみで示唆された、ロボットの振る舞いがユーザの感情を増幅するという主張を支持することを目的として行われる。評価に使用した情動共有手法の変更点を以下にまとめる。

- 雰囲気度を推定するための特徴量として、盛り上がり判定区間から 15 分前までのコメント数の平均 μ と標準偏差 ρ を採用せず、判定したい 5 秒区間のコメント数 C_i と直近 30 分間のコメント数の平均 μ に変更：実験に使用する映像に対して、ground truth として用いた人間の盛り上がりレベルの時間変化により近い形で追従できるため。
- 盛り上がりレベルに応じてロボットの発話頻度と韻律情報（話速、ピッチ、ボリューム）を動的に変更：人同士のサッカー中継の番組視聴時の様子を観察した際、動的に変動していたパラメータであったため。
- ロボットを pip 製のうなずきかぼちゃんから、SHARP 製の人型ロボット RoBoHoN[82] に変更：予め設定された腕振りなどのモーションがうなずきかぼちゃんよりもより自然で豊富。モーションと動作に発話が可能であることに加え、モビリティに優れていたため。
- ロボットの最大数を 4 体に変更：情動共有の効果を評価するに当たり、ロボットの盛況感を演出するために十分であると判断したため。また、4 体で十分な効果があることが確認できた場合、それ以上の数のロボットを用いた場合も同様の効果が期待できるため。

なお本章で述べる実験では人型であり、実体を持つロボットのみを用いて行った。新たに実装した SNS 解析に基づく TV 番組の盛り上がりレベル E をリアルタイム

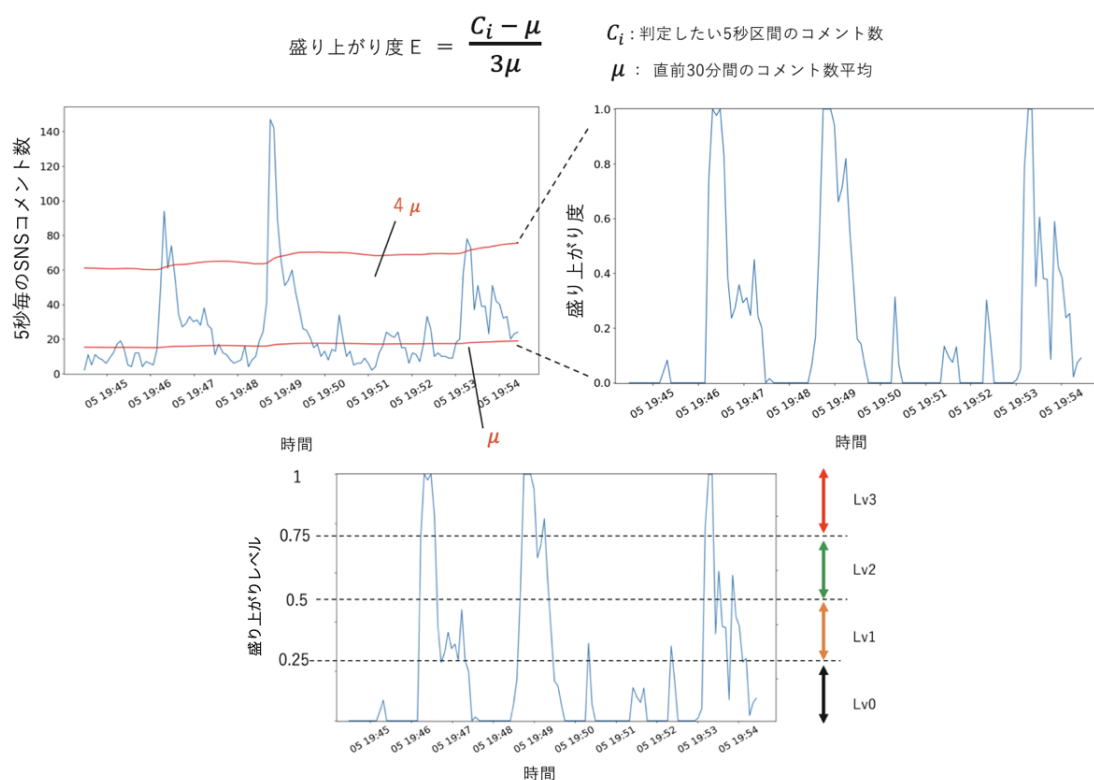


図 38 SNS コメント数の推移とモデルの出力例 (上図), 4 段階に分類された TV 番組 (サッカー中継) 中の盛り上がりレベルの例 (下図)

μに検出する推定モデルを示す.

$$E = \begin{cases} 1 & (4\mu \geq \frac{C_i - \mu}{3\mu} : \text{more than max threshold}) \\ \frac{C_i - \mu}{3\mu} & (\text{other}) \\ 0 & (\frac{C_i - \mu}{3\mu} \leq \mu : \text{less than min threshold}) \end{cases} \quad (14)$$

TV 雑談エンジンを用いて取得されたニコニコ実況 [65] および Twitter[55] 上のコメントをロボットの発話文および盛り上がりレベル推定のための入力として用いる. このモデルで出力される盛り上がりレベルは μ を下限, 4μ を上限とした比率で表現される. 盛り上がりレベルは図 38 に示されるように 4 段階に分類され, 各段階に応じてロボットの振る舞いを変化させる. 想定するレベルごとのロボットの振る舞いのイメージと voice を図 39 に示す. 表 11 は, サッカー中継 (FIFA ワールドカップ 日本 vs ポーランド, 2018 年 6 月 18 日放送) を共に視聴した 23

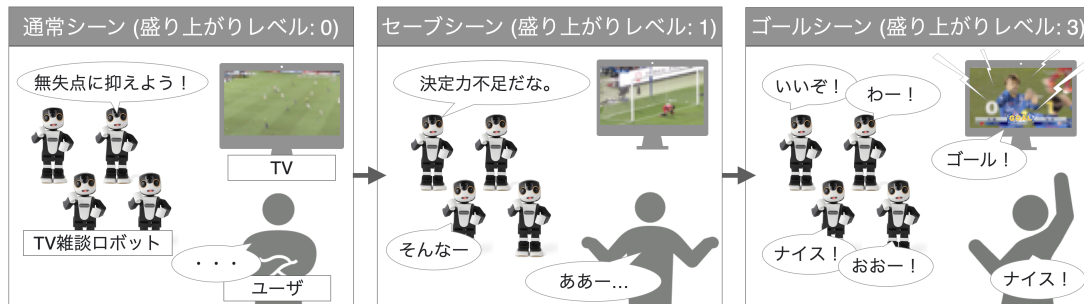


図 39 TV 雑談ロボットとユーザ間のインタラクションの例

から 30 歳の 5 名の男性を観察した際の盛り上がり時、非盛り上がり時に測定された韻律情報に基づく、盛り上がりレベルごとの韻律特徴を示す。発話間隔は発話間の無音声区間の長さの平均値、話速はモーラ数 [83] を発話間隔の長さで割って得られる 1 秒あたりの平均モーラ数、ピッチは発話音声から抽出される基本周波数 (f_0) を対数領域で平均化した値、ボリュームは発声区間のパワースペクトルの平均値から取得された。発話間隔以外の特徴については、上述した値からロボットの合成音声を作成するツールである VoiceText[59] のパラメータに入力する値に変換された値が表 11 中に示される。レベル 1 と 2 の値は、レベル 0 と 3 の間の線形補間によって取得された。

今回提案する盛り上がりレベル推定モデルおよび、モデルを元に情動共有を行うロボットの印象を評価する実験を行った。モデルの精度を評価するために、ベースラインとしてマニュアルで作成された TV 番組により引き起こされる盛り上がりレベルとの比較を行った。TV 番組には 2019 年 9 月 5 日に放送されたサッカー日本代表 vs パラグアイ戦の一部を使用し、ground truth となるデータは、23 から

表 11 盛り上がりレベルと対応するロボットの発話時の韻律特徴

	発話間隔 (秒)	ピッチ	ボリューム	話速
レベル 3: 盛り上がり時	4	123	132	100
レベル 2	7	117	118	
レベル 1	10	111	105	
レベル 0: 非盛り上がり時	13	105	92	

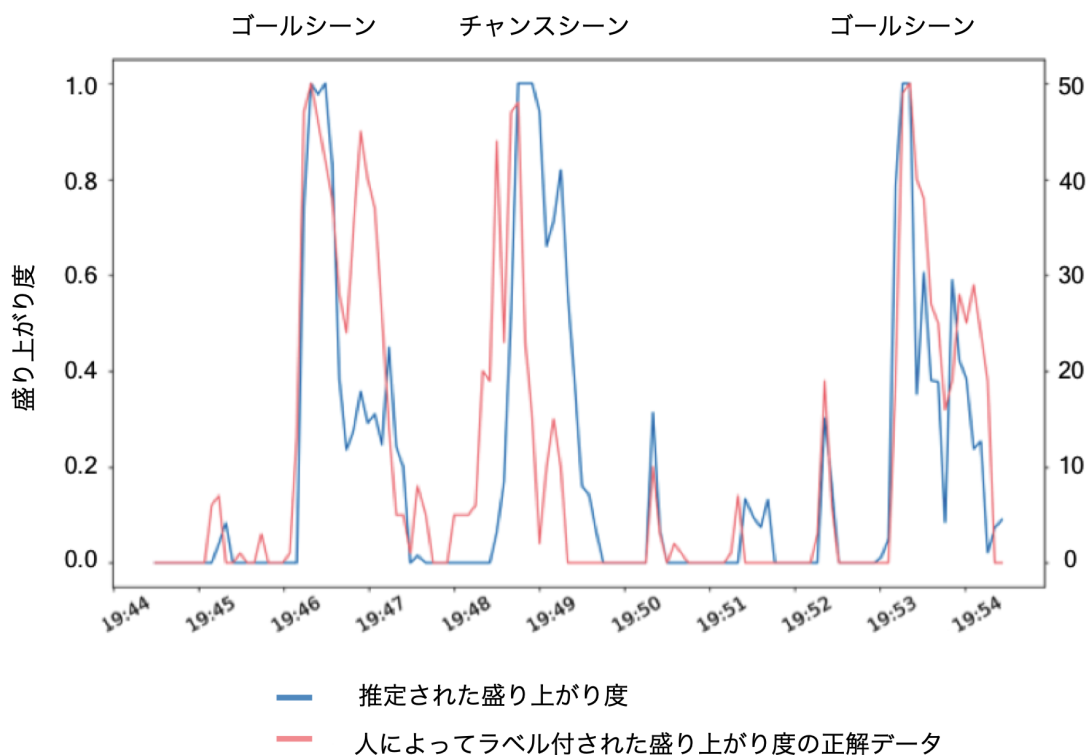


図 40 提案モデルにより推定された盛り上がり度 (青線) と 10 秒間遅延させた ground truth の推移 (赤線)

30歳の男性10人により、作成された。毎秒、TVシーンに対して盛り上がったと感じたら1を、そうでなければ0を入力するよう指示をした。このアンケートにより1秒ごとのデータが10人分、つまり5秒間を1区間として区間あたり0から50のスコアで盛り上がりレベルを表す ground truth を作成した。作成されたデータの傾向として、ゴールなどの決定的なシーンだけでなく、チャンスに繋がりそうなパス回しの段階から盛り上がりと判定されやすいことがわかった。これに比べ、SNS上では盛り上がり一連のプレイが終了してからコメントを投稿される傾向があるため、盛り上がりの判定タイミングにずれが生じると考えられる。そこで、提案する盛り上がりレベルの推定結果を10秒間シフトした際のデータを比較した結果を図40に示す。2つのデータ間には、複数の箇所でも盛り上がり発生タイミングの一致が確認でき、相関係数は0.714であった。以上より、

提案する盛り上がりレベルの推定モデルはリアルタイムに番組を視聴する環境に比べて、10秒遅れで盛り上がりを推定できることが示された。

次に、提案するモデルを適用されたTV雑談ロボットが、人に与える印象や感情増幅の可否について評価を行う。評価には、ロボットの印象について問うアンケートおよびアフェクトグリッド [84] による主観評価と、対話中の覚醒度を測定するためのスキンコンダクタンス、感情価を測定するための大頬骨筋と皺眉筋の筋電位生理指標による客観評価を用いる。今回は被験者が映像のみを視聴するロボット無し条件 (No agent)、韻律のパラメータが動的に変化しない4体のロボットと映像を視聴する条件 (Neutral agents)、情動共有によって動的に振る舞いを変える提案手法を適用した4体のロボットと映像を視聴する条件 (Excited agents) の3条件について比較を行い、覚醒度およびユーザの利用意欲を評価する。ロボットにはSHARP製のRoBoHoN[82]を使用し、発話時は両腕を頭の高さまで同時に振り上げるモーション、両腕を前後に振って頷くモーション、右腕を頭の高さまで振り上げ頭をかしげるモーション、モーション無しの4つの中から発話ごとにランダムに一つ選ばれるように設定した。実験時の実装として、Excited agents条件では提案する盛り上がりレベル推定手法をリアルタイムに動作させた際の評価を行うため、10秒間のシフトはせずに行う。実験の仮説として、No agent, Neutral agents, Excited agentsの順に評価値が高くなると予想した。実験に使用するTV番組は全てサッカー中継とし、ひいきのチームが被験者にとって偏らないよう日本代表の出場するのワールドカップ戦を選んだ。1つの映像につき1.5分から2分間で、かつチャンスシーン、ファインプレー、ゴールシーンといった盛り上がるタイミングが明確であることが事前に判断できる内容を含む映像を12本用意した。実験に参加する被験者は、番組自体への興味度が印象評価に影響する可能性があるため、日本人でありかつスポーツ(サッカー)に興味があるという条件を満たす20代の男女24名を対象として実験を行なった。実験は防音室内で行い、被験者は生理指標計測装置をつけた状態で録画したサッカー中継をロボットともに視聴した。実験の様子を図41に示す。映像の視聴回数は被験者1名に対して、3回(各条件1回ずつ)を1セットとして、4セットつまり12回行われた。映像及び条件の順番は被験者間でランダムとし、カウンターバランスをとった。各映像の



図 41 防音室内での実験の様子

視聴後、被験者はアフェクトグリッドに感情のプロットと、以下の質問に回答する：「エージェントは盛り上がっていたか」「エージェントに共感できたか」。さらに、各セット終了後にまた利用したいと思った条件を1つ選んだ。

3条件について、アフェクトグリッドによる覚醒度と感情価の評価結果を図42に示す。結果より、覚醒度のみがNo agentとExcited agents間で有意差が認められた ($t(23) = 2.99, p < 0.005$)。

図43に示すように生理指標による客観評価の結果、覚醒度を測定するためのスキンコンダクタンスレベルのみにおいて、No agentとExcited agents間で、有意差が認められた ($t(22) = 1.65, p < 0.05$)。

図44は、Neutral agentsおよびExcited agents間の差に関して、アンケートにより評価されたロボットに対して感じた覚醒度と感情価に関する結果を示す。結果から、覚醒度のみが2つの条件間で有意差が認められ ($t(23) = 4.84, p < 0.001$)、被験者から見てExcited agentsがNeutralよりも興奮していると認識されたことを示した。

各セッションの終了時に回答された、被験者が選ぶ再び利用したい条件を比較

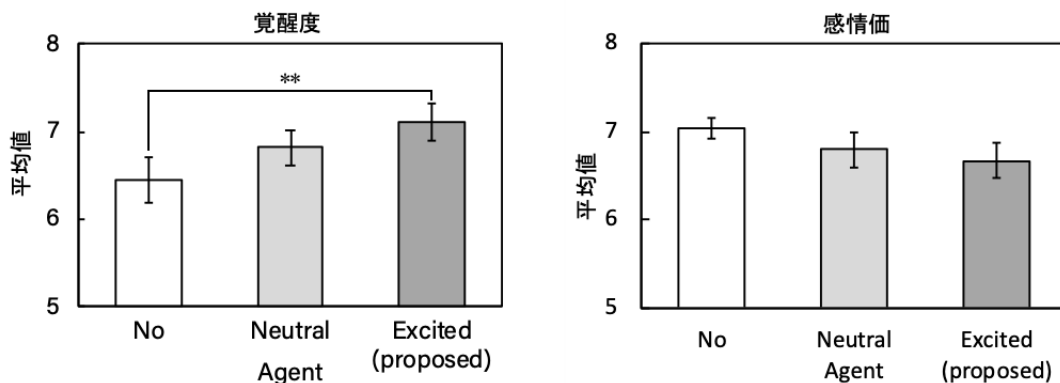


図 42 アフェクトグリッドにより評価された被験者の覚醒度（左図）と感情価（右図）の平均値と標準偏差 (**, $p < 0.005$)

した結果を図 45 に示す。No agent と Excited agents の間 ($t(23) = 2.98, p < 0.005$), および No agent と Neutral agents の間 ($t(23) = 2.26, p < 0.05$) にそれぞれ有意差が認められた。これらの結果から、独りではなく TV 雑談ロボットと共に TV コンテンツを視聴すること、つまり提案する TV 雑談ロボットに対する利用意欲が促進されたことが示唆された。

実験を通して、今回実装した盛り上がり雰囲気における情動共有手法が適用された TV 雑談ロボットが、サッカー中継という TV 番組においてユーザに与える効果を主観的・客観的に評価した結果、提案手法がロボットを使用しない条件および提案手法を適用しないロボットと比較して、ユーザの覚醒度を有意に高めることを示した。しかし、条件間で感情価においては有意差は認められなかった。

研究の限界として、実験によって得られた結果が、盛り上がりレベルの推定に必要な 10 秒の遅延を解決した場合を考慮しない条件で得られたものであるという点である。この遅延を含めた上でロボットがユーザと対話を行った場合、ユーザがロボットから受け取る印象に影響を与える可能性があるため、遅延を解決する方法が必要になると考えられる。遅延を取り除くために 3.2 にて提案した手法である、テレビ番組のタイムシフト再生機能を介してエージェントの発話タイミングや興奮のタイミングを制御することが挙げられる。あるいは、事前にエージェントに状況に適した固定フレーズを与えることも挙げられる。今回の実験に用い

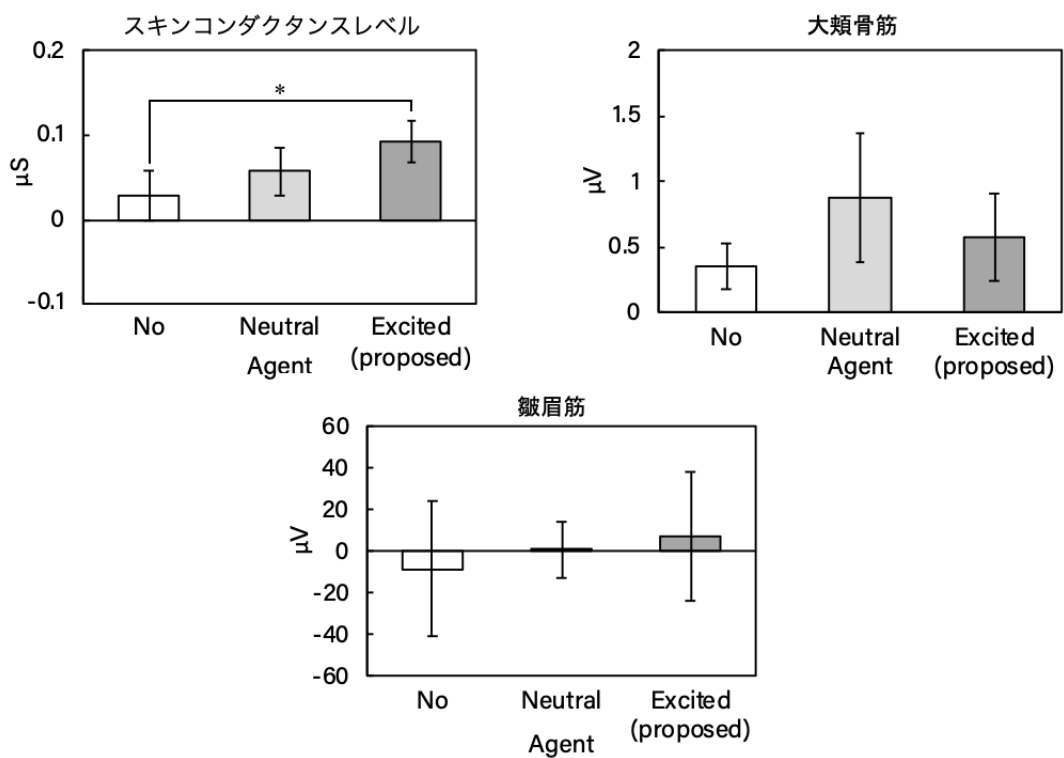


図 43 生理学的反応の平均値と標準偏差：スキンコンダクタンスレベル（左上図），大頬骨筋（右上図），および皺眉筋（下図）の筋電位 (*, $p < 0.05$)

たサッカー中継のように，味方チームの得点など，予め予想される状況が検出された際に SNS 上でコメントされやすいキーワードをロボットに発話させることで，SNS 解析による盛り上がりレベルの推定結果が出力される前に，ロボットに発話させることができる。

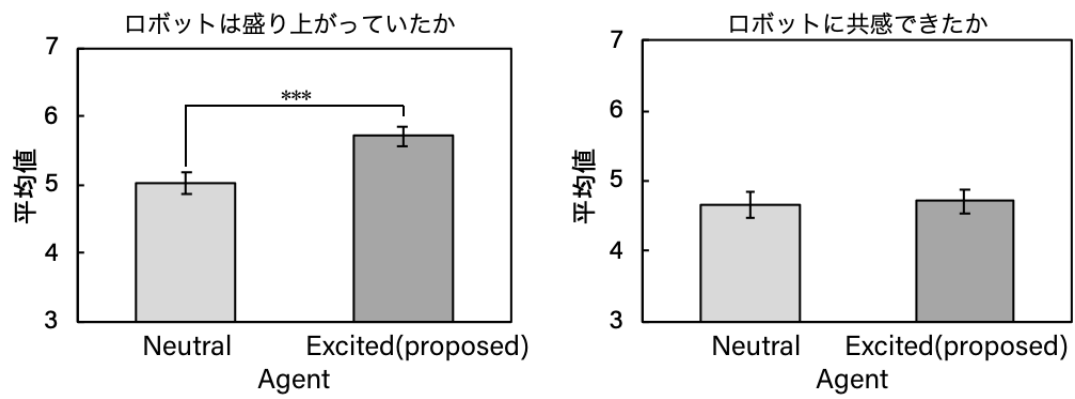


図 44 被験者がロボットに対して盛り上がっていると感じた度合い (左図) と、ロボットが共感してくれていると感じた度合いの平均値と標準偏差 (右図) (***, $p < 0.001$)

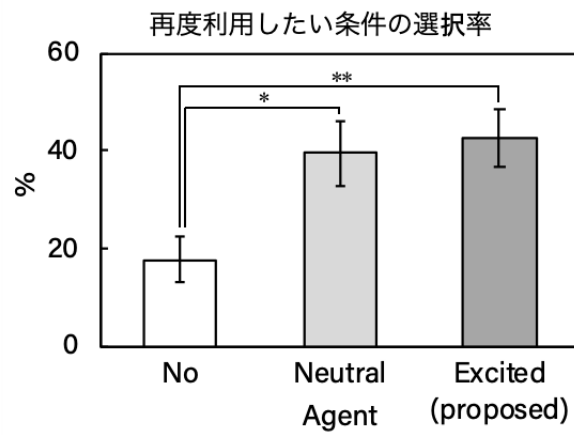


図 45 各実験条件下での被験者の再選択率の平均値と標準偏差 (**, $p < 0.01$; *, $p < 0.05$)

4.3 韻律模倣

4.3.1 非言語コミュニケーションにおける同調対話に関する先行研究

非言語コミュニケーションの同調には成人の対話者同士においてラポールを構築する役割があると考えられている [85], [86]. また, 非言語・パラ言語情報の同調を意図的に使用すると, 話者の共感に関する評価を向上することができる [87]. この共感を人間同士のインタラクションにおいて表現することで, 互いに親密な関係を構築することに繋がると考えられている. [88].

この非言語・パラ言語情報における同調をロボットに適用した例によると, 人とロボットのインタラクションにおいても人同士のインタラクションと同様の効果が得られる可能性が示唆されている [89, 90]. ここで韻律情報においても同様の効果が期待される [91, 92, 93] が, 韻律の同調が人間の感情変化に与える影響に関する科学的証拠に基づく特定の効果を検証した例はない. 従って, 本節はロボットが韻律的な同調対話によって対話者に共感性を感じさせ, ラポールを構築することに貢献することができるか, という疑問を科学的に調査することを目的とする. 対話者の話し方から韻律情報を推定し, 同様の韻律情報によって返答を行うロボットを用いることで, 人同士の韻律模倣を観察するといった外乱の多く発生する環境と比較して再現性の高い実験環境の構築を行った. このような実験環境下において, 対話ロボットに対する対話者の印象を主観的および客観的に評価する.

4.3.2 韻律模倣の実装

韻律情報には話速 [94], ピッチ [95], 音量 [95], アクセント [96], [97], 発話長 [98], 応答時間 [99], 間の周期と長さ [98], [99], 笑い [98] といった要素が含まれる. これらの特徴の中から話速, ピッチ, 音量の3つを対話者の発話から推定し, 模倣するシステムを開発する. このシステムは動的に実行され, 人が発話を行なうとその韻律に合わせた合成音声生成され, 直後のロボットの発話音声として出力される. 本システムの構成にあたって, 大語彙連続音声認識エンジンである Julius [52], 音声信号処理ツールである SPTK [100], 音声合成ソフトウェアであ

る VoiceText [59] の3つを用いた。このシステムによって、人の発話を認識してからロボットの音声発話を生成するまでの過程を以下に示す。

- (1) 人の発話区間を Julius によって認識・録音する。
- (2) 話速を Julius によって推定する。
- (3) ピッチ・音量は SPTK を用いて推定する。
- (4) 各推定値が VoiceText に特有の話速・ピッチ・音量に関するパラメータに変換され、合成音声作成のリクエストを送信する。

話速、ピッチ、および音量の推定方法の詳細を以下に説明します。

話速を表す特徴量として、音声認識処理の前処理で得られる発話区間長でモーラ数を除算して得られる1秒あたりの平均モーラ数を用いた。モーラとは、日本語における「拍」であり、各モーラは時間的に等価である。基本的には「(子音+)母音」の構造を持つが、例外として、母音を持たない撥音「ん」や、音そのものにかかる促音「っ」がある [83]。ピッチと音量を推定する SPTK では、ピッチの推定値を得るためにまず、Julius によって取得された人の発話の wave ファイルから基本周波数 (F_0) を抽出し、 F_0 の対数領域の平均値を計算することで平均ピッチが推定される。そこから10ミリ秒毎に分析窓をシフトして F_0 を出力する。ここから得られた F_0 値から無声音区間を取り除いた F_0' の平均値を人のピッチとして推定する。音量の推定対話者の発話音量推定にはピッチ推定と同様、SPTK を用いる。ここでは簡単のために音声区間のパワー項の平均値を用い、これを人の音量として推定する。一方、音声合成は、VoiceText による男性と女性の両方の音声モデルを使用して実行される。各性別の音声合成によって出力される音声は、推定された韻律情報に基づいて VoiceText の話速・ピッチ・音量に関するパラメータに校正されます。

この校正に関して、推定された韻律情報に対して VoiceText の合成音声は客観的に一致している必要がある。推定した人の韻律情報から VoiceText の音声合成パラメータへの変換式を求める作業を行う。作業内容を以下で説明する。音声入力にはソニー株式会社のエレクトリックコンデンサーマイク ECM-674/9X を用

いた。音声出力には対話ロボットとして後述する印象評価実験でも用いるヒューマノイドロボット NAO v4 [101] の内蔵スピーカを用いた。なお、VoiceText の音声合成パラメータの値は話速 50~400(%), ピッチ, 音量 50~200(%) に制限されている。VoiceText を用いて合成音声を生成するためにはテキスト情報が必要である。この発話文章には以下に示す 5 つの特定の日本語文章を用いた。

- あらゆる現実をすべて自分のほうへねじまげたのだ。
- テレビゲームやパソコンでゲームをして遊ぶ。
- 救急車が十分に動けず救助作業が遅れている。
- 老人ホームの場合は健康器具やひざ掛けだ。
- 嬉しいはずがゆっくり寝てもいられない。

これら 5 種類の文章は様々な音素をバランスよく含んでおり [102], [103], 様々な発話に対応する必要がある本システムに適していると考えた。変換式の導出では、今回用いる音声合成パラメータである話速, ピッチ, 音量においてそれぞれ行う。以下にその導出の手順を具体的に示す。

- (i) VoiceText の該当する音声合成パラメータを 50 (話速は 70) に設定し, 合成音声を出力する。このとき変換式を導出するパラメータ以外は 100 に固定する。
- (ii) Nao のスピーカから出力された合成音声をマイクで入力し, その時の韻律情報を推定する。このとき Nao の音量設定は 65% で固定された。
- (iii) 推定した韻律情報とその時の音声合成パラメータを記録する。
- (iv) 1 から 3 を既に示した 5 種類の文章それぞれにおいて 3 回ずつ行う。
- (v) 該当する音声合成パラメータを 10 変化させ, 1 から 4 の作業を行う。
- (vi) 1 から 5 をパラメータが 200 (話速は 150) になるまで行う。

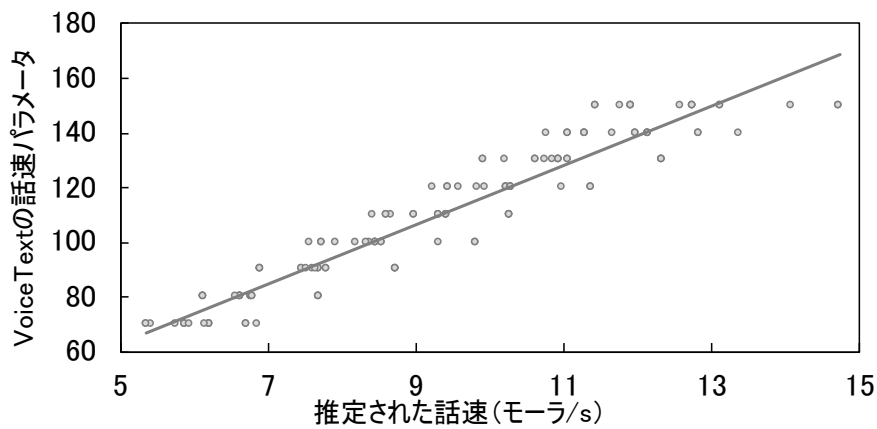


図 46 推定された話速の値と VoiceText の話速に関するパラメータの関係

ここから得られた音声合成パラメータと韻律情報の関係性を用いて変換式の導出を行う。

VoiceText が提供している音声モデルとして、男性用の音声モデルには “kenta” を、女性モデルには “hikari” を用いて音声合成を行う。

話速

図 46 に示されるように、推定された話速を x 軸，VoiceText の話速パラメータの値を y 軸にプロットした。話速の変換式は線形回帰を用いて式 15 に示される。

$$y = 10.843x + 8.7307 \quad (15)$$

決定係数 R^2 は 0.91 であった。話速については、男性・女性の音声モデルによる VoiceText パラメータへの影響がなかったため、同一の変換式を用いるものとした、

ピッチ

人間の知覚では音の高さは基本周波数の対数軸上で等間隔に知覚される。このため、推定したピッチには対数を取る処理を施している。よって、ピッチの校正

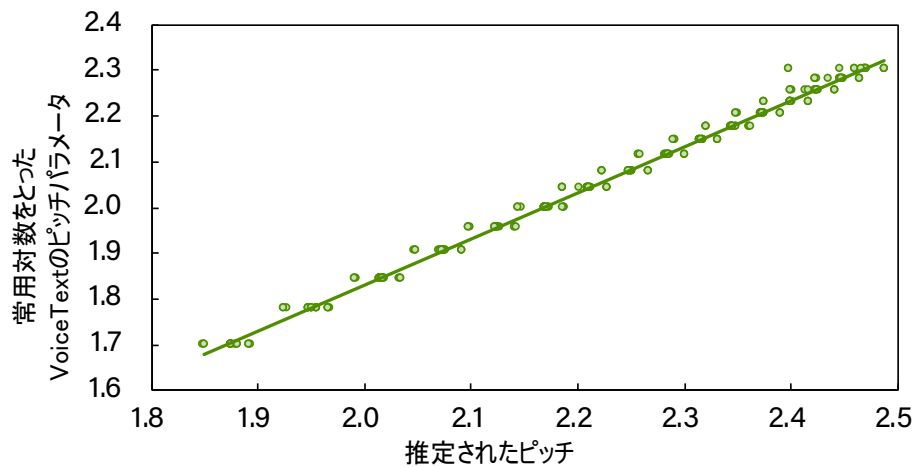


図 47 推定されたピッチの値と VoiceText のピッチに関するパラメータの関係（音声モデル: takeru）

式を求める際も、VoiceText のピッチパラメータに関して事前に常用対数を取った値を用いた。図 46, 46 に示されるように、推定されたピッチを x 軸，常用対数をとった VoiceText のピッチパラメータの値を y 軸にプロットした。ピッチの変換式は線形回帰を用いて式 16, 17 に示される。

$$y = 1.0065x - 0.1826 \quad (16)$$

$$y = 1.0123x - 0.4827 \quad (17)$$

男性モデルに用いる校正式 16 における R^2 は 0.99 であった。女性モデルに用いる校正式 17 では R^2 は 0.99 であった。音声合成を行う際には求めた値の真数を音声合成パラメータとして用いるものとする。

音量

音の高さと同様に人間の知覚では音量も対数化された上で等間隔に知覚されるため、推定した音量には対数を取る処理を施している。よって、音量の校正式を求める際も、VoiceText の音量パラメータに関して事前に常用対数を取った値を用いた。図 49, 50 に示されるように、推定されたピッチを x 軸，常用対数をとっ

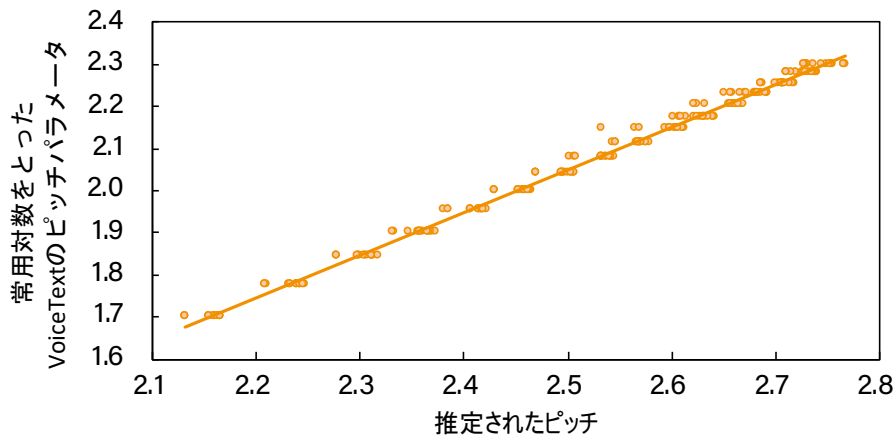


図 48 推定されたピッチの値と VoiceText のピッチに関するパラメータの関係（音声モデル: hikari）

た VoiceText の音量パラメータの値を y 軸にプロットした。音量の変換式は線形回帰を用いて式 18, 19 に示される。

$$y = 0.2472x - 2.9784 \quad (18)$$

$$y = 0.2565x - 3.2278 \quad (19)$$

男性モデルの式 18 の R^2 は 0.98, 女性モデルの式 18 の R^2 は 0.95 であった。

ここで、VoiceText の特性上、ピッチが変化することで、それに伴い音量が変化することが判明したため、ピッチの変化に伴い音量が変化する際の倍率を求める式の導出を行った。実際のペーシングシステムでは推定したピッチから求めた倍率で音量パラメータの値を割るような処理を加えた。式の導出作業としては、基準となる音量を固定し、ピッチを徐々に変化させることで、基準となる音量に対する算出された音量の倍率を求めた。そして、ピッチとこの倍率との関係性から倍率を求める式を非線形回帰を用いて以下の式 20, 21 ように導出した。

$$y = 0.0000001x^3 - 0.0000517x^2 + 0.0121527x + 0.2149448 \quad (20)$$

$$y = 0.0000000003x^5 - 0.0000002x^4 + 0.0000542x^3 - 0.0059065x^2 + 0.3058218x - 5.2385902 \quad (21)$$

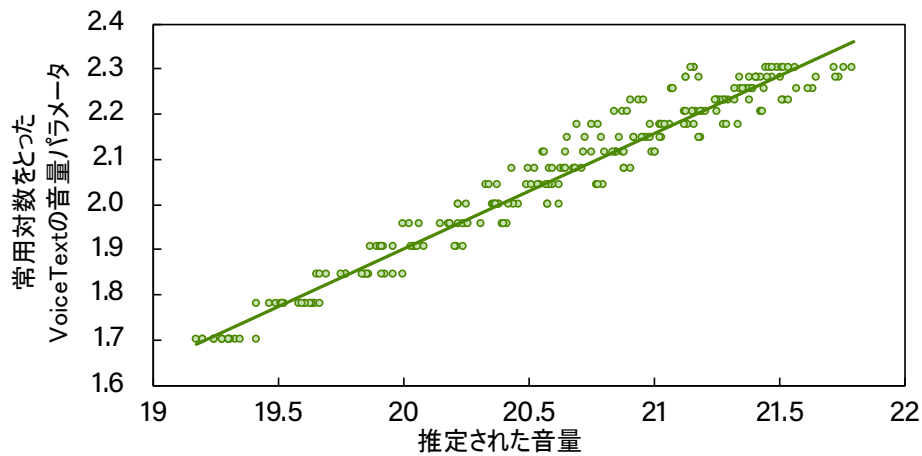


図 49 推定された音量の値と VoiceText の音量に関するパラメータの関係（音声モデル: takeru）

男性モデルの式 20 の R^2 は 0.99，女性モデルの式 21 の R^2 は 0.99 であった。

ここで作成した各変換式を用いて，人の話速・ピッチ・音量の3つの韻律特徴を模倣する音声合成を作成し，それぞれの特徴を反映できているかを検証するための評価実験を，男性モデルにおいて行った．評価を行うにあたって，以下の手順で人と合成音声の音声サンプルのセットを作成した．

1. 一人の特定の人物によって7種類の韻律的な特徴を (fast, slow, high, low, big, small, neutral) 持つ音声サンプルを作成する．このときの発話内容は「あらゆる現実をすべて自分のほうへねじまげたのだ」とした．
2. 各音声から，変換式を用いて音声合成パラメータを取得する．発話内容は人の音声と同じとする．
3. 上記の試行を5回ずつ繰り返し，音声サンプルごとに平均を取ったものを各音声における人の韻律情報とする．
4. 人の韻律情報から得られた3つの推定値を音声合成パラメータとして，VoiceText を用いてロボットの合成音声を7種類生成する．

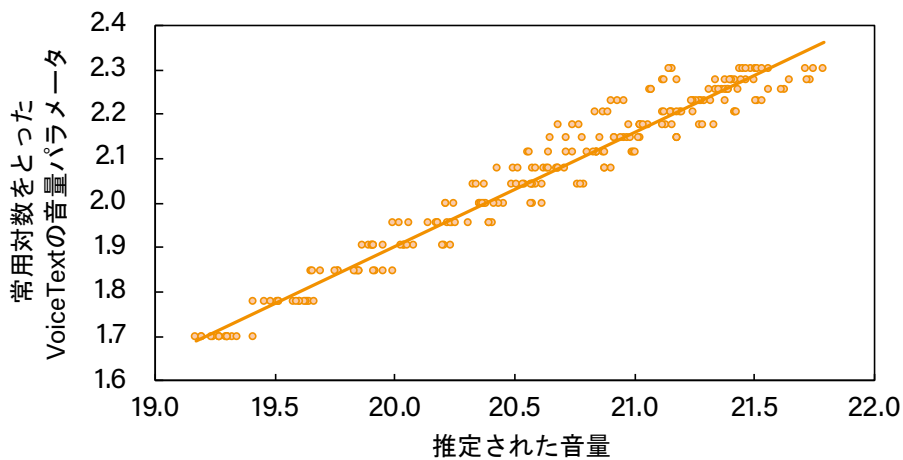


図 50 推定された音量の値と VoiceText の音量に関するパラメータの関係（音声モデル: hikari)

5. 人の音声と同様に，この各合成音声からもペーシングシステムを用いて音声合成パラメータを取得する。
6. 人の音声と同様に，5回の試行の平均値を取得し，ロボットの韻律情報とする。

人の音声と合成音声のペアを7セット用意し，主観・客観の両側面から評価を行った。客観評価では，推定された韻律情報の値の類似度を音声ペアごとに求め，比較する。

5回推定された各音声の話速・ピッチ・音量の平均値・分散および，各ペアにおけるコサイン類似度と相関係数を図 12 に示す。結果として，全てのサンプルにおいて十分に高いコサイン類似度，相関係数を示した。

主観評価では第三者に，各音声ペアを聴き比べた際の印象を調査する質問を7段階で評価してもらい，そのスコアを比較した。結果を図 51 に示す。以上から，提案するシステムの妥当性を示した。

表 12 韻律同調システム評価のための客観データ

サンプル	話者	話速	ピッチ	音量	コサイン類似度	相関係数
Fast	人	124.24	98.16	101.02	0.9999	0.9954
	ロボット	125.00	98.05	103.59		
slow	人	83.43	98.74	100.53	0.9997	0.9944
	ロボット	81.11	98.93	103.68		
high	人	98.53	123.85	102.76	0.9998	0.9799
	ロボット	97.30	123.96	106.77		
low	人	102.03	79.13	98.75	0.9999	0.9969
	ロボット	98.53	78.95	97.33		
large	人	101.58	98.78	124.22	0.9994	0.9905
	ロボット	97.71	99.06	129.51		
small	人	101.58	98.81	73.82	0.9998	0.9936
	ロボット	98.53	99.14	75.46		
neutral	人	97.71	99.33	99.97	0.9999	0.8868
	ロボット	97.71	99.03	102.18		

4.3.3 評価方法

話者の声の調子に合わせて対話を行うロボットが話者に与える影響を主観的・客観的に評価する。評価指標として主観的評価にはアンケートを、客観的評価には皮膚筋電位・大頬骨筋・皺眉筋を用いる。被験者にロボットと対話を行う際に使用する対話シナリオを14個独自に作成した。これは雑談ではなく決められたシナリオを用いて対話を行うことで、実験の再現性を保つためである。各対話は図52に示す例のように、ロボットと人が交互に7回のターンテイキングを行う構造であり、シナリオを読んだ際に人がポジティブな印象を抱くよう設計された。シナリオが読み手に与える印象を評価するために、アフェクトグリッドを用いた実験を行う[84]。アフェクト・グリッドとは、図4.2に示すように感情の地図のようなものである。図の真ん中は、感情的にneutral（中性）、つまりポジティブでもなくネガティブでもなく平常的な日常の感情を表す。図の右半分は、快感情を

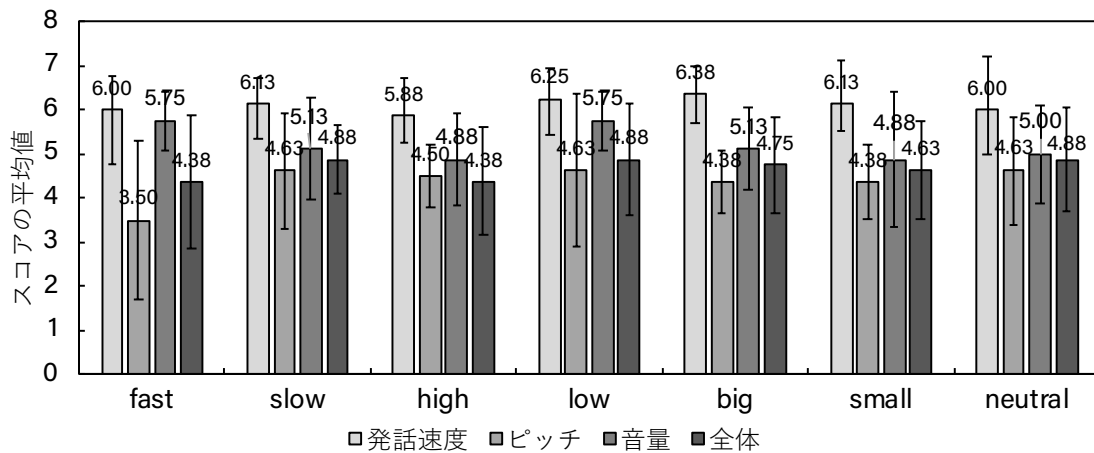


図 51 韻律同調システム評価のためのアンケートスコア

表し、右に行くほど快となる。逆に左半分は不快感情を表す。上下は覚醒の程度を表す。ポジティブ、ネガティブとは独立に、どれだけ覚醒し活性化していると感じるかを示す。上に行くほど覚醒的となり、下の方は眠気が強いということになる。右上は、エキサイト、喜びといった感情を表す。逆に左下は、抑うつ、悲しみ、憂うつといった感情を表す。左上は、ストレス、緊張を表す。逆に右)下は、静穏、リラックスを表す。

我々はポジティブな印象はアフェクトグリッド上の快かつ覚醒を示す範囲に位置すると考え、作成したシナリオがどの点に位置するかを評価する。20代の男性7名にロボットと人の対話シナリオを読んでもらい、対話終了後に図53に示すアフェクトグリッド上に感情をプロットさせた。アフェクトグリッドによる各シナリオの評価値の平均値を図54に示す。アフェクトグリッドのスコアを、快感を示す横軸において中央の値を0、軸の右側をプラス、反対がマイナスとなるよう1刻みで変化するものと定義した(最高値4、最低値-4)。同様に、覚醒度を示す縦軸においても中央の値を0、軸の上側をプラス、反対がマイナスとなるよう1刻みで変化するものと定義した。結果より、用意した14種類の対話シナリオのうち12種類において、覚醒-眠気の軸においてプラスの評価を示した。また、この12種類のシナリオにおいては、非常に快-非常に不快の軸においてもプラスの評価を示した。

Nao : こんにちは!

被験者 : こんにちは!

Nao : 好きな食べ物は何ですか?

被験者 : 私, ラーメンがすごい好きなんですよ.

Nao : 私もラーメンは大好きです! 何ラーメンが好きなんですか?

被験者 : 醤油ラーメンが好きです.

Nao : 私も醤油ラーメンはよく食べに行きますよ!

被験者 : え, 今度一緒に食べに行きませんか?

Nao : 本当ですか!

私も誰かと行きたいと思っていたところなんですよ!

被験者 : 実は, 美味しいお店を知ってるんですよ (内緒話みたいに)

Nao : どこにあるんですか? 教えてください.

被験者 : 大阪にあるんで, 今度紹介しますね!

Nao : ありがとうございます! 楽しみにしています!

被験者 : はい, 私も楽しみです!

図 52 シナリオの例.

皮膚筋電位から, 人の覚醒の程度を測ることができる. また, 大頬骨筋・皺眉筋からは人が感じる快感情の程度が示される. これらの指標を用いることによって, 覚醒度と快感情を喚起するシナリオをロボットと対話した際, 被験者の覚醒度と快感情がそれぞれどのように変化するのか調査することができる.

これらのシナリオを用いて人の韻律を模倣するロボットの印象を評価する実験の構成について詳細を説明する. 本実験の目的は対話者の話し方に同調するロボットがユーザに与える印象を主観・客観的に調査することである. 実験に際し以下の4つの仮説を設定した.

H1 対話者は自分の話し方に同調するロボットに対して, そうでない場合と比

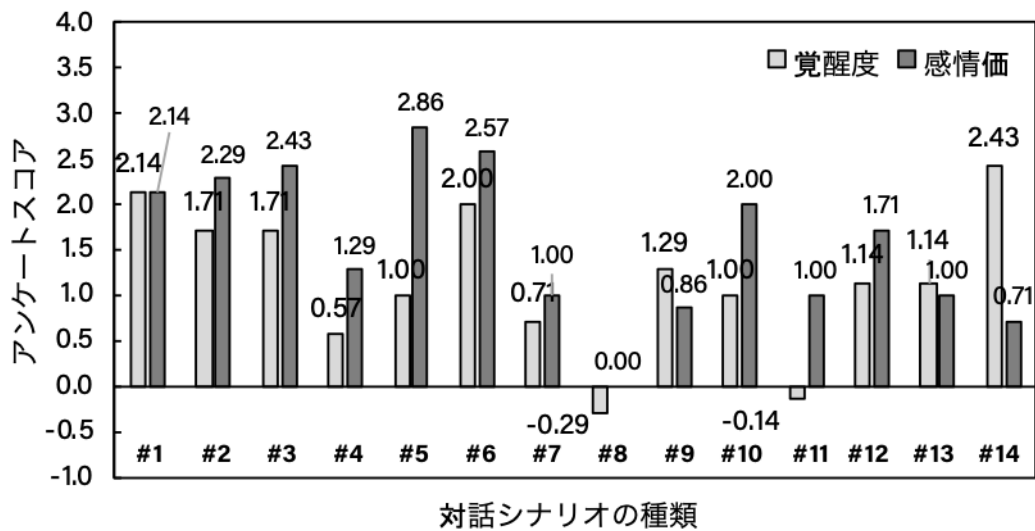


図 54 シナリオの評価結果

被験者は以下の手順で各条件下のロボットと対話を行う。まず被験者は事前に評価された14のシナリオを読むことでロボットと対話を行うことを説明される。これらのシナリオに書かれた注意事項に従って、シナリオを読むよう導入が行われる。図 55 に示すように、被験者がポジティブかつ活性な発話をするべき箇所を赤字で、反対にネガティブかつ不活性な発話をするべき箇所に青字で恣意的にシナリオにセリフを表記した。快かつ活性の感情を喚起する12のシナリオはランダムに3つのセットに分けられ、計4セットの対話においてロボットの印象を評価する。4つのシナリオセットのそれぞれにおいて、ロボットの同調あり・なしの条件を被験者によってランダムになるよう対話を行う。そのため、被験者は用意された4つのシナリオセットをそれぞれ2回ずつ読む。14のシナリオのうち印象評価には使われない2つのシナリオを用いて、被験者の慣らしを行う。慣らしでは同調あり条件のロボットを用いる。ロボットとの対話を開始する前に、被験者に左手と顔に電極を装着する。これにより、ロボットの印象を客観的に評価するために用いる皮膚筋電位・大頬骨筋電位・皺眉筋電位を取得する。被験者はロボットとの対話が終わるたびに、以下の7つのアンケートを7段階の尺度で回答してもらう。

Nao : こんにちは!

被験者 : こんにちは!

Nao : 好きな食べ物は何ですか?

被験者 : 私, ラーメンがすごい好きなんですよ.

Nao : 私もラーメンは大好きです! 何ラーメンが好きなんですか?

被験者 : 醤油ラーメンが好きです.

Nao : 私も醤油ラーメンはよく食べに行きますよ!

被験者 : え, 今度一緒に食べに行きませんか?

Nao : 本当ですか!

私も誰かと行きたいと思っていたところなんですよ!

被験者 : 実は, 美味しいお店を知ってるんですよ (内緒話みたいに)

Nao : どこにあるんですか? 教えてください.

被験者 : 大阪にあるんで, 今度紹介しますね!

Nao : ありがとうございます! 楽しみにしています!

被験者 : はい, 私も楽しみです!

図 55 シナリオの読み方に対する指示の表記例

- Q1 ロボットは親しみやすかったですか?
- Q2 ロボットと話していて楽しかったですか?
- Q3 ロボットとどれくらい繋がりを感じましたか?
- Q4 ロボットはどれくらい自分に同調していましたか?
- Q5 ロボットを継続的に利用したいと感じましたか?
- Q6 ロボットはあなたの話を聞いている感じがしましたか?
- Q7 2回の対話のうちどちらのロボットが好きでしたか?

Q1 から Q6 では、7 段階のリッカート尺度（1：まったくそう思わない、7：まったくそう思う）によってスコアをつけてもらうものとした。質問 7 に関しては、同一対話シナリオにおける 2 回のうち、どちらのロボットが好きであったか選択してもらった。そのため、1 回目の対話では回答せず、2 回目の対話後のみ回答してもらった。

Q1 に関しては、柿井 [104] による関連研究において、人同士の信頼関係構築を示す質問として利用されていたものを用いた。Q2, 3, 4 に関しては Tarr ら [105] によるダンスにおける同調が社会的結合に与える影響について調査した研究において利用していた質問を利用した。Q5 は H2 を評価するための使用する。Q6 は音声の同期が信頼関係の構築に寄与することを示す指標として使用される。これは、対話エージェントが積極的なリスニング行動を示すことが話し手との信頼感を生み出すという知見に基づいている [56, 106]。Q7 は H4 を評価するために使用する。質問の他に、被験者にはロボットとの対話が終わるたびに自由記述を行ってもらった。

4.3.4 評価結果

図 56 に Q1 から Q7 に対する同調あり・なし条件におけるスコアの平均値を示す。t 検定（両側）により、各質問において同調あり条件が同調なし条件と比較して有意に高いスコアを示した。各質問における効果量を表 13 に示す。

図 57 は、Q7 で設定された各シナリオにおける同調あり条件を好む回答者の割合を示す。結果から、同調あり条件が好きと答えた割合も、しない場合が選ばれた割合に比べ多い結果となった。

皮膚筋電位、大頬骨筋、皺眉筋による客観評価の結果として、図 58 に同調あり・なし条件における平均値のグループを示す。また、効果量について表 14 に示す。t 検定（片側）により、同調あり条件下で皮膚筋電位、大頬骨筋がともに同調なしに比べ有意に高い結果となった。同様に、t 検定（片側）により、同調あり条件下で皺眉筋が同調なしに比べ有意に低い結果となった。

次に、男性と女性の間でロボットの印象差に関する結果を示す。12 名の男性と 15 名の女性のそれぞれにおいて Q1 から Q6 へアンケートスコアの平均値を図

表 13 各質問に対する対応ありの数値的結果と効果量

	t 値	p 値	d 値
Q1	-4.93	4.1×10^{-5}	1.85
Q2	-6.56	5.9×10^{-7}	1.81
Q3	-5.28	1.6×10^{-5}	1.83
Q4	-5.13	2.4×10^{-5}	2.06
Q5	-5.16	2.2×10^{-5}	1.50
Q6	-5.09	2.7×10^{-5}	1.87
Q7	-5.93	2.9×10^{-6}	3.29

表 14 各生理指標における対応ありの t 検定の数値的結果と効果量

	t 値	p 値	d 値
感情価 皺眉筋	2.02	0.027	0.54
大頬骨筋	-1.73	0.047	0.28
覚醒 スキンコンダクタンスレベル	-1.72	0.049	0.48

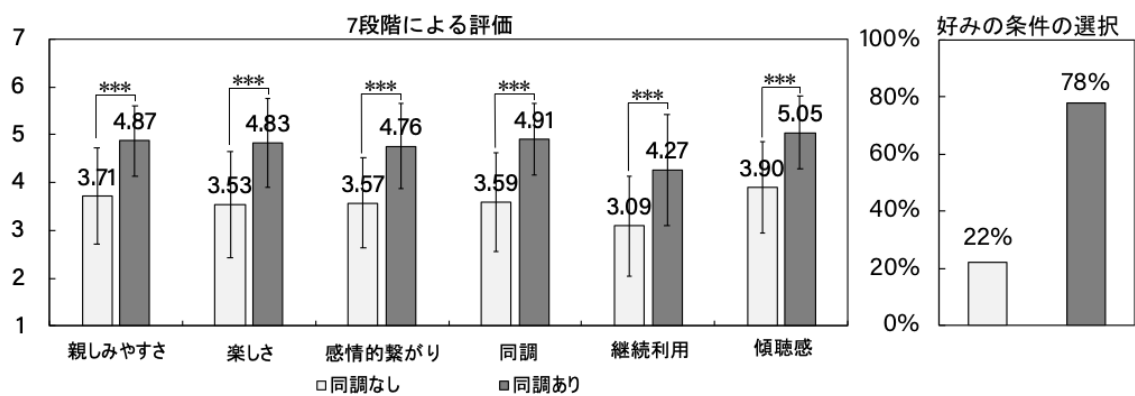


図 56 Q1 から Q7 の平均および分散

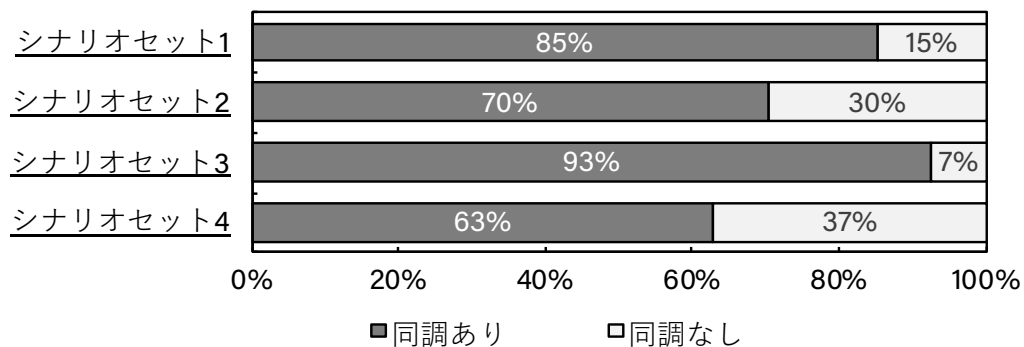


図 57 Q7 のシナリオセットごとの結果

59, 60 にそれぞれ示す。これらの結果は図 56 の結果と概ね一致している。また、質問 7 にて同調なし条件が好ましいと感じた被験者の男女の内訳は以下の通りである。

対話シナリオセット 1 : 4 名 (男性 1 名, 女性 3 名)

対話シナリオセット 2 : 8 名 (男性 4 名, 女性 4 名)

対話シナリオセット 3 : 2 名 (男性 1 名, 女性 1 名)

対話シナリオセット 4 : 10 名 (男性 4 名, 女性 6 名)

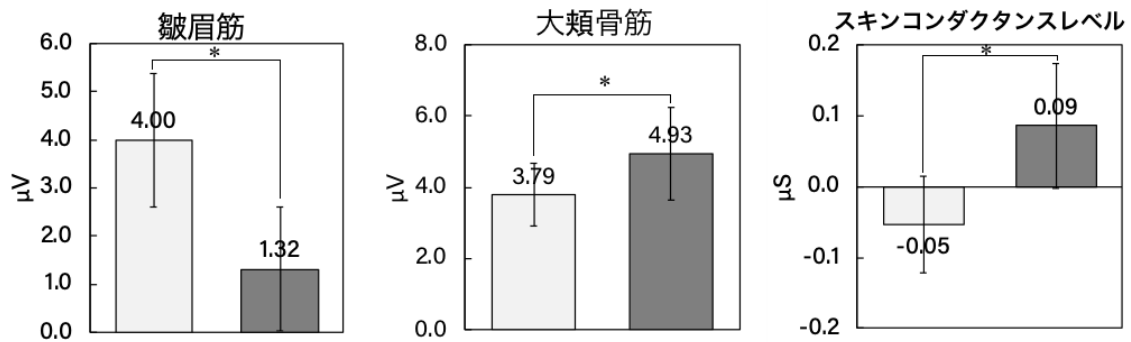


図 58 各生理指標における平均値および分散

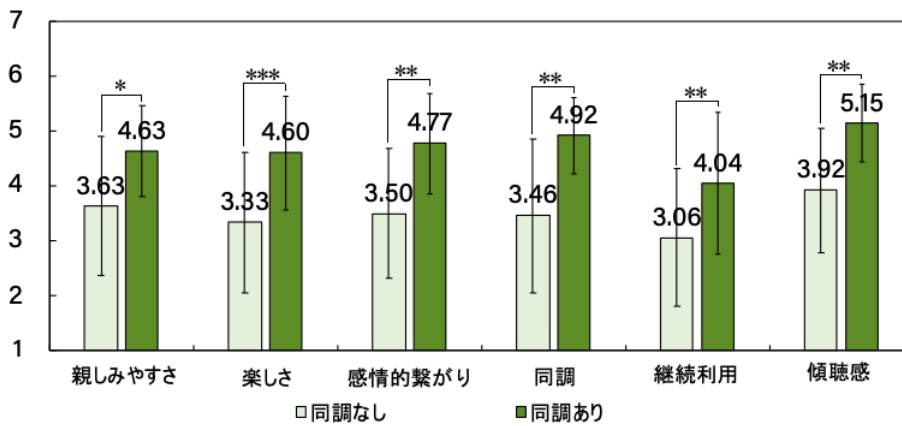


図 59 男性における Q1 から Q6 の平均値と分散

以上から、各条件の対話ロボットが被験者に与える印象は、男女間で大きな差はないと考えられる。

4.3.5 考察

Q2：ロボットと話していて楽しかったですか？ から、被験者が対話ロボットとの対話を通して喚起された感情である楽しいという感情は、アフェクトグリッド上で示される活性且つ快到位置と考えられる。つまり、対話者の話し方に同調するロボットはこの活性且つ快到の感情をより増幅したと言える。Q3,4の結果から、対話者の話し方に同調するロボットは、ラポール構築に必要とされる同調

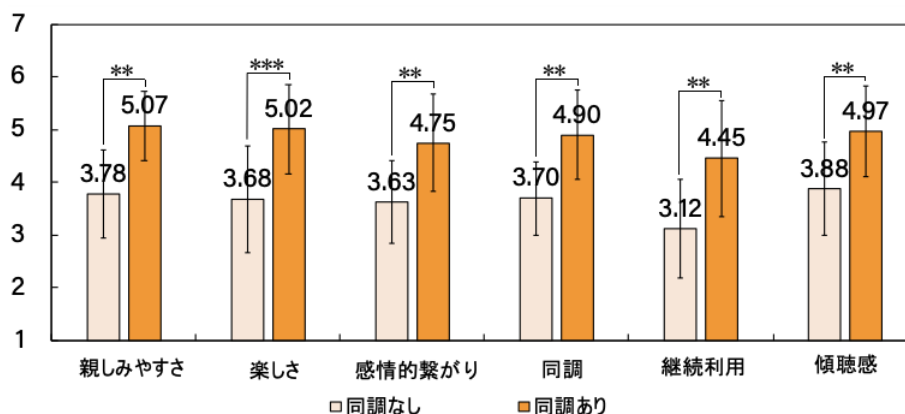


図 60 女性における Q1 から Q6 の平均値と分散

を被験者自身が実感することに貢献することがわかった。以上から、H1「対話者は自分の話し方に同調するロボットに対して、そうでない場合と比べ同調しているように感じる」が支持された。また、客観評価の結果から H3 も同様に支持された。Q5：ロボットを継続的に利用したいと感じましたか？ から、ユーザのロボットを継続的に利用するモチベーションを喚起する上で同調対話が有効であることが示された。以上から、H2「対話者は自分の話し方に同調するロボットに対して、そうでない場合と比べ継続利用意欲を示す」が支持された。客観的な評価の結果から、H3 がサポートされ、同期条件でのロボットの有効性を科学的に証明することができます。質問への評価結果から、H4 については支持されない結果となった。

H4 が支持されなかった原因を考察するために、実験環境を確認する。実験の説明や電極の装着・アンケートへの記述を含め、実験時間は被験者ごとに1時間程度であった。この間に被験者は対話シナリオセット1~4までを2回ずつ読み上げるため、被験者がロボットとのインタラクションを行う実時間は20から30分程度であった。このことから、H4：インタラクションを行う中でロボットに与える印象がどのように変化するかを調査するには、さらに継続的にロボットとのインタラクションを行うことが可能な環境を用いて調査する必要があると考えられる。以下では、今回の実験を通して得られた結果を用いてあらゆる視点から考察をさらに深める。シナリオセットによって被験者が同調ありを選ぶ割合に影響す

ることがわかった。このことから、今回用いたシナリオセットについて考察する。ロボットの印象評価に用いた各シナリオセットのトピックは、1シナリオセットごとに3つずつ与えられた。

対話シナリオセット 1： 自己紹介・好きな食べ物・映画の誘い

対話シナリオセット 2： 自炊・ご飯の誘い・寝不足の原因

対話シナリオセット 3： 好きなスポーツ・夏休みの出来事・サウナ

対話シナリオセット 4： アルバイト・お盆休み・無くした財布

質問7に対して同調あり条件ではなく、同調なし条件が好ましいと答えた被験者の詳細な内訳は以下の通りである。全シナリオセットに対して一貫して同調なしの条件が好きと答えた被験者はいなかった。4つのシナリオセットのうち3つにおいて同調なしが好ましいと答えた被験者は27名中1名であった。4つのシナリオセットのうち2つにおいて同調なしが好ましいと答えた被験者は27名中8名であった。4つのシナリオセットのうち1つにおいて同調なしが好ましいと答えた被験者は27名中5名であった。

今回の実験では被験者の共感特性を計測するテストは行なっておらず、共感特性に基づく被験者のスクリーニングできていない。同調あり条件を好ましくないと答えた被験者の自由記述からは、シナリオについての否定的な意見は得られなかったが、ロボットの声に対する印象がいくつか得られた。被験者は同調あり条件下のロボットに対して、以下のように自由記述を述べている。

「会話が早口であまりいい印象を受けなかった」、「自分の話し方に合わせてロボットも話してくれたが、気を遣われてる感じがして話しづらかった」、「財布のトピックは良かったが、アルバイトのトピック時はなんだかロボットが退屈そうだった気がした」。これら被験者からのコメントから、同調なしが場合によってはユーザに好印象を与えることがあることがわかった。ユーザに違和感や不快感を与える要因が対話中にわずかでも入ってしまうと、その条件の印象が悪くなる可能性がある。しかしながら、今回の実験で同調あり条件に適応された韻律特徴は話速・ピッチ・大きさの3つであったが、自由記述からこの3要素のみでも被

験者はロボットの抑揚を感じることができていた。本来韻律特徴に含まれる要素はこれら3つの他にも、アクセント [96], [97], 発話長 [98], 応答時間 [99], 間の周期と長さ [98], [99], 笑い [98] といった要素が含まれる。今後はこれらを複合した同調を用いた際の被験者の印象についても調査する必要があると考えられる。

一方で被験者の発話においても問題がある可能性がある。今回我々は実験導入時に被験者にロボットの韻律変化をより感じやすくするためのコントロールとして、被験者の話し方に抑揚をつけるよう彼らに指示を行った。各被験者の自由記述と実験時の動画観察の結果、被験者が発話を行う際に感情の高まり・低まりが韻律に反映されていない場合が見られた。つまり、同調していない時にロボットの韻律に大きな変化を感じたことが原因の一つとして考えられる。これは被験者の以下の同調なし条件のロボットへの自由記述から示されている。

「ロボットの話し方に抑揚があって面白かった」, 「声量の点で抑揚が大きく感じた」, 「1回目よりロボットにテンションの差があった気がした」。今回の同調なし条件の実装上、被験者の話し方に抑揚がない場合に同調がない場合においてロボットの韻律特徴がユーザにとって好印象を与える場合があることがわかった。これを解決するために、同調なし条件において、ロボットの韻律情報がある一定の値に固定するというアプローチが考えられる。

本章で提案する手法である韻律の同調は、アンケートによる主観評価だけでは適用された手法が従来手法と比較して差が大きく認められにくいという仮定の元、客観的指標に重きを置いた実験環境を作成した。しかしながら実験設計によって、潜在意識に働きかけるアプローチでも十分に主観評価に差が見られることが今回の調査を通して明らかになった。非同調条件においてロボットの合成音声を作成するために、今回の実験では被験者の平均的な韻律特徴のパラメータを求め、そこから同調する際に変化させる量のマイナス値を加算することで生成される音声を、非同調条件において使用した理由を述べる。提案する同調対話と対となるベースラインの作成手法を決定する際に、今回の実験で用いた手法のほかにパラメータを全てフラットにする条件や、ランダムなパラメータを用いて合成音声を作成するといった、方法についての検討を行っていた。その結果、まず明らかに不自然な発話を生む可能性を持つランダムなパラメータが排除され、続いてフラット

なパラメータを使用することについての検討が行われた。実験ではシナリオベースの対話において被験者に抑揚をつける箇所を指示したが、抑揚をつける程度によってはフラットなパラメータを用いた際には、パラメータの実数値に差が付きにくいと考えられる。ここで、マイナス値を加算する今回の手法を用いることで、たとえば被験者の発話する際の抑揚が小さい場合でも、パラメータの実数値として差が付きやすくなるため、今回の調査では後者の手法をベースラインに使用することが決定した。結果的に生理指標を用いた客観評価のみならず、主観による評価からも全ての項目で有意差が認められたため、今回選択した非同期条件に用いたパラメータの決定方法には一定の価値があったと言える。研究の今後の課題として、この比較対象にフラットなパラメータを用いたとき、あるいはランダムなパラメータを用いて生成される合成音声を使用した場合の調査の必要性が考えられる。

以上の結果を元に、主観による影響の比較的少ないと考えられるアプローチに対して適切に実験設計を行うために特に重要な点は、予備検討を十分に行うことが挙げられる。今回の韻律情報について同調対話を行うロボットの評価実験では、予備検討によって得られる結果から、次回の実験設計時に役立つ情報を精査しどのように改善すべきかを議論することで、主観的な視点からは差がわかりにくいと考えられる潜在的なアプローチに対して、有効な実験設計を構築することができた。

5. 考察と研究の限界

本研究ではTV雑談ロボットに対し、ユーザの継続利用意欲を向上すると予想される振る舞いとして、自然な振る舞い、および共感を示す振る舞いの大きく2種類に分類された中から、重要であると考えられる手法を選出し、手法の提案と評価実験による調査を行った。その結果、提案された5つの手法はそれぞれにおいてユーザの継続利用意欲の向上に貢献することが示唆された。ここで、提案する全ての手法で期待する効果が得られた理由について考察する。この要因としては、今回提案された各手法において想定される意図が、ユーザのニーズとマッチしていたことが要因であると考えられる。例として、3にて説明した自然な振る舞いとして今回検証した要素である、応答を示す発話とトピックに同期する発話において継続利用意欲の促進に貢献した理由について考える。今回実装のテーマとした、自然な振る舞いを実装するにあたり、継続利用に対しネガティブに働くと考えられる要因（ユーザの発話に対して応答しないこと・発話内容と番組トピックとの不一致）を排除することができ、ポジティブな要因を阻害しない発話が設計できていたことが共通して言える。同様に、嗜好を反映する発話、情動共有、韻律模倣についても、こちらの意図するロボットの振る舞いがユーザに受け入れられる傾向が見られた。まとめると、こちらの提案に対しユーザのニーズがマッチしていたため、提案する手法の設計が妥当であったことが明らかになった。

一方、提案した各機能が全ての被験者にとって期待する効果を与えたかという点、必ずしもそうではない。全ての被験者が提案する手法に納得することはなかった理由を説明する、これは先ほどの例とは反対に、こちら側が想定する意図とユーザのニーズがマッチしなかったことが理由として考えられる。例として、

例1: TVは一人で見たいという意見を持つ被験者がいた

例2: ロボットが自分の話し方に同調することを嫌がる被験者がいた

例3: ロボットの音声や見た目について違和感を感じることから離れられなくなってしまう被験者がいた

などが挙げられる。これらの被験者についての性別や年齢などの共通する特徴は得られなかったが、この分析により、図61に示すように、提案する手法ごとにそ

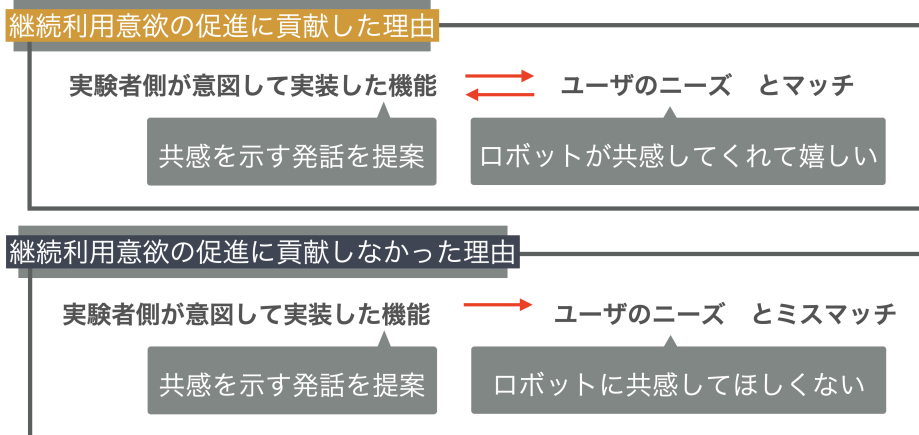


図 61 提案した振る舞いが継続利用意欲を促進した理由

れ自体にニーズを感じないどころか、返ってネガティブな印象を与える要因となる可能性があることが明らかになった。このような特徴を持つ被験者に共通して対処するために、ロボットに対するネガティブな印象を与える要因を排除する必要がある。具体的な対処の例として、例1においては、ユーザが今話しかけられたいかどうかという対話意欲をロボットが認識し、その時点で発話を行うかどうかを判断する機構を設けるなどが考えられる。今回のケースから、ユーザによる個人差に対応する手段として、システムが提供する要素とユーザのニーズがマッチしているかどうかを判断する取り組みが必要であることがわかった。

次に本研究全体の限界について述べる。本研究で提案した全ての手法に当てはまる研究の限界として、今回のTV雑談ロボットをユーザが「使い続けたい」と思うかどうかを、長期的なロボットとの関係から評価していないことが挙げられる。よって、今回の調査によってTV雑談ロボットにおいて継続利用に繋がる要素を個別に検討した結果を踏まえ、長期的な関係の調査を行った際の効果を調査する必要がある。図62に本研究では調査していない範囲を示す。短期間の利用でユーザがロボットに初見時に抱く印象を調査するフェーズから、中長期的にユーザにロボット利用を促進する課程の中で、ユーザのニーズにマッチしたサービスを提供し続けることによる影響については、本研究では明らかにしていない。また、「自然な振る舞い」や「共感を示す振る舞い」という分類の中で、今回提案した手法以外にも調査が必要な手法として、自然さを高手法には、ロボットのアピ

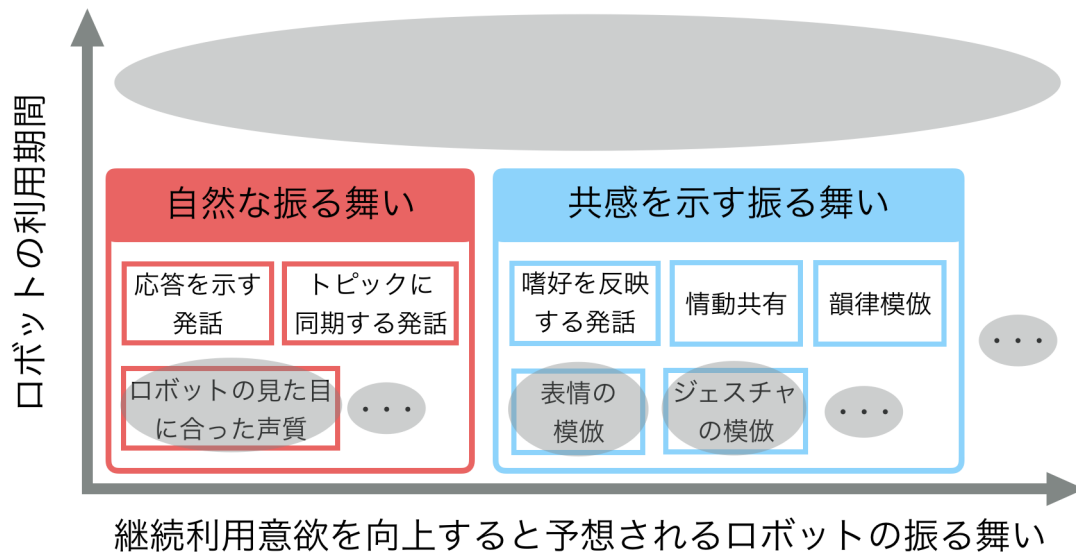


図 62 今後の課題として調査すべき範囲（黒影の範囲内）

アランスにあった声質を選択することや、共感を示すに人の表情やジェスチャを模倣するなどが今回の調査を通して考えられた。これらの追加要素についても今後調査を行う必要がある。

ユーザが日常生活シーンのどの場面においてもロボットと雑談を行うモチベーションが保たれる、という研究の最終目的にたどり着くためにはどの段階から、TV 視聴以外の日常的な生活シーンを追加していくかなど、追加検討すべき内容は多い。そのような調査を行う際、今回提案した各機能をどのように組み合わせるべきかについて、これまでの結果を踏まえて考察する。今回提案した各手法をどのように組み合わせるべきかについて、これまでの結果を踏まえて考察する。各手法間でシナジーが強いと考えられる組み合わせとして、3.2 トピックに同期する発話と 4.2 情動共有が挙げられる。リアルタイムでの運用に関して問題を抱える情動共有のデメリットを発話タイミングをトピックに同期するよう、発話タイミングを同期させることで、ユーザの継続利用意欲の更なる向上が期待される。一方、その他の手法である 3.1 応答を示す発話や 4.3 韻律模倣、4.1 嗜好を反映した発話に関する組み合わせを考えた際も、互いに他の手法により得られるメリットを阻害せずに組み合わせることができる可能性が高いと考えられる。この理由

は各手法がTV雑談ロボットに関わらず、対話を行う上でのテクニックとして人同士のコミュニケーション中にも取り入れられており、掛け合わせることでトレードオフになる要素ではないためである。しかし、これらを組み合わせたTV雑談ロボットを評価する際にも、ユーザのニーズにマッチした振る舞いを行う必要がある、個人差に対応することや各手法をどの程度の頻度で発話に適用させるか、など今回調査の範囲に含まれていなかった点を考慮して組み合わせる必要があると考えられる。さらに、長期的な関係の中で組み合わせの評価を行う必要がある。

6. 結論

本研究は人とコミュニケーションを行う対話ロボットがユーザとの関係構築の初期段階において、ロボットの継続的な利用に繋げるための必要な要素の調査を行なった。そのために、ユーザがロボットを利用する上で心理的障壁が低いと考えられる、特定の日常的な生活シーンとしてTV視聴に着目した。本研究の目的である、TV視聴というシーンにおいて継続的に利用され続けるTV雑談ロボットとはどのようなロボットであるべきか明らかにするため、継続利用意欲を向上すると予想されるロボットの振る舞いとして選択された5つの手法について短期間における関係性の評価実験を行い、継続利用に繋がる要因を調査した。その結果TV雑談ロボットにおいて継続利用意欲を向上させる上で2つの設計指針が効果的であることが明らかになった。1つは、ロボットの自然な振る舞いの実現を目指した設計、もう1つはロボットがユーザに対し共感を示す振る舞いの実現を目指した設計である。以下に本研究でTV雑談ロボットの振る舞いを決定するための個々の提案手法についてまとめる。

「自然な振る舞い」を実現するための2つの手法について、「3.1 応答を示す発話」では、TV雑談ロボットユーザへの応答に着目し、SNSコメント以外の相槌・復唱・自動応答文生成という3つの機械応答を組み合わせることで、ユーザに対し自然な応答を返すTV雑談ロボットの提案を行った。提案する手法を適用したロボットは、SNSコメントのみで発話を行うロボットよりも自然な応答性を実現し、この手法によってユーザの継続利用意欲が向上することが示された。「3.2 トピックに同期する発話」では、TV雑談ロボットがTV番組のトピックに同期する発話を実現することによる、影響を調査した。具体的にはSNSコメントを発話する内容を遷移する番組のトピックに同期させた場合とそうでない場合を比較した。これにより、ユーザに対しTV雑談ロボットが発するトピックに同期しない発話によって与える不快さを低減し、TV雑談ロボットの継続利用意欲を高めることが示された。また、音声指向型のコンテンツとして分類されるニュース番組が映像指向型コンテンツであるスポーツ番組よりも、トピックに同期するロボット発話が効果的であることが示された。

次に「共感を示す振る舞い」を実現するための3つの手法について、「4.1 嗜好を

反映する発話」では、膨大な量の SNS コメントからユーザ個人の趣向にあったコメントを適切に選択するためのフィルタの生成手法を提案した。個人の趣向を反映した発話文の選択手法では大量の SNS コメントに対する評価データから、フィルタリングされたコメントの抽出を行うための基盤を構築し、フィルタを通してどのようなコメントが抽出されるかを示した。この手法によってロボットが発話する SNS コメントをユーザが好ましいと感じ、結果としてロボットの継続利用意欲が向上することが期待されることが示唆された。「4.2 情動共有」では、盛り上がりや笑いの雰囲気や SNS コメントから推定する手法を提案し、各雰囲気において TV 雑談ロボットがどのように振る舞うかを決定した。各雰囲気において情動共有を適用されたロボットは、バーチャルな見た目のロボットにおいて最も評価の高い結果となり、ユーザの継続利用意欲の向上に貢献することを示した。また盛り上がりや笑いの雰囲気において、生理指標を用いた客観的な評価の結果、提案する情動共有によってユーザの覚醒度が有意に高まり、ロボットとポジティブな感情を共有できることが示された。「4.3 韻律模倣」では、ノンバーバルコミュニケーションにおいて韻律情報に着目し、ユーザの韻律情報に同調して発話を行う機能においては、人の韻律を模倣する対話ロボットを主観的・客観的な指標を用いて評価した結果、共に同調対話が適用された群が適用されない群と比較して有意に実験参加者に共感しているという印象を与えた。また韻律同調を適用したロボットが継続利用意欲の促進に貢献することが示された。

個々の研究の分析を統合した結果、人とロボットの短期的な関係性の調査から、ロボットの振る舞いにおいて、ユーザにネガティブな印象を与えると考えられる要因を検討し、排除することの重要性が示唆された。この要因はユーザの持つ個人差によって異なり、ユーザの特性を認識し対応する手段として、ロボットが提供する振る舞いとユーザのニーズがマッチしているかどうかを判断する取り組みが必要であることがわかった。以上から、本研究で TV 雑談ロボットに対し実装した提案手法が継続利用意欲を高める上で効果的であり、今回の明らかになった設計指針に従い TV 雑談ロボットのインタラクションデザインを設計することの妥当性が示された。

日常的な継続利用を促進した先には、ユーザとロボットの間で愛着のような深

い関係が構築されると考えられる。このようにして築き上げられた関係に基づき、ロボットからの能動的な働きかけによって、ユーザの行動変容といった健康促進の支援、QoLの向上等の促進を実現するという研究の最終的なゴールに対し、本研究で調査し結論づけたことに加え、「継続利用意欲を向上すると予想されるロボットの振る舞い」を網羅する要素についての検討を行うことや、明らかとなった要素同士を適切に組み合わせた際の影響を考慮することが今後の課題である。

本研究ではこれまで述べてきた各手法をTV雑談ロボットに適用することのみ焦点を絞って考えてきたが、4.3同調対話においてはTV雑談ロボット以外の非タスク指向型の対話ロボットに応用可能な技術であるといえる。その他、自然で共感を示す振る舞いというTV雑談ロボットを用いて明らかになった知見を一般化するための調査を行うことで、TV以外のコンテンツに対しても、同様の対話システムを構築することが可能である。このように、TV雑談システムが他のコンテンツにも対応可能となり応用されていく流れが、今後広がりを見せることを期待している。

謝辞

研究活動全般にわたり、指導教員の加藤 博一教授には幅広い視野で多くの助言をいただきました。ここに感謝の意を表します。丁寧かつ熱心なご指導、ご鞭撻をさせていただいた神原 誠之准教授に深く感謝致します。副指導教官としての確なご指摘をしてくださった小笠原 司教授に熱く御礼申し上げます。博士論文審査員として、また旧環境知能学研究室客員教授として、熱心な議論を通し的確なご指摘をしてくださった萩田 紀博教授に深く御礼申し上げます。論文作成にあたり丁寧かつ的確なご指摘をしてくださった藤本 雄一郎助教に熱く御礼申し上げます。心理学的な面から実験設計や論文作成についてご指導、ご鞭撻をさせていただいた佐藤 弥先生に深く感謝致します。研究についてのアドバイスだけでなく、精神的なサポートをさせていただいた宮下 敬宏先生に深く感謝致します。近い立場で研究に関する相談を快く引き受けてくれた博士研究員の澤邊太志さんに深く御礼申し上げます。本研究の実験を実施するにあたり、多くの事務手続きを一手に引き受けてくださった南 あずさ秘書、上野 真紀子秘書、小川 暁子秘書に深くお礼申し上げます。本研究で使用したTV視聴型雑談アプリケーションの根幹となる、TV雑談サーバの特許出願に対して尽力してくださった久保 浩三先生をはじめ、国立大学法人 奈良先端科学技術大学院大学 研究推進機構参観学連携推進部門の皆様にも深くお礼申し上げます。共に研究活動を歩み、日々切磋琢磨したインタラクティブメディア設計学研究室および理化学研究所ロボティクスプロジェクト心理プロセス研究チームの皆様、また研究の実験に参加・協力してくださった被験者の皆様に心から感謝申し上げます。最後に、今までお世話になった両親に深く感謝致します。

参考文献

- [1] 内閣府. 令和元年版高齢社会白書. Technical report, Cabinet Office Government of Japan, 2019. <https://www8.cao.go.jp/kourei/whitepaper/w-2019/html/zenbun/index.html> [15 Mar, 2021 Accessed].
- [2] 内閣府. 令和元年版子供・若者白書. Technical report, Cabinet Office Government of Japan, 2019. <https://www8.cao.go.jp/youth/whitepaper/r01honpen/index.html> [15 Mar, 2021 Accessed].
- [3] Martha E Pollack. Intelligent technology for an aging population: The use of ai to assist elders with cognitive impairment. *AI magazine*, Vol. 26, No. 2, pp. 9–9, 2005.
- [4] Julianne Holt-Lunstad, Timothy B. Smith, and J. Bradley Layton. Social relationships and mortality risk: A meta-analytic review. *PLOS Medicine*, Vol. 7, No. 7, pp. 1–1, 07 2010. <https://doi.org/10.1371/journal.pmed.1000316> [15 Mar, 2021 Accessed].
- [5] 総務省. 令和2年版情報通信白書. 2005. <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r02/pdf/02honpen.pdf> [15 Mar, 2021 Accessed].
- [6] Timothy W Bickmore and Rosalind W Picard. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)*, Vol. 12, No. 2, pp. 293–327, 2005.
- [7] Takayuki Kanda, Rumi Sato, Naoki Saiwaki, and Hiroshi Ishiguro. A two-month field trial in an elementary school for long-term human-robot interaction. *IEEE Transactions on robotics*, Vol. 23, No. 5, pp. 962–971, 2007.
- [8] Iolanda Leite, Carlos Martinho, and Ana Paiva. Social robots for long-term interaction: a survey. *International Journal of Social Robotics*, Vol. 5, No. 2, pp. 291–308, 2013.

- [9] Robert B Zajonc. Attitudinal effects of mere exposure. *Journal of personality and social psychology*, Vol. 9, No. 2p2, p. 1, 1968.
- [10] William R Kunst-Wilson and Robert B Zajonc. Affective discrimination of stimuli that cannot be recognized. *Science*, Vol. 207, No. 4430, pp. 557–558, 1980.
- [11] Machiko Kusahara. The art of creating subjective reality: an analysis of japanese digital pets. *Leonardo*, Vol. 34, No. 4, pp. 299–302, 2001.
- [12] A Kerepesi, E Kubinyi, GK Jonsson, MS Magnusson, and A Miklosi. Behavioural comparison of human–animal (dog) and human–robot (aibo) interactions. *Behavioural processes*, Vol. 73, No. 1, pp. 92–99, 2006.
- [13] Frederic Kaplan. Talking aibo: First experimentation of verbal interactions with an autonomous four-legged robot. In *Learning to behave: interacting agents CELE-TWENTE Workshop on Language Technology*, Vol. 22. Cite-seer, 2000.
- [14] Google. Google home-スマートスピーカーホーム アシスタント-googleストア. https://store.google.com/product/google_home?hl=ja [15 Mar, 2021 Accessed.].
- [15] Amazon. Amazon — echo - スマートスピーカー. <http://amzn.asia/8Xp41Ga> [15 Mar, 2021 Accessed.].
- [16] Ryuichiro Higashinaka, Kotaro Funakoshi, Yuka Kobayashi, and Michimasa Inaba. The dialogue breakdown detection challenge: Task description, datasets, and evaluation metrics. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pp. 3146–3150, 2016.
- [17] Richard S Wallace. The anatomy of alice. In *Parsing the Turing Test*, pp. 181–210. Springer, 2009.

- [18] Kosuke Tsujino, Shinya Iizuka, Yusuke Nakashima, and Yoshinori Isoda. Speech recognition and spoken language understanding for mobile personal assistants: A case study of "shabette concier". In *2013 IEEE 14th international conference on mobile data management*, Vol. 2, pp. 225–228. IEEE, 2013.
- [19] Kango Iwama and Yoshinobu Kano. Japanese advertising slogan generator using case frame and word vector. In *Proceedings of the 11th International Conference on Natural Language Generation*, pp. 197–198, 2018.
- [20] T Takahashi, M Kanbara, and N Hagita. A social media mediation robot to increase an opportunity of conversation for elderly: mediation experiments using single or multiple robots. *Technical Committee on Cloud Network robotics (CNR)*, Vol. 113, No. 84, pp. 31–36, 2013.
- [21] Christian Jones and Andrew Deeming. Affective human-robotic interaction. In *Affect and emotion in human-computer interaction*, pp. 175–185. Springer, 2008.
- [22] Christoph Bartneck and Jodi Forlizzi. A design-centred framework for social human-robot interaction. In *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No. 04TH8759)*, pp. 591–594. IEEE, 2004.
- [23] 宮下善太, 神田崇行, 塩見昌裕, 石黒浩, 萩田紀博. 顧客と顔見知りになるショッピングモール案内ロボット. *日本ロボット学会誌*, Vol. 26, No. 7, pp. 821–832, 2008.
- [24] Alessandra Maria Sabelli, Takayuki Kanda, and Norihiro Hagita. A conversational robot in an elderly care center: an ethnographic study. In *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 37–44. IEEE, 2011.

- [25] Marcel Heerink, Ben Kröse, Bob Wielinga, and Vanessa Evers. Enjoyment intention to use and actual use of a conversational robot by elderly people. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pp. 113–120. ACM, 2008.
- [26] Kouki Miyazawa, Toru Tokoyo, Yusuke Masui, Norimichi Matsuo, and Hideaki Kikuchi. Factors of interaction in the spoken dialogue system with high desire of sustainability. *Trans. IE-ICE (A)*, Vol. 95, No. 1, pp. 27–36, 2012.
- [27] Nicole Krämer, Astrid Marieke Rosenthal-von der Pütten, and Sabrina Eimler. Human-agent and human-robot interaction theory: Similarities to and differences from human-human interaction. *Studies in Computational Intelligence*, Vol. 396, pp. 215–240, 01 2012.
- [28] Justine Cassell, Joseph Sullivan, Elizabeth Churchill, and Scott Prevost. *Embodied conversational agents*. MIT press, 2000.
- [29] 山田誠二, 角所考, 小松孝徳ほか. 人間とエージェントの相互適応と適応ギャップ (特集 hai ヒューマンエージェントインタラクションの最先端). *人工知能学会誌*, Vol. 21, No. 6, pp. 648–653, 2006.
- [30] Seiji Yamada and Takanori Komatsu. Designing simple and effective expression of robot’s primitive minds to a human. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2614–2619. IEEE, 2006.
- [31] Kazuhiko Shinozawa, Futoshi Naya, Junji Yamato, and Kiyoshi Kogure. Differences in effect of robot and screen agent recommendations on human decision-making. *International journal of human-computer studies*, Vol. 62, No. 2, pp. 267–279, 2005.
- [32] 神原誠之, 向井田一平, 西村祥吾. 特開 2019-045978, 対話制御装置, 学習装置, 対話制御方法, 学習方法, 制御プログラム, および, 記録媒体, 2019-3-22.

- [33] 株式会社 amirobo tech. Dabelive. <http://www.amirobo.tech/service/> [15 Mar, 2021 Accessed.].
- [34] Tanya L Chartrand and John A Bargh. The chameleon effect: the perception–behavior link and social interaction. *Journal of personality and social psychology*, Vol. 76, No. 6, p. 893, 1999.
- [35] Michael J Hove and Jane L Risen. It’s all in the timing: Interpersonal synchrony increases affiliation. *Social cognition*, Vol. 27, No. 6, pp. 949–960, 2009.
- [36] Rick B Van Baaren, Rob W Holland, Kerry Kawakami, and Ad Van Knippenberg. Mimicry and prosocial behavior.
- [37] 神田崇行, 平野貴幸, イートンダニエル, 石黒浩. 日常生活の場で長期相互作用する人間型対話ロボット. 日本ロボット学会誌, Vol. 22, No. 5, pp. 636–647, 2004.
- [38] Kazuyoshi Wada, Takanori Shibata, Tomoko Saito, and Kazuo Tanie. Analysis of factors that bring mental effects to elderly people in robot assisted activity. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vol. 2, pp. 1152–1157. Ieee, 2002.
- [39] 神田崇行, 石黒浩, 小野哲雄, 今井倫太, 中津良平. 人間と相互作用する自律型ロボット robovie の評価. 日本ロボット学会誌, Vol. 20, No. 3, pp. 315–323, 2002.
- [40] 中田亨, 森武俊, 佐藤知正. ロボットの身体動作表現と生成される印象とのラバン特徴量を介した定量的相関分析. 日本ロボット学会誌, Vol. 19, No. 2, pp. 252–259, 2001.
- [41] Yuki Kamiya, Tomohiro Ohno, and Shigeki Matsubara. Coherent back-channel feedback tagging of in-car spoken dialogue corpus. In *Proceedings*

- of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue, pp. 205–208. Association for Computational Linguistics, 2010.
- [42] Yuka Kobayashi, Daisuke Yamamoto, Toshiyuki Koga, Sachie Yokoyama, and Miwako Doi. Design targeting voice interface robot capable of active listening. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, pp. 161–162. IEEE Press, 2010.
- [43] Nikolaos Mavridis, Chandan Datta, Shervin Emami, Andry Tanoto, Chiraz BenAbdelkader, and Tamer Rabie. Facebots: robots utilizing and publishing social information in facebook. In *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 273–274. IEEE, 2009.
- [44] Nikolaos Mavridis, Michael Petychakis, Alexandros Tsamakos, Panos Toulis, Shervin Emami, Wajahat Kazmi, Chandan Datta, Chiraz BenAbdelkader, and Andry Tanoto. Facebots: Steps towards enhanced long-term human-robot interaction by utilizing and publishing online social information. *Paladyn, Journal of Behavioral Robotics*, Vol. 1, No. 3, pp. 169–178, 2010.
- [45] Y Yoshida and M Hagiwara. A dialogue system with humor using multiple linguistic resources. *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, Vol. 26, No. 2, pp. 627–636, 2014.
- [46] Alan Ritter, Colin Cherry, and William B Dolan. Data-driven response generation in social media. In *Proceedings of the conference on empirical methods in natural language processing*, pp. 583–593. Association for Computational Linguistics, 2011.
- [47] Raymond E Barber and Henry C Lucas Jr. System response time operator productivity, and job satisfaction. *Communications of the ACM*, Vol. 26, No. 11, pp. 972–986, 1983.

- [48] T Goodman and Robert Spence. The effect of system response time on interactive computer aided problem solving. *ACM SIGGRAPH Computer Graphics*, Vol. 12, No. 3, pp. 100–104, 1978.
- [49] Jan L Guynes. Impact of system response time on state anxiety. *Communications of the ACM*, Vol. 31, No. 3, pp. 342–347, 1988.
- [50] Robert B Miller. Response time in man-computer conversational transactions. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, pp. 267–277, 1968.
- [51] Thad Starner. The challenges of wearable computing: Part 2. *Ieee Micro*, Vol. 21, No. 4, pp. 54–67, 2001.
- [52] Akinobu Lee and Tatsuya Kawahara. Recent development of open-source speech recognition engine julius. In *Proceedings: APSIPA ASC 2009: Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference*, pp. 131–137. Asia-Pacific Signal and Information Processing Association, 2009 Annual ..., 2009.
- [53] 西宏之. キーワードネットワークを用いた電話取り次ぎ対話処理. 信学技報, pp. SP88–30, 1988.
- [54] Yahoo! JAPAN Developer Network. Text analysis. [Online] Available: <http://developer.yahoo.co.jp/webapi/jlp> (in Japanese) [15 Mar, 2021 Accessed].
- [55] twitter. Welcome to twitter. [Online] Available: <https://twitter.com/> (in Japanese) [15 Mar, 2021 Accessed].
- [56] J Smith. *GrandChair: Conversational collection of family stories*. PhD thesis, MS Thesis, MIT Media Arts & Sciences, 2000.
- [57] K. Yamamoto, K. Inoue, S. Nakamura, K. Takanashi, and T. Kawahara. Dialogue behavior control model for expressing a character of humanoid

- robots. In *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 1732–1737, Nov 2018.
- [58] Ltd. PIP Co. Unazuki-kabochan (chat robot). [Online] Available: <http://www.pip-club.com/kabo/index.html> (in Japanese) [15 Mar, 2021 Accessed.].
- [59] HOYA Corporation. *Speech synthesis voiceText web API*. [Online] Available: <http://voicetext.jp/products/vt-webapi/> (in Japanese) [15 Mar, 2021 Accessed.].
- [60] Bilyana Martinovsky and David Traum. The error is the clue: Breakdown in human-machine interaction. Technical report, UNIVERSITY OF SOUTHERN CALIFORNIA MARINA DEL REY CA INST FOR CREATIVE ..., 2006.
- [61] Japanese tts system: Open jtalk. [Online] Available: <http://open-jtalk.sourceforge.net> [15 Mar, 2021 Accessed.].
- [62] Kotaro Funakoshi, Kazuki Kobayashi, Mikio Nakano, Seiji Yamada, Yasuhiko Kitamura, and Hiroshi Tsujino. Smoothing human-robot speech interactions by using a blinking-light as subtle expression. In *Proceedings of the 10th international conference on Multimodal interfaces*, pp. 293–296. ACM, 2008.
- [63] Apple. App store. <https://www.apple.com/jp/ios/app-store/> [15 Mar, 2021 Accessed.].
- [64] Google. Google play store. <https://play.google.com/store> [15 Mar, 2021 Accessed.].
- [65] ドワンゴ. ニコニコ実況. <http://jk.nicovideo.jp/> [15 Mar, 2021 Accessed.].

- [66] Apple. Avspeechsynthesizer - avfoundation — apple developer documentation. <https://developer.apple.com/documentation/avfoundation/avspeechsynthesizer> [15 Mar, 2021 Accessed].
- [67] Google. Texttospeech — android developers. <https://developer.android.com/reference/android/speech/tts/TextToSpeech.html> [15 Mar, 2021 Accessed].
- [68] 有安香子, 藤沢寛, 砂崎俊二. テレビ番組に関するメッセージ解析手法 (hybridcast 特集号). NHK 技研 R&D, No. 133, pp. 28–34, 2012.
- [69] Kudo Taku, Yamamoto Kaoru, and Matsumoto Yuji. Applying conditional random fields to japanese morphological analysis. In *EMNLP*, Vol. 4, pp. 230–237, 2004.
- [70] 佐藤敏紀, 橋本泰一, 奥村学. 単語分かち書き辞書 mecab-ipadic-neologd の実装と情報検索における効果的な使用方法の検討. 言語処理学会第 23 回年次大会 (NLP2017), pp. NLP2017–B6–1. 言語処理学会, 2017.
- [71] James A Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, Vol. 39, No. 6, pp. 1161–1178, 1980.
- [72] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- [73] Takamasa Iio, Yuichiro Yoshikawa, and Hiroshi Ishiguro. Starting a conversation by multi-robot cooperative behavior. In *International Conference on Social Robotics*, pp. 739–748. Springer, 2017.
- [74] Takahiro Matsumoto, Shunichi Seko, Ryouyuke Aoki, Akihiro Miyata, Tomoki Watanabe, and Tomohiro Yamada. Affective agents for enhancing emotional experience. In *Proceedings of the second international conference on Human-agent interaction*, pp. 169–172, 2014.

- [75] Saif M Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu. Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. *arXiv preprint arXiv:1308.6242*, 2013.
- [76] CHE Gilbert and Erric Hutto. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*. Available at (20/04/16) <http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf>, Vol. 81, p. 82, 2014.
- [77] Norjihhan Abdul Ghani, Suraya Hamid, Ibrahim Abaker Targio Hashem, and Ejaz Ahmed. Social media big data analytics: A survey. *Computers in Human Behavior*, Vol. 101, pp. 417–428, 2019.
- [78] David A Shamma, Lyndon Kennedy, and Elizabeth F Churchill. Tweet the debates: understanding community annotation of uncollected sources. In *Proceedings of the first SIGMM workshop on Social media*, pp. 3–10, 2009.
- [79] Newton’s method. [Online] Available: https://en.wikipedia.org/wiki/Newton's_method (in Japanese) [15 Mar, 2021 Accessed].
- [80] 中澤昌美, 帆足啓一郎, 小野智弘. Twitter によるテレビ番組重要シーン検出及びラベル付与手法. In *DEIM Forum*, pp. 517–519, 2011.
- [81] Shogo Nishimura, Masayuki Kanbara, and Norihiro Hagita. Atmosphere sharing with tv chat agents for increase of user ’s motivation for conversation. In *International Conference on Human-Computer Interaction*, pp. 482–492. Springer, 2019.
- [82] SHARP Corporation. *RoBoHon, Product information*. [Online] Available: <https://robohon.com/product/robohon.php> (in Japanese) [15 Mar, 2021 Accessed].

- [83] Kumada Masanobu, Kobayashi Takeo, Kosaki Hiroko, and Niimi Seiji. "mora method" for objective evaluation of severity of spasmodic dysphonia. *The Japan Journal of Logopedics and Phoniatrics*, Vol. 38, No. 2, pp. 176–181, 1997.
- [84] James A Russell, Anna Weiss, and Gerald A Mendelsohn. Affect grid: a single-item scale of pleasure and arousal. *Journal of personality and social psychology*, Vol. 57, No. 3, p. 493, 1989.
- [85] Marianne LaFrance. Posture mirroring and rapport. *Interaction rhythms: Periodicity in communicative behavior*, Vol. 279298, , 1982.
- [86] Xiaofan Sun and Anton Nijholt. Multimodal embodied mimicry in interaction. In *Analysis of Verbal and Nonverbal Communication and Enactment. The Processing Issues*, pp. 147–153. Springer, 2011.
- [87] Richard E Maurer and Jeffrey H Tindall. Effect of postural congruence on client's perception of counselor empathy. *Journal of counseling psychology*, Vol. 30, No. 2, p. 158, 1983.
- [88] Claire Schmais and Aaron Schmais. Reflecting emotions: The movement-mirroring test. *Journal of Nonverbal Behavior*, Vol. 8, No. 1, pp. 42–54, 1983.
- [89] Frank J Bernieri and Robert Rosenthal. Interpersonal coordination: Behavior matching and interactional synchrony. 1991.
- [90] Joseph N Cappella. Coding mutual adaptation in dyadic nonverbal interaction. *The sourcebook of nonverbal measures: Going beyond words*, pp. 383–392, 2005.
- [91] Theodora Chaspari and Jill Fain Lehman. An acoustic analysis of child-child and child-robot interactions for understanding engagement during speech-controlled computer games. In *INTERSPEECH*, pp. 595–599, 2016.

- [92] Najmeh Sadoughi, André Pereira, Rishub Jain, Iolanda Leite, and Jill Fain Lehman. Creating prosodic synchrony for a robot co-player in a speech-controlled game for children. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 91–99. ACM, 2017.
- [93] Noriko Suzuki, Yugo Takeuchi, Kazuo Ishii, and Michio Okada. Effects of echoic mimicry using hummed sounds on human–computer interaction. *Speech Communication*, Vol. 40, No. 4, pp. 559–573, 2003.
- [94] Richard L Street Jr. Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research*, Vol. 11, No. 2, pp. 139–169, 1984.
- [95] Emilie Delaherche and Mohamed Chetouani. Multimodal coordination: exploring relevant features and measures. In *Proceedings of the 2nd international workshop on Social signal processing*, pp. 47–52. ACM, 2010.
- [96] Howard Giles. Accent mobility: A model and some data. *Anthropological linguistics*, pp. 87–105, 1973.
- [97] Richard Y Bourhis and Howard Giles. The language of intergroup distinctiveness. *Language, ethnicity and intergroup relations*, Vol. 13, p. 119, 1977.
- [98] Frances R Bilous and Robert M Krauss. Dominance and accommodation in the conversational behaviours of same-and mixed-gender dyads. *Language & Communication*, 1988.
- [99] Joseph N Cappella and Sally Planalp. Talk and silence sequences in informal conversations iii: Interspeaker influence. *Human Communication Research*, Vol. 7, No. 2, pp. 117–132, 1981.
- [100] *SPTK working group, Examples for Using Speech Signal Processing Toolkit Ver. 3.9*, 2015.

- [101] David Gouaillier, Vincent Hugel, Pierre Blazeovic, Chris Kilner, Jerome Monceaux, Pascal Lafourcade, Brice Marnier, Julien Serre, and Bruno Maisonnier. The nao humanoid: a combination of performance and affordability. *CoRR abs/0807.3223*, 2008.
- [102] Ken-ichi Iso. Design of a japanese sentence list for a speech database. *Proc. Acoust. Soc. Japan (March)*, 1988, Vol. 1, pp. 89–90, 1988.
- [103] Akira Kurematsu, Kazuya Takeda, Yoshinori Sagisaka, Shigeru Katagiri, Hisao Kuwabara, and Kiyohiro Shikano. Atr japanese speech database as a tool of speech recognition and synthesis. *Speech communication*, Vol. 9, No. 4, pp. 357–363, 1990.
- [104] Toshiaki Kakii. Characteristics of multimedia counseling: A study of an interactive tv system. *Shinrigaku kenkyu: The Japanese journal of psychology*, Vol. 68, No. 1, pp. 9–16, 1997.
- [105] Bronwyn Tarr, Jacques Launay, and Robin IM Dunbar. Silent disco: dancing in synchrony leads to elevated pain thresholds and social closeness. *Evolution and Human Behavior*, Vol. 37, No. 5, pp. 343–349, 2016.
- [106] Jonathan Gratch, Anna Okhmatovskaia, Francois Lamothe, Stacy Marsella, Mathieu Morales, Rick J van der Werf, and Louis-Philippe Morency. Virtual rapport. In *International Workshop on Intelligent Virtual Agents*, pp. 14–27. Springer, 2006.

付録

A. 研究実績

A.1 学術論文誌

- 1 Shogo Nishimura, Daiki Kimata, Wataru Sato, Masayuki Kanbara, Yuichiro Fujimoto, Hirokazu Kato, and Norihiro Hagita. “Positive Emotion Amplification by Representing Excitement Scene with TV Chat Agents,” *Sensors* published by Multidisciplinary Digital Publishing Institute (MDPI), 20(24), 7330, 2020. (4.2 章に関連)
- 2 Shogo Nishimura, Takuya Nakamura, Wataru Sato, Masayuki Kanbara, Yuichiro Fujimoto, Hirokazu Kato, and Norihiro Hagita. “Vocal Synchrony of Robots Boosts Positive Affective Empathy,” *Applied Sciences* published by Multidisciplinary Digital Publishing Institute (MDPI), 11(6), 2502, 2021. (4.3 章に関連)

A.2 査読付き国際会議・ワークショップ

- 1 Shogo Nishimura, Hiromichi Kawanami, Masayuki Kanbara, and Norihiro Hagita. “TV Chat Robots Allowing Daily-Use Chat with a User by Synchronizing TV Enthusiasm,” In *IEEE International Symposium on Robot and Human Interactive Communication (IEEE RO-MAN)*, 3rd Workshop on Groups in Human-Robot Interaction, 2017.
(<https://grouprobot.wordpress.com/2017submissions/>) [15 Mar, 2021 Accessed.] (4.2 章に関連)
- 2 Shogo Nishimura, Hiromichi Kawanami, Masayuki Kanbara, and Norihiro Hagita. “A TV Chat Robot with Time-Shifting Function for Daily-Use Communication,” In *International Conference on Social Robotics*, pp. 516-525, published by Springer, 2017. (3.2 章に関連)

- 3 Shogo Nishimura, Masayuki Kanbara, and Norihiro Hagita. “Atmosphere Sharing with TV Chat Agents for Increase of User ’ s Motivation for Conversation,” In International Conference on Human-Computer Interaction (HCI), pp. 482-492, published by Springer, 2019. (4.2 章に関連)
- 4 Shogo Nishimura, Takuya Nakamura, Masayuki Kanbara, Wataru Sato, and Norihiro Hagita. “Evaluation of Pacing for Dialog Robots to Build Trust Relationships with Human Users,” In Proceedings of the 7th International Conference on Human-Agent Interaction (HAI), pp. 300-302, 2019. (4.3 章に関連)