

音声対話インタフェースにおけるロボットエージェントの優位性の評価

玉川 聡¹ 山本 一公^{2,1} 中川 聖一¹

概要: 音声対話の発話対象は、ロボット(実物)と画面エージェント(仮想人物・生物)に二分できるが、話しやすさ・親しみやすさの観点から考慮したとき、身体性による特徴から実物のロボットの方が良いのではないかと考えられる。そこで、本研究では親しみやすいインタフェースであるペットロボット Phyno を話し相手とした対話システムを構築し、インタフェースとしての評価実験を行った。本システムのタスクは建物内の案内とし、設置場所を本学研究棟 1 階のエレベータ前と想定してシステムを設計した。発話対象の違いによるユーザの印象の比較をするため、発話対象をロボット(Phyno)、PC ディスプレイ、携帯ディスプレイとし、システムの印象の比較を行った。実験結果より、ロボットの身体性による利点を実証された。また、理解の補助として携帯端末を併用する効果や、ロボットがユーザの方を向くことによる効果の実証実験を行い、親しみやすさなどにおいて高評価を得ることができた。

キーワード: 音声対話システム, ロボット, エージェント, 発話対象, 目的指向対話

Evaluation of Superiority of Robot Agent in Spoken Dialog Interface

AKIRA TAMAGAWA¹ KAZUMASA YAMAMOTO^{2,1} SEIICHI NAKAGAWA¹

Abstract: The target of speech input can be an agent on the screen (virtual person or creature) or robot (real). When considered from the standpoint of easiness of talk and friendliness, a real robot would be better on the physicality. In this study, we built a dialog system that was set to talk with the pet robot “Phyno”. The task of this system is to guide rooms in the building, where the location is assumed at the front of the elevator on the first floor of our university building C. In order to compare the impression of the user due to the target’s devices in the utterance, we carried out the comparison of the impression for the system with the robot, computer display, and mobile display. From the experimental results, the benefit of physicality of the robot was demonstrated. We also designed the system to display the route on mobile devices and confirmed in the experiment the effect of the robot facing the user. In these experiments, it was possible to obtain a high evaluation rate in such friendliness.

Keywords: spoken dialog system, robot, agent, target device, task oriented dialog

1. はじめに

人間と対話を行うコンピュータやロボットの実現は、人工知能研究において大きな課題の一つである。特に 90 年代以降、計算機機能の飛躍的な向上により音声認識技術が発展し、コンピュータを話し相手とする多くの音声対話システムが構築された。現在では観光案内やカーナビゲーションシステムなど様々なシーンで音声認識技術が用いられ、将来これらのサービスの質はさらに向上すると思われる。しかし、音声対話においては、音声認識率などの音声処理性能とともに、ユーザの話しやすさ・使いやすさ・親しみやすさといったユーザフレンドリーさの観点も忘れて

はならない。

音声対話システムは使用目的により目的指向と非目的(雑談)指向に分けられる。目的指向では、認識精度の向上により音声入力の実用段階になり、その需要が高まってきた。また、スマートフォンの普及もこれを後押ししており、携帯端末向けの音声入力機能が開発された。これらの音声入力精度の向上により、いかに効率良く対話の目的を達成するか焦点が移り、最近では大量のコーパスに基づく POMDP 法 (Partially Observable Markov Decision Process)[1] などが提案されている。また、複数のドメインタスクを管理する方法や言語モデルの構成法 [2] などが研究されている。

対話システムによる発話対象は、ロボット(実物、ビジュアル型)と画面エージェント(仮想人物・生物、バーチャル型)に二分できる [3] が、ユーザフレンドリーさの観点から考

¹ 豊橋技術科学大学
Toyohashi University of Technology

² 豊田工業高等専門学校
Toyota National College of Technology

慮したとき、身体性による特徴から実物のロボットの方が良いのではないかと考えられる。発話対象に関する先行研究は [4], [5], [6], [7] などで行われている。

三宅らは、拡張現実感を用いた音声対話システムを構築し、画面のエージェントの有無によるユーザの発話の変化について調べた [4]。彼らはエージェントが存在する場合には命令形などが多用されることで、発話とのマッチングに単語の原形を利用でき、応答正解率が向上することを確認した。また、この結果よりエージェントの存在が、円滑に対話を進行する良い役割を果たしていると述べている。

角らは、ディスプレイによる人物のアニメーションと、顔面に人物のアニメーション画像を投影したロボットとの比較を行なっている [5]。被験者は「マス目色塗り課題をユーザが拒否するまで続ける」という作業を続けた結果、ディスプレイのエージェントだと課題を続けた回数が多く、思考力がアップした印象で説得力が高く、対して、ロボットの場合だと嫌味でない、優しい、謙虚な印象があったと報告され、モチベーションを維持するため説得を行う場面においてはキャラクターエージェントの方が影響力が強いと結論付けている。

田中らは、遠隔会議での対話において、対話相手の音声のみを聞かせる場合、画面上で動くエージェントを対話相手として示す場合、画面上で対話相手の実映像を写す場合、静止したロボットを対話相手とする場合、動作するロボットを対話相手とする場合、対話相手と直接対面した場合を比較し、実体を持つロボットと対話する場合に、画面上のエージェントと対話するよりも、遠隔会議の相手と対面して対話している感覚が強くなることを示している [7]。

また、奈良先端大学では、音声による駅周辺情報案内システム「キタちゃん」「キタちゃんロボット」を開発し、近鉄けいはんな線学研北生駒駅に設置した [8]。「キタちゃん」は CG エージェントが応答するがタッチパネル機能によりユーザからのウェブ上の情報の参照も可能としている。「キタちゃんロボット」はロボット型インタフェースを持ちユーザが親しみやすいシステムを目指している。両システムともサービス内容は基本的に同一で、駅や周辺の設備などの情報を一問一答形式でウェブ表示と合成音声を用いて回答していたが、ロボット型インタフェースである「キタちゃんロボット」が大人から子供まで幅広くのユーザに使用されたことがわかった。

海外の案内をするロボットの研究では、Autonomous Systems Lab の自律型移動対話ロボット「Robox」が挙げられる [9]。この「Robox」は、マルチモーダルな対話機能を有した自律型の移動ロボットであり、ナビゲーションロボットとして Swiss National Exhibition expo.02 で実際に運用された。ロボットの右目の搭載されるカメラにより人の追跡を行い、左目の LED ディスプレイにより記号やアイコンなどの表示を行う。また、ロボットの正面にあるボ

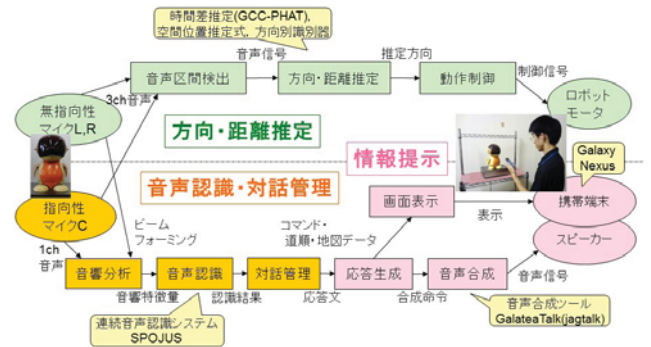


図 1 システムの流れ

タンで選択することで、四ヶ国語を話すことが可能であり、音声認識も四ヶ国語の Yes/No に該当する単語を認識できるようになっている。

本研究では、このような背景をふまえ、音声対話システムの構築を行う。研究の目的はユーザフレンドリーな(話しやすい・使いやすい・親しみやすい)システムの研究・開発であり、ユーザが対話を楽しみながらタスクを達成することを目標とする。本研究では親しみやすいインタフェースであるペットロボット Phyno[10] を話し相手とした対話システムを構築する。本システムのタスクは建物内の案内とし、設置場所を本学研究棟 1 階のエレベータ前と想定してシステムを設計した。

本稿では、本システムによる対話実験を行い、発話対象の違いやその動作による影響について分析した結果を報告する。

2. 対話システム

2.1 システムの概要

本システムは、ユーザの要求を満たす建物内の部屋へのルートを案内することをタスクとして想定している。設置場所は本学 C 棟 1 階のエレベータ前、対応する部屋は C 棟全体の研究室などを想定している。対応する文法は、挨拶、行きたい場所の提案・質問受付、問いに対する肯定・否定であり、語彙は 200 語程度となった。ただし、外部テキストファイルで対応する単語、文法、案内場所を簡単に追加できるよう実装しているので、理論上名前が完全に被らない限り、拡張は可能である。

また、ロボットと人間がより円滑に対話できるようにするため、話しかけた方向を向く方向推定機能も実装している。

本システムの構成は図 1 に示すように方向・距離推定機能、音声認識・対話管理機能、情報提示機能の 3 つに分けられる。システムの構成は本研究室の音声対話システム [11] を参考に行った。

2.2 インタフェース

本システムで用いているロボットは図 2 に示す株式会社



図 2 ペットロボット Phyno の外観と機能

ユウビ造形の Phyno[10] というロボットである。口には指向性マイクが搭載され、本体下のボックスにはスピーカー、そして各部位のモータにより首振り、手の上げ下げ、胴体ひねりといった動作が可能である。本システムでは口の指向性マイクを音声認識に用いている。また、方向推定をする場合、2チャンネルの音声で最低限必要なため、両耳に無指向性マイク (SONY ECM-88B) を追加で設置した。Phyno の動作制御は、コマンド受付サーバに対象モータと移動量のコマンドを送信する形で行う。

音声のみの案内では、ユーザがすぐに道順を理解するのは難しいのではないかと問題が挙がったため、理解の補助として携帯端末に道順を表示し、音声案内と連動した道順表示システムを設計した。使用するのは Samsung 製の Galaxy Nexus という Android OS 端末である。通信は無線 LAN 経由で行い、通信量を削減するため、現状では再生する動画は予め携帯端末側に用意しておき、再生コマンドのやりとりのみを行う。将来的に動画ストリーミングも可能であれば、サーバから携帯端末に動画を配信する。

2.3 方向推定・距離推定機能

方向・距離推定機能において、さらに細かく音声区間検出部、方向・距離推定部、動作制御部の3つに分かれる。まず、音声区間検出部においてユーザの発話を取り込む。この際、雑音にできるだけ反応しないよう、音声のパワーから音声区間検出 (VAD: Voice Activity Detection) を行い、音声区間のみを方向推定部に送る。方向・距離推定部では、音声の到達時間差 (TDOA: Time Difference Of Arrival) より、音の到来方向やユーザとの距離を推定する。最後に動作制御部で Phyno に取り付けられたモータに制御信号を送ることにより、Phyno が音声到来方向を向くことができる。

2.3.1 方向推定評価実験

方向推定では提案手法として、まず 3 マイクの TDOA を用いて粗く前後判別を行い、次に推定精度の良い両耳の 2 マイク (L,R) の TDOA による方向推定を行うという 2 段階の手法をとった。

実験に用いるデータは、防音室と実環境 (本学 C 棟 1 階のエレベータ前) で収録した。発話者は大学生・大学院生

表 1 距離別の前後判別正解率 [%]

環境	全方向の平均
防音室	98.7
実環境	94.9

表 2 条件別の方向推定正解率 [%](防音室)

	前方向	後方向	50cm	100cm	全方向/距離	
許容誤差	$\pm 15^\circ$	$\pm 15^\circ$	$\pm 15^\circ$	$\pm 15^\circ$	$\pm 15^\circ$	$\pm 30^\circ$
2ch	93.7	-	-	-	-	-
3ch	47.2	96.8	64.5	75.6	70.1	96.6
前後→方向	96.7	93.4	92.3	97.0	94.7	99.6

表 3 条件別の方向推定正解率 [%](実環境)

	前方向	後方向	50cm	100cm	全方向/距離	
許容誤差	$\pm 15^\circ$	$\pm 15^\circ$	$\pm 15^\circ$	$\pm 15^\circ$	$\pm 15^\circ$	$\pm 30^\circ$
2ch	84.5	-	-	-	-	-
3ch	43.3	74.5	56.8	58.5	57.7	87.0
前後→方向	82.8	81.3	77.4	86.3	81.8	93.6

の男性 6 名であり、発話文は IPA100 文 (毎日新聞) から選んだ 9 文を 3 セット (A、B、C) に分け、各セットを 2 人ずつに 30° 刻みの方向 ($-180^\circ \sim +180^\circ$ の 13 方向)、2 距離 (50cm、100cm) 別に発話データを収録した。各環境における収録データは計 468 発話文である。

方向推定の実験において、第 1 段階の 3 マイクの TDOA を用いた前後判別の正解率を表 1 に示す。また、第 2 段階である 2 マイク (L,R) の TDOA による方向推定の結果と他手法の結果の比較を表 2,3 に示す。表中の「前後→方向」と表記されたものが、3ch で前後方向を判別しその結果より 2ch の方向推定を行った結果である。「前方向」とは -60° から $+60^\circ$ 、「後方向」とは -180° から -90° と $+90^\circ$ から $+180^\circ$ 、「全方向」とは -180° から $+180^\circ$ の結果の平均値である。

結果より、前後の判定を組み合わせた時、防音室のとき最大許容誤差 $\pm 15^\circ$ で 95%、許容誤差 $\pm 30^\circ$ で 99% 以上の精度であり、実環境のとき許容誤差 $\pm 15^\circ$ で 82%、許容誤差 $\pm 30^\circ$ で 94% の精度を達成した。

2.3.2 距離推定評価実験

任意のユーザの距離を推定するのは困難なため、ユーザの発話位置とマイクとの距離が 50cm と 100cm のどちらであるかの 2 クラス識別問題とした。これにより、音量の調整、ユーザを近づくように要求、音響モデルの切り替えができる。

事前実験として、以下に示す音響的特徴量のうち、どの特徴量が距離識別に有用であるかを調べた。

- マイク L,R 間のポイント差 p_{LR}
- マイク L,C 間のポイント差 p_{LC}
- マイク R,C 間のポイント差 p_{RC}
- GCC-PHAT プログラム内で求まるマイク L,C 間の音声波形の相関のピーク値

表 4 距離識別 (50cm と 100cm) における識別率 [%]

学習・テストデータ	テスト前方向平均	テスト全方向平均
防音室	91.0	78.6
実環境	82.5	67.9

→ 距離が遠いほど残響の影響を受けて波形が歪むため、相関値が小さくなると予想した。

- GCC-PHAT プログラム内で求まるマイク L, R 間の音声波形 (前処理として音声に高域通過フィルタを通す) の相関のピーク値
- マイク L, R の音声パワー比
- 前処理として音声に低域通過フィルタを通したマイク L, C の音声波形のパワー比
- 前処理として音声に高域通過フィルタを通したマイク R, C の音声波形のパワー比
- $\frac{\text{マイク L の高域周波数と低域周波数のパワー比}}{\text{マイク R の高域周波数と低域周波数のパワー比}}$
→ 距離が長くなるほど音声の高域の方が低域よりも減衰するので、これを利用するために用いた。
- 上式の分子のみを用いたもの

その結果、各マイクの音声の到達サンプリングポイント差 p_{mn} のみを学習・テストデータに用いた場合が距離識別に一番精度良く距離推定ができることがわかった。

距離識別のため、事前に収録したデータより全方向に共通な 1 つの SVM (Support vector machine) 識別器を作成したが、60% 程度の結果しか出なかった。そこで、方向別に SVM 識別器を構成し、方向推定結果をもとに用いる SVM を選択し識別する手法をとった。距離識別において、SVM 識別器の学習・テストには SVM^{light} version 6.02[12] の RBF (Radial Basis Function) カーネルを用いた。

話者オープンの推定結果を出すため、前節と同じ収録データを学習用の 5 人分のデータ (390 発話文) とテスト用の 1 人分のデータ (78 発話文) に分け、交差検定を行った。その結果を表 4 に示す。方向の誤認識の対処、および任意の方向識別 (例えば 10° 刻み) の頑健性を検証するため、 30° 方向の場合は学習・テストデータともに 30° と 60° 方向のデータを用いる様にしており、他の方向も同様である。防音室の距離識別正解率は前方向 (-60° から $+60^\circ$) のみだと 91%、全方向で 79% であり、実環境では前方向 (-60° から $+60^\circ$) のみだと 83%、全方向で 68% であった。

今回の実験では、方向別の学習データが 30° 刻みしかなかったため、例えば 30° と 60° の学習データで識別器を学習したが、防音室の学習データを 15° 刻みにして識別器を学習させれば、よりよい性能が得られると考えられる。

2.4 音声認識・対話管理機能

音声認識・対話機能については、さらに細かく音響分析部、音声認識部、対話管理部に分けられる。

音声認識にとって音声区間検出と背景音排除は、非常に重要な要素である。そのため、精度向上のため認識に適する指向性マイクによる音声波形を用いる。

音響分析部では、音声のパワーから VAD が行われた後、音声区間から音響特徴量であるメル周波数ケプストラム係数 MFCC (Mel-Frequency Cepstral Coefficient) が算出され、音声認識部へ情報が送られる。

音声認識部には、本研究室で開発された文脈自由文法 (CFG: Context Free Grammar) 駆動型連続音声認識システム SPOJUS[13][14] を用いた。

対話管理部では、音声認識部から認識結果 (文字列情報) を受け取り、後述の応答生成部と連携し、状況に応じて応答文を生成する。

2.5 情報提示・音声合成機能

情報提示機能は、応答生成部、音声合成部、画面表示部に分けられる。

応答生成部では、前述の対話管理部と連携し、認識結果から状況に応じて応答文を生成する。

音声合成部では応答生成部より受け取った応答文から人工的な音声を生成することにより、Phyno の仮想音声を生成し音声応答する。本システムでは GalateaTalk[15] というオープンソースの日本語の音声合成エンジンの派生版である jagtalk[16] を用いている。

画面表示部では、音声応答の補助として画面の表示を行う。詳しくは次節で述べる。

3. 被験者実験

実験は全て実環境 (本学 C 棟 1 階エレベータ前) で行った。

3.1 発話対象の違いによるユーザの印象評価実験

話しかける対象別のユーザの印象の違いを話しやすさの観点から評価する目的で、実験を行った。ロボット (Phyno)、PC ディスプレイ、携帯端末のディスプレイの 3 つを比較した研究はまだないと思われる。

ロボットが発話対象の場合のシステムは今まで述べたシステムを用いた。携帯端末の場合、画面にはロボットの代わりとなるエージェントの対話内容に応じた動画を事前に収録し、それを適宜表示させた。PC ディスプレイの場合は、実物大相当のロボットのエージェントの動画を表示させた。対話中のジェスチャは各発話対象とも共通のものとする (ユーザの発話によるうなずき、案内中に手を上げるなど)。また、これら 3 つの対話システムは全てソケット通信を用いて情報のやりとりを行ない、PC ディスプレイ・携帯端末に関しては無線 LAN 経由でシステムと双方向通信を行う。

システムのタスクは本学 C 棟の建物案内とし、ユーザに課題として「言語関係の研究室の場所の情報をシステムか

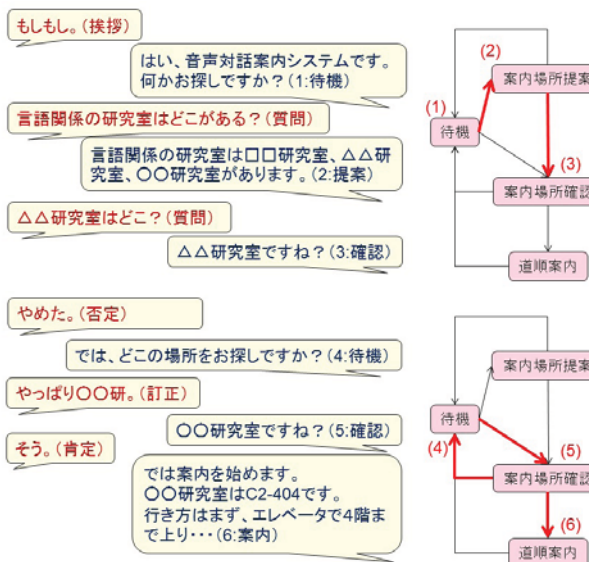


図 3 発話対象比較実験の対話例と対話状態



図 4 携帯端末を介した音声対話

ら聞き出してください。」などと課題を与え、自発的にシステムと対話するよう促した。また、本実験では音声認識用のマイクをヘッドセットに統一した。これは各発話対象の印象が音声認識率によって左右することを避けるためである。被験者には実験開始前にシステムの使用練習を一度だけ行ってもらった後、順番による影響が出ないように各発話対象と2回対話してもらい、その対話相手の順番はランダムにした。評価アンケートは全実験が終わった後、まとめて行った。

発話対象による印象の違いを比較するには、ある程度の対話のターン数が必要になると考えられる。そこで対話のターン数を増やすため、対話内容の拡張を行った。拡張設計した対話例とその状態遷移を図3に示す。図中において「案内場所提案」の状態を拡張することで、ユーザが知りたい情報の提案をシステム側から行うよう設計した。

携帯端末にロボットのエージェントが表示されている様子を図4に示し、各実験の様子を図5に示す。



図 5 発話対象別の対話実験の様子
 (上段:ロボット、中段:PC ディスプレイ、下段:携帯端末)

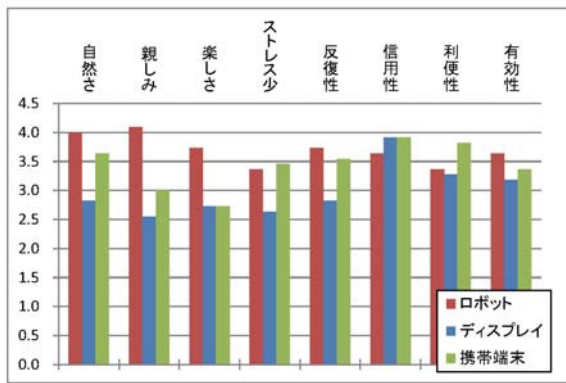
表 5 対話実験での評価指標 (5段階 絶対評価)

項目	質問内容
自然さ	対象に話しかけることに違和感があったか。
親しみやすさ	対象に親しみを持てるか。
楽しさ	対話をしていて楽しく感じたか。
ストレス少なさ	対話をしていてストレスを感じなかったか。
反復性	また使いたいと思うか。
信用性	その情報が信用できると思うか。
利便性	使ってみて便利だと感じたか。
有効性	この対話をすることに有効性を感じるか。

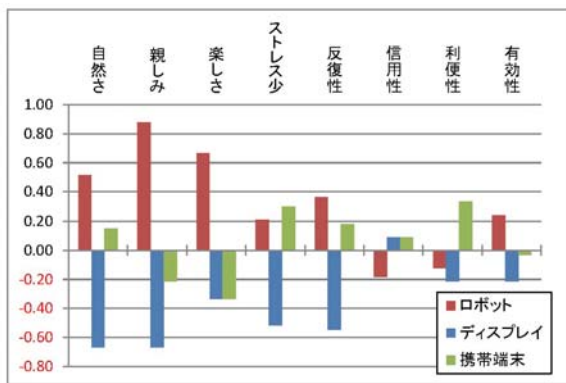
表 6 対話実験での評価指標 (相対評価)

項目	質問内容
自然さ	一番自然に話せたのはどれか。
親しみやすさ	一番親しみを持てたのはどれか。
楽しさ	対話をしていて一番楽しく感じたのはどれか。
ストレス少なさ	対話をしていて一番ストレスを感じなかったのはどれか。
反復性	また使いたいと思えるのはどれか。
信用性	一番信用できると思ったのはどれか。
利便性	使ってみて一番便利だと感じたのはどれか。
有効性	対話をすることに一番有効性を感じたのはどれか。

被験者のアンケートによる評価指標を表5、6に示す。



a) 実験結果



b) 相対値化した実験結果

図 6 発話対象の違いによるユーザの印象評価の実験結果

本実験は大学生・大学院生の男性 11 名で行った。アンケートの結果を図 6 に示す。図 6 の相対値は、以下のようにして求めた。

(1) ユーザの評価指標毎に平均値 \bar{x} を求める。

$$\text{例: } \bar{x} = \{(\text{ユーザ A の PC ディスプレイの自然さ}) + (\text{ユーザ A のロボットの自然さ}) + (\text{ユーザ A の携帯端末の自然さ})\} / 3$$

(2) 各評価とその平均値 \bar{x} の差を計算する。

$$\text{例: ユーザ A の PC ディスプレイの自然さの相対値 } x_{A,PC} = (\text{ユーザ A の PC ディスプレイの自然さ}) - \bar{x}$$

(3) 各ユーザの相対値の平均値を求める。

$$\text{例: PC ディスプレイの自然さの相対値} = \frac{1}{N} \sum x_{i,PC}$$

アンケートの結果より、ロボットは他の発話対象と比べ、自然さ、親しみやすさ、楽しさ、反復性に 1 名を除く全員がポジティブな印象を持ったことがわかり、特に自然さ、親しみやすさ、楽しさに関してその傾向が顕著であった。この結果より、研究の前提として考えた身体性によるメリットを実証することができた。しかし、ロボットの起動音・動作音にネガティブな印象を持つユーザも多かった。これは実物であるロボットにどうしても生じる制約であり、身体性によるメリットと動作音のデメリットがトレードオフの関係になっている。ただし、近い将来には、ロボット製作技術の向上によりこの音は解消できると思われる。

PC ディスプレイの場合、全体的にユーザがネガティブ

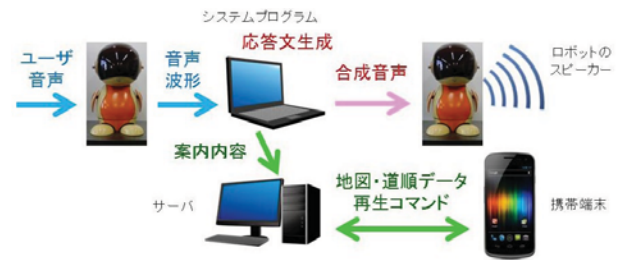


図 7 音声案内と連動した道順の表示

な傾向を持っており、特に自然さ、親しみやすさ、楽しさ、ストレスの少なさ、反復性にネガティブな印象を持つ傾向があることがわかった。ユーザの感想より、ロボットに比べ身体性に欠くことや、ディスプレイに実物のロボット動画が表示される不自然さが指摘されていた。今回、できるだけインターフェイス以外の条件を揃えるため、収録したロボットの動画をディスプレイに映したが、収録動画ではなく CG アニメーションにするとまた違う傾向が現れるかもしれない。

携帯端末の場合、反復性、利便性にポジティブな印象を 7 割以上のユーザが持っていたことがわかった。また、ストレスの少なさに関しては 3 つの発話対象の中で最も良かった。これは昨今のスマートフォンの普及により、携帯端末の操作・閲覧に対する違和感が軽減されたためと考えられる。将来、これらはさらに普及することが考えられるので、携帯端末の評価は今後さらに上昇するものと思われる。

3.2 対話内容理解の補助としての携帯端末併用の効果の実証実験

音声のみの対話では、ユーザが道順を理解するのは難しいのではないかという問題が挙げた。そこで、理解の補助として携帯端末に道順を表示し、音声案内と連動した道順表示システムを設計した。携帯端末のアプリケーションの開発はスマートフォン向けに行う。スマートフォンを用いる利点は、現在所持者が急速に増えているという機器の普及率、案内表示のまま持ち歩けるという可搬性、搭載されるセンサを用いて今後新しい機能を追加しやすいという拡張性の 3 点である。

本研究の対話システムの発話対象の主体はあくまでロボットであり、携帯端末の扱いは補助として地図の表示を行うだけであることに注意してもらいたい。比較用に実装した携帯端末を介した音声対話については前節で述べている。

音声と連動表示するための処理の流れを図 7 に示す。サーバと携帯端末間の通信は WebSocket[17] を用いて無線 LAN 経由で行う。

対話の流れは、案内場所の決定後、サーバから専用のアプリケーションの入ったユーザの携帯端末へ案内情報を送信し、事前に用意された動画を音声案内と同期して再生さ

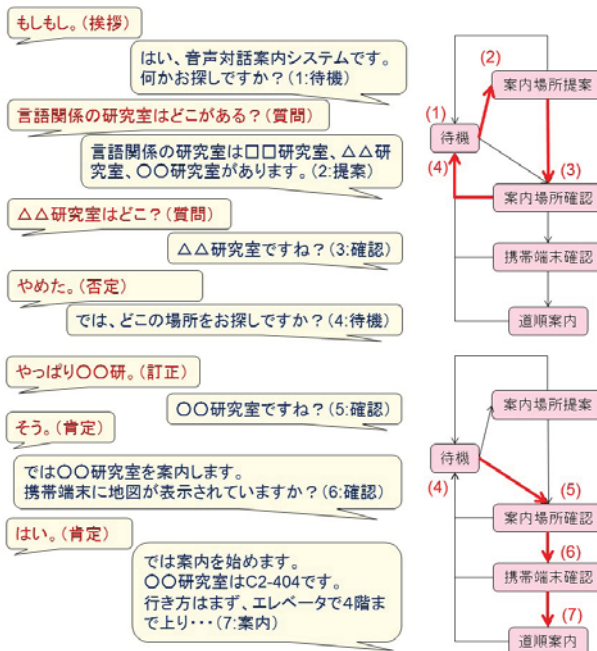


図 8 携帯端末併用実験の対話例と対話状態



図 9 携帯端末による地図表示

せるというものである。案内終了後、ユーザは画面を見ながら実際に目的地に移動してもらった。

実験の条件は前節と同様である。本実験を行うにあたり、前節の対話内容をさらに拡張した。拡張設計した対話例とその状態遷移を図 8 に示す。図中において「携帯端末確認」の状態を拡張し、ユーザに携帯のディスプレイの準備をするように促す。

また、携帯端末に地図が表示されている様子を図 9 に示す。

本実験は大学生・大学院生の男性 11 名で行った。実験後のアンケートは、3.1 節のロボットとの対話実験に対する相対評価として行った。実験の様子を図 10 に示す。また、被験者のアンケートによる評価指標とその結果を表 7



図 10 ロボットとの対話に携帯端末を併用した対話実験の様子

表 7 ロボットに携帯端末を併用した対話実験での評価指標 (相対評価) とその結果

項目	質問内容	平均値
自然さ	対象に話しかけることの違和感は減ったか。	3.1
親しみやすさ	さらに対象に親しみを持てたか。	3.6
楽しさ	対話をしていてさらに楽しく感じたか。	3.5
ストレス少なさ	対話をしていてストレスを感じにくくなったか。	3.5
反復性	さらにまた使いたいと思ったか。	4.3
信用性	その情報がさらに信用できると思ったか。	4.0
利便性	使ってみてさらに便利だと感じたか。	4.3
有効性	この対話をすることにさらに有効性を感じたか。	4.0

に示す。アンケートの評価値は、携帯端末の併用により、1:「悪くなった」、2:「少し悪くなった」、3:「変わらない」、4:「少し良くなった」、5:「良くなった」とした。

結果より、ユーザが特に反復性、利便性において高い評価をしていることがわかった。被験者の感想より、複数のデバイスが連動していることがこの高評価の要因になったようである。

3.3 ロボットの方向転換による効果の実証実験

対話開始時に、ロボットがユーザの方向を向く (実際は方向推定により話しかけられた方向を推定することによる効果を調べるための実験を行った。本実験では、対話開始前にロボットをユーザがいる方向とは別の方向を向けさせ、ユーザが話しかけるとロボットがユーザの方向を向く。対話内容は 3.1 節の実験と同様である。図 11 にロボットの方向転換前後の様子を示す。

本実験は大学生・大学院生の男性 6 名で行った。実験後のアンケートは、3.1 節のロボットとの対話実験 (ロボットがユーザの正面に向いている状態で対話開始) に対する相対評価として行った。被験者のアンケートによる評価指標とその結果を表 8 に示す。アンケートの評価値は、ロボットが方向転換することにより、1:「悪くなった」、2:「少し悪くなった」、3:「変わらない」、4:「少し良くなった」、5:「良くなった」とした。

結果より、ユーザが特に親しみやすさ、楽しさにおいて高い評価をしていることがわかった。被験者の感想より、



a) ユーザが話しかける前



b) ユーザが話しかけた後

図 11 ロボットの方向転換の様子

表 8 ロボットが方向転換する対話実験での評価指標 (5段階 相対評価) とその結果

項目	質問内容	平均値
自然さ	対象に話しかけることの違和感は減ったか。	3.5
親しみやすさ	さらに対象に親しみを持てたか。	4.5
楽しさ	対話をしていてさらに楽しく感じたか。	4.0
ストレス少なさ	対話をしていてストレスを感じにくくなったか。	3.5
反復性	さらにまた使いたいと思ったか。	3.8
信用性	その情報がさらに信用できると感じたか。	3.5
利便性	使ってみてさらに便利だと感じたか。	3.8
有効性	この対話をするのにさらに有効性を感じたか。	3.7

ロボットがユーザの方向を向くことで、より人間らしい動作に近づいたように感じたことがこの高評価の要因になったようである。今後、さらに人間らしい動作をするにはどうすれば良いかを検討していきたい。

4. まとめ

本稿では、親しみやすいインタフェースであるペットロボット Phyno を話し相手とした対話システムを開発し、インタフェースとしての評価実験を行った。本対話システムでは建物内の案内を通して、ユーザフレンドリーな対話システムの構築を図った。発話対象の違いによるユーザの印象の比較をするため、発話対象をロボット (Phyno)、PC ディスプレイ、携帯ディスプレイとしたシステムの印象の比較を行い、実験結果よりロボットの身体性による利点を実証できた。また、理解の補助として携帯端末を併用する効果や、ロボットがユーザの方を向くことによる効果の実証実験を行い、親しみやすさなどにおいて高評価を得ることができた。

今後の発展として、ディスプレイに表示するエージェンツの違い (動画やアニメーション) による印象の違いの調査、ロボットをより人間のように動かせる (対話の違和感を低減する) ための手法の調査などが挙げられる。また、音声の割り込み入力 (バーズイン) の対応など、対話機能の向上も行なっていきたい。

参考文献

- [1] Jason D. Williams and Steve Young. Partially observable Markov decision processes for spoken dialog systems. *Comput. Speech Lang.*, Vol. 21, No. 2, pp. 393–422, 2007.
- [2] 森祥二郎, 駒谷和範, 佐藤理史. 大規模コーパスへのクラス付与に基づく音声対話システム用言語モデルの構築. 情報処理学会研究報告. SLP, 音声言語情報処理, Vol. 2012, pp. 1–8, 2012.
- [3] 徳田英幸. ユビキタスコンピューティング環境の進化とネットワークロボット ~スマートフォン, クラウド, IoT, スマートシティとの連携~. 情報処理学会誌, Vol. 54, No. 7, pp. 686–689, 2013.
- [4] 三宅真司, 伊藤彰則. 拡張現実感を用いた音声対話エージェンツの評価. 日本音響学会研究発表会講演論文集, 2012.
- [5] 角薫, 長田瑞恵. モチベーション維持におけるキャラクターエージェンツとロボットの比較. HAI シンポジウム 2011, 2011.
- [6] 志和敏之, 奥野佑将, 神田崇行, 今井倫太, 石黒浩, 萩田紀博. コミュニケーションロボットによる道案内ジェスチャーの有用性と発話タイミングのモデル化. 情報通信学会論文誌 D Vol.J95-D No.10 pp.1818-1828, 2012.
- [7] 田中一品, 中西英之, 石黒浩. 実体で身体動作を提示するロボット会議によるソーシャルテレプレゼンスの強化. インタラクシオン 2013, pp. 125–132, 2013.
- [8] 川波弘道, 木田学, 早川直樹, ツインツアレクトピアス, 北村任宏, 加藤智之, 鹿野清宏. 駅構内音声対話システム「キタちゃん」「キタちゃんロボットの開発」. 電子情報通信学会技術報告, SP2006-14, 2006.
- [9] Roland Siegwart, Kai O. Arras, Samir Bouabdallah, Daniel Burnier, Gilles Froidevaux, Xavier Greppin, Björn Jensen, Antoine Lorotte, Laetitia Mayor, Mathieu Meisser, Roland Philippsen, Ralph Piguët, Guy Ramel, Gregoire Terrien, and Nicola Tomatis. Robox at Expo.02: A Large-Scale Installation of Personal Robots. *Robotics and Autonomous Systems*, Vol. 42, pp. 203–222, 2003.
- [10] Phyno. <http://www.yuvi.co.jp/robot/phyno.htm>.
- [11] 西村良太. 人間同士の対話現象を模倣した音声対話システムの研究. 2007.
- [12] SVM-Light Support Vector Machine. <http://svmlight.joachims.org/>.
- [13] 甲斐充彦, 中川聖一. 日本語連続音声認識システム SPOJUS-SYNO の改良と評価. 電子情報通信学会技術報告 SP93-20, 1993.
- [14] SPOJUS-SYNO. <http://www.slp.cs.tut.ac.jp/SPOJUS/>.
- [15] GalateaTalk プロジェクト. <http://sourceforge.jp/projects/galateatalk/>.
- [16] jagtalk [ja.nishimotz.com]. <http://ja.nishimotz.com/project:jagtalk>.
- [17] WebSocket. <http://ajf.me/websocket/>.