

コミュニケーションロボットの交替潜時と動作が対話の円滑性に与える影響

朱 凱[†] 北村 達也^{††} 梅谷 智弘^{††}

[†] 甲南大学大学院自然科学研究科知能情報学専攻 〒658-8501 兵庫県神戸市東灘区岡本 8-9-1

^{††} 甲南大学知能情報学部 〒658-8501 兵庫県神戸市東灘区岡本 8-9-1

E-mail: [†]m2424006@s.konan-u.ac.jp, ^{††}t-kitamu@konan-u.ac.jp

あらまし 本研究では、コミュニケーションロボットとの対話において、交替潜時(ロボット発話後に音声認識が起動するまでの時間)の長短と発話を促す動作の有無が対話の円滑さに及ぼす影響を調査した。交替潜時と動作の有無を組み合わせた4条件で対話実験を行い、音声認識失敗の回数を測定した。その結果、交替潜時と動作の有無によって認識失敗の回数が増減し、これらがヒトとロボットとの対話の円滑性に影響を及ぼすことが示された。

キーワード 対話ロボット, 話者交替, 音声合成, 音声認識

Effects of turn-taking latency and behavior of communication robot on dialogue smoothness

Kai ZHU[†], Tatsuya KITAMURA^{††}, and Tomohiro UMETANI^{††}

[†] Graduate School of Natural Sciences, Konan University, 8-9-1 Okamoto, Higashinada-ku, Kobe, Hyogo 658-8501, Japan

^{††} Faculty of Intelligence and Informatics, Konan University, 8-9-1 Okamoto, Higashinada-ku, Kobe, Hyogo 658-8501, Japan

E-mail: [†]m2424006@s.konan-u.ac.jp, ^{††}t-kitamu@konan-u.ac.jp

Abstract In this study, we investigated effects of the alternation latency (the time it takes for speech recognition to start after the robot speaks) and a motion to prompt speech on the smoothness of a conversation with communication robot. We conducted a dialogue experiment under four conditions combining the alternation latency and the presence or absence of motion, and measured the number of speech recognition failures. The results showed that the number of recognition failures changed depending on the alternation latency and the presence or absence of motion, and that shown that these factors affect the smoothness of dialogue between humans and robots.

Key words Dialogue robot, turn-taking, speech synthesis, speech recognition

1. はじめに

近年、日本社会では少子高齢化や労働力不足が深刻化しており、これらの課題に対応するための新たな手段として、コミュニケーションロボットの活用が注目されている。特に、日常的な対話を通じて人間と自然なやり取りができるロボットは、学習支援、接客、教育、介護など多岐にわたる分野において大きな可能性があり、今後の社会において重要な役割を果たすことが期待されている。

こうしたコミュニケーションロボットには、単なる機械的な応答にとどまらず、より人間らしい対話の実現が求められている。そのためには、人間との対話の中で適切なタイミングで発話を切り替える「交替潜時」の自然さが重要な要素となる。先

行研究[1][2]によれば、交替潜時の自然さに影響を与える要素として、「交替潜時」や「ロボットの動作」が挙げられている。これらの要因を適切に制御することで、ロボットとの対話における違和感を軽減し、よりスムーズなコミュニケーションを実現することが可能となる。

本研究では、既存のコミュニケーションロボットシステム[3]を用いて、ロボットの発話終了から音声認識が可能になるまでの時間を「交替潜時」と定義し、この認識潜時が対話に及ぼす影響について調査を行う。また、ロボットが腕を挙げるなどの発話交替を示すシグナル(ロボット動作)[4]が対話に与える影響も調査する。著者らの研究[2]では初めは参加者数が不足していたため、実験の精度と信頼性を高めるために追加の参加者を増やし、より信頼性の高いデータを収集する。

また、対話交替中で身ぶりの中でも行っている身ぶりが発話予備動作を有するため話者交替において重要であると考え[5]. 対話のスムーズさも主に非言語的な要素に依存していることが示されている[6]. したがって、ロボットが適切に非言語的な情報を提供することで、対話の質が向上する可能性が高いと考えられる。

また、多様なロボット動作を活用することで、よりスムーズにな対話を実現する身体的コミュニケーション技術の研究も進んでる[5][7].

本研究では、ロボットとの対話における「交替潜時」の有無、およびロボットの「動作」の有無が、対話の円滑さにどのような影響を与えるかを明らかにする。特に、音声認識の成功率や対話の自然さに焦点を当て、これらの要因がどのようにコミュニケーションの質に関わるのかを検討する。最終的に、これらの要因を改善することで、より自然でストレスの少ない人間とロボットとの対話を実現し、コミュニケーションロボットが社会で広く活用されるための基盤を築くことを目指す。

2. 研究方法

2.1 対話ロボットシステム

本研究で使用した対話ロボットシステム[3]は、卓上型コミュニケーションロボット(ヴイストーン Sota[8]), 制御ソフトウェア(Vstone Magic), および音声認識・音声合成システム(Microsoft Azure)から構成された。

使用したロボットは、280(H)×140(W)×160(D)mm, 重量800g, 合計8自由度(胴体1軸, 腕2軸×2, 首3軸)を有する。また、カメラ, モノラルマイク, スピーカー, LED(両目×2, 口×1, 電源ランプ×1), スイッチ(電源ボタン×1, 音量ボタン×2)を搭載している。CPUにはIntel Edisonが搭載されている。

本研究では、使用したロボットには音声認識および音声合成機能が搭載されているが、本研究では、より高性能なMicrosoft Azureの音声認識・音声合成機能をノートPC上で実行した。通信状況が音声合成に影響を及ぼす可能性を考慮し、ロボットの発話にはリアルタイム合成を使用せず、事前に音声を合成してWAVファイルとして保存し、それを再生する方式を採用した。

ロボットの動作は、ノートPCからUDP通信を介して、ロボットに登録された指令を呼び出すことで実行した。参加者の音声はノートPCに接続したマイクで認識し、ロボットの音声はノートPCのスピーカーから出力した。ロボットを起動させるため、キーボードをロボットの右側に配置した。



図1 実験の状態

2.2 実験参加者データ

実験参加者は28名で、内訳は男性25名、女性3名、平均年齢は22.5歳であった。すべて甲南大学の学部生と大学院生である。

2.3 タスク

参加者には、以下の説明でロボットと対話するよう教示した。

- あなたは甲南大学の学生であり、甲南大学の施設について詳しく知りたいと考えています。
- ロボットは、「図書館」「食堂」「ファミリーマート」「ジム」の4つの施設に関する質問に答えることができます。
- この4施設の中から2施設を選び、それぞれの施設について3つの質問を選択し、合計6つの質問を行ってください。
- あらかじめiPadで6つの質問を選び、そのiPadを実験室に持ち込むことができます。
- テーブル上のキーボードのエンターキーを押すとロボットが起動し、「こんにちは」と発話します。ユーザも「こんにちは」と返してください。最初は簡単な対話から始まります。
- 選んだ質問をロボットに聞いてください。ロボットが質問を認識しない場合は、もう一度同じ質問をしてください。何度か試しても反応がない場合は、次の質問に進んでください。
- 対話を終えたいときは、「ありません」「もういい」「ない」「いいえ」「大丈夫」などの発話をしてください。
- 対話が終わったら、実体験に基づいたアンケートへのご記入をお願いします。

質問リストには、図書館に関しては「開館時間」や「場所」など、食堂に関しては「値段」や「おすすめメニュー」など、ファミリーマートに関しては「支払方法」や「利用できるポイントカード」など、ジムに関しては「場所」や「利用できるマシン」など、施設別に計32個の質問を記載した。実験中に質問内容を忘れないよう、テーブル上に実験の質問が表示されているiPadを設置し、参加者から見える状態で進行した。実験の様子はカメラ(DJI Osmo Pocket 3)で記録した。本研究で使用した動画は、すべて2K解像度(2048×1080ピクセル)、60フレーム/秒(fps)で撮影されたものである。各動画のファイルサイズは約800MB~1.1GBの範囲である。実験環境の全体像は図1に示す。

質問リストは以下に示す：

(1) 図書館に関する質問:

- 開館時間は何時ですか？
- 場所はどこですか？
- 本は何冊ありますか？
- 何日間借りることができますか？
- 何冊まで借りることができますか？
- 自習(勉強)できる場所はありますか？
- コンセント(電源)はありますか？
- プリンターはありますか？
- Wi-Fiは使えますか？
- ノートパソコンを借りることはできますか？
- パソコンを使える場所はありますか？

(2) 食堂に関する質問

- 営業時間は何時ですか？
- 場所はどこですか？
- おすすめのメニューは何ですか？
- 座席数はどのくらいありますか？
- どんな料理がありますか？
- 値段はいくらですか？
- 支払い方法を教えてください。
- 混雑する時間帯はいつですか？

(3) ファミリーマート（コンビニ）に関する質問

- 営業時間は何時ですか？
- 場所はどこですか？
- どんな商品がありますか？
- 支払い方法を教えてください。
- ポイントは使えますか？
- ATM はありますか？
- コピー機はありますか？
- プリンターはありますか？

イトインスペースはありますか？

(4) トレーニングルーム（ジム）に関する質問

- 営業時間は何時ですか？
- 場所はどこですか？
- どんなマシンがありますか？
- 誰が利用できますか？（利用対象者は誰ですか？）

2.4 実験条件

実験は表 1 の 4 条件にて実施した。各条件に 7 人ずつ参加させた。

表 1 実験の 4 条件

	交替潜時無	交替潜時有
ロボット動作有	A	C
ロボット動作無	B	D

本研究では、交替潜時とは、ロボットの発話が終了した後に音声認識が起動し、ユーザが発話可能になるまでの時間を指す。交替潜時付き音声は Azure で合成された音声を無加工のまま使用しており、各発話の末尾に約 1 秒の無音区間が存在する（図 2 参照）。一方、条件 A および B では、音声編集ソフトウェア Audacity を用いて、音声の末尾に含まれる無音区間を手動で削除した。

ロボットは neutral 状態で：図 3 に示すように、ロボットは両腕を下に構えた状態で静止しており、ユーザとのインタラクションの開始を待っている。この姿勢は、視覚的にユーザに対して「待機中」であることを明示する役割を果たす。LED リングの点灯（緑色）は、システムがアクティブであり、対話準備が整っていることを示している。一方、図 4 に示すように、Sota の目が青色に発光している状態は、「待機状態」または「音声入力の認識中」を意味する。これはユーザーの発話を待っていることを示す非言語的フィードバックである。同図では、ロボットが右手を挙げている姿勢が確認される。また、ロボットの頭部はやや左方を向いた。

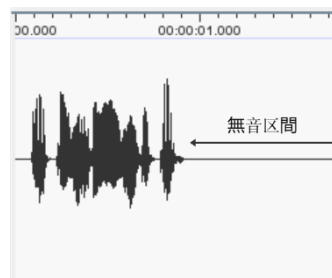


図 2 Azure で作成された音声の例。音声データには、発話終了後に約 1 秒間の無音区間を付加されている。



図 3 neutral 状態



図 4 question 状態

2.5 実験手続き

実験開始前に、参加者には実験内容の説明を行い、内容を理解した上で同意書に署名してもらい、参加の意思を確認した。その後、参加者を高さ 0.7m のテーブル上に設置されたロボットの前に着席させた（ロボットとの距離は約 0.55m）。その状態で、2.3 節に示した質問項目に基づき、約 3 分間ロボットとの対話を行わせた。対話終了後、参加者には Google Forms を用いてアンケートに回答してもらった。アンケートは 5 段階リッカート尺度による多肢選択式と自由記述式を組み合わせた形式で実施した。

2.5.1 アンケート

- (1) ロボットと対話中で何回対話が途切れましたか？
(0 回, 1 回, 2 回, 3 回, 4 回, 5 回以上)
- (2) 対話中でロボットの反応スピードがどうですか？
(1: 早かった～5: 遅かった)
- (3) ロボットから必要な情報を得られますか？
(1: 得られた～5: 得られなかった)

- (4) ロボットとの対話はスムーズにいましたか？
(1: いけた～5: いけなかった)
- (5) ロボットと対話についてどう思いますか？
(1: 分かりやすかった～5: 分かりにくかった)
- (6) ロボットと対話中で違和感を感じましたか？
(1: すごく感じた～5: 全然感じなかった)
- (7) 今回の実験ようなロボットサービスがあれば使いたいですか？
(1: 使いたい～5: 使いたくない)
- (8) その他、気づいた点があればご記入をお願いします (自由記述)

3. 実験結果

3.1 交替潜時と動作の影響

ロボットが交替潜時中に実験参加者が発話を始めると音声認識が失敗となる。各条件で実験参加者の音声認識が失敗となった回数の平均値および標準偏差を表2に示す。参加者がロボットに質問した平均回数も下の表に示されている。

表2 音声認識の失敗となった回数の平均値と標準偏差と実験参加者がロボットに質問した平均回数 (回)

条件	平均値	標準偏差	質問した平均回数
A	1.00	1.15	7.00
B	1.43	0.98	7.57
C	2.00	3.21	11.71
D	4.86	3.63	18.57

本調査では、参加者がロボットに対して質問した平均回数を比較した。参加者がロボットに質問した回数に対して、各グループの平均回数は、Aが7.00回、Bが7.57回、Cが11.71回、Dが18.57回であった。この結果から、グループAおよびBは比較的少ない回数でロボットとのやり取りを行っており、スムーズなコミュニケーションが実現されていた可能性がある。一方、グループCおよびDでは質問回数が多く、特にDは他のグループと比べて顕著に高い平均回数を示している。

表2に関して、一元配置分散分析(ANOVA):4つの条件(A～D)の間には統計的に有意な差があった $F(3, 24) = 5.93, p = 0.0035$ 。つまり、条件によって測定された数値の平均に違いがある可能性が高い。

(1) 条件A(交替潜時なし、ロボット動作あり)では、平均失敗回数は1.00回であり、条件B(交替潜時なし、ロボット動作なし、平均1.43回)よりもやや少ない結果となった。しかし、実験の前半では認識失敗が発生しており、これは参加者がロボットの動作に慣れていなかったことが要因と考えられる。その後、参加者がロボットの動作に慣れるにつれて、失敗回数は減少した。条件Cおよび条件D(交替潜時あり)に比べて、全体的に対話はより円滑であった。

(2) 条件B(交替潜時なし、ロボット動作なし)では、平均1.43回で比較的少ない失敗回数が記録されており、条件Cおよび条件Dと比べて対話交替がスムーズに進んだと考えられる。

(3) 条件C(交替潜時あり、ロボット動作あり)では、平均2.00回の失敗回数があり、条件D(交替潜時あり、ロボット動作なし、平均4.86回)と比較して失敗回数は少なかった。全体的に失敗回数は5回未満が多かったが、5回以上のケースも見られた。ロボットの動作があったことにより、参加者は対話のタイミングをつかみやすくなったと考えられる。

(4) 条件D(交替潜時あり、ロボット動作なし)では、平均失敗回数が4.86回であり、他の全条件の中で最も高い数値を示した。特に5回以上の失敗が見られ、10回以上失敗するケースも存在した。これはロボットの発話後に認識潜時があるため、参加者がすぐに反応すると音声が入らずに認識されないケースが多く見られたことに起因する。

3.2 対話の例

条件C(交替潜時あり、ロボット動作あり)における対話内容およびロボットの動作を以下に示す。ここで、Rはロボット、Uは参加者を表す。認識に失敗した場合には、「FALSE」と記載した。また、ロボットの動作には2種類あり、「neutral」(図3)と「question」(図4)である。

R: こんにちは。

U: こんにちは。

R: 私は Sota.

あなたのお名前はなんですか?(question)

U: ○○です。

R:(neutral) ありがとうございます。本日はよろしく申し上げます。では、あなたが知りたい大学の施設の名前を教えてください。(question)

U: 図書館です。

R:(neutral) はい。わかりました。図書館についての質問を受け付けます。なんでも聞いてくださいね。(question)

U: 場所は。「FALSE」場所はどこですか？

R:(neutral) 図書館は西門に入ってすぐ右手側にあります。学生証がないと利用できないので気を付けてくださいね。他に図書館についての質問はありますか?。(question)

U: ですか? 「FALSE」

R:(neutral) 他に図書館についての質問はありますか?。(question)

U: 時間を教えてください。

R:(neutral) 平日は9時から21時、土曜日は9時から18時まで空いています。日曜日は基本的に閉まっています。他に図書館についての質問はありますか?。(question)

U: ええ。混雑状況。「FALSE」

R:(neutral) 他に図書館についての質問はありますか?。(question)

U: コンセントありますか？

R:(neutral) 図書館の地下1階、1階、2階のそれぞれの階にコンセントがある机がいくつかあります。特に、地下1階には多くあります。他に図書館についての質問はありますか?(question)

U: です。ないです。？

R:(neutral) わかりました。他に知りたい大学の施設はありますか?(question)

U: 食堂.

R:(neutral) はい。わかりました。食堂についての質問を受け付けます。何でも聞いてくださいね。(question)

U: おすすめを教えてください。

R:(neutral) おすすめのメニューはネギトロ丼です。ネギトロ丼を食べるためだけに学校に来る人もいます。他に食堂についての質問はありますか?(question)

U: 支払い方法を教えてください。

R:(neutral) 現金または生協の電子マネーが利用できます。他に食堂についての質問はありますか?(question)

U: 何時から混雑状況になるんですか？

R:(neutral) 平日は 10 時 30 分から 15 時 30 分まで営業しています。土日は基本的に営業していません。他に食堂についての質問はありますか?(question)

U: いです。「FALSE」

R:(neutral) 他に食堂についての質問はありますか?(question)

U: ない。

R:(neutral) わかりました。他に知りたい大学の施設はありますか?(question)

U: ありません。

R:(neutral) わかりました。本日はご利用ありがとうございました。実験を終了します。(question)

3.3 アンケートの結果

対話の流暢さを相対的に反映した参加者の回答を以下の表 3 と表 4 と表 5 に示す。

表 3 ロボットと対話中で違和感を感じましたか

評価	条件 A	条件 B	条件 C	条件 D
1 (good)	0	0	1	0
2	1	1	1	1
3	2	3	0	1
4	3	2	3	4
5 (bad)	1	1	2	1

表 4 対話中でロボットの反応スピードがどうですか

評価	条件 A	条件 B	条件 C	条件 D
1 (good)	2	3	0	2
2	2	3	3	2
3	3	0	4	2
4	0	1	0	1
5 (bad)	0	0	0	0

表 5 ロボットとの対話はスムーズにいけましたか

評価	条件 A	条件 B	条件 C	条件 D
1 (good)	4	3	3	2
2	2	2	4	1
3	1	1	0	1
4	0	1	0	3
5 (bad)	0	0	0	0

本研究では、先行研究 [2] と比較して参加者数を 28 名に増加させた。人数の増加によりデータの信頼性が向上し、評価の精度も高まった。また、条件 A および B は、条件 C および D と比較して全体的に高い評価を得た。

4. 実験結果および考察

本研究では、コミュニケーションロボットとの対話交替において、音声の交替潜時(発話終了後に 1 秒間の無音区間)とロボット動作の有無が、人とロボットの対話のスムーズさに与える影響を調査した。過去に著者らが行った研究 [2] と比較して、参加者数を増加させた。

その結果、交替潜時が存在する場合、対話のスムーズさが著しく阻害され、対話中に参加者に悪い印象を与える傾向が見られた。一方、ロボット動作は対話中に参加者へ交替タイミングを示唆する役割を果たしていることがわかった [9]。ロボットの動作で、参加者はテストの後半でミスが少なくなった。

また、実験参加者はロボットの動作を音声認識開始の標識として捉える可能性があることも示唆された [10]。さらに、音声に交替潜時が伴うことで対話がスムーズに進まない原因をより深く考察すると、参加者が日常生活で高速かつ途切れのない対話に慣れていることが一因であると考えられる。

5. おわりに

本研究では、音声提示後の認識潜時およびロボットの動作が対話交替に与える影響について調査を行った。結果から、認識潜時が人とロボットの対話交替のスピードに遅れをもたらすことが分かったが、ロボットの動作によってその遅延がやや軽減されたことも確認された。ただし、効果の差は小さく、有意とは言い難かった。今後は、より多様な影響要因を加えて検討を続けたい。

謝 辞

本研究の一部は、日本学術振興会科学研究費 (JP20H04287, JP22K04020)、甲南大学 KONAN プレミア・プロジェクト、甲南大学デジタルツイン研究所および私立大学等経常費補助金特別補助「大学間連携等による共同研究」の支援を受けた。

文 献

- [1] 大藤聖菜, 馮建美, 小山大幾, 今井倫太, “人間とロボットの会話における「間」のデザイン,” 情報処理学会第 79 回全国大会講演論文集, 243–244, 2017.
- [2] 櫻井裕真, 榎本佐知子, 北村達也, 梅谷智弘, “コミュニケーションロボットとの対話における交替潜時および交替シグナルの影響,” 信学技報 CNR2023-27, 8–12, 2024.
- [3] 梅谷智弘, 的場瞳, 榎本佐知子, 陸鳴宇, 北村達也, “クラウド音声 API を利用した図書館ヘルプデスク支援ロボットの試作,” 日本機械学会ロボティクス・メカトロニクス講演会, 2022.
- [4] 横山真男, 青山一美, 菊池英明, 帆足啓一郎, 白井克彦, “人間型ロボットの対話インタフェースにおける発話交替時の非言語情報の制御,” 情報処理学会論文誌, 40(2), 487–496, 1999.
- [5] 長谷川孔明, 中内靖, “テレプレゼンスロボットによる無意識的身ぶりの表出が発話交替に与える影響,” 日本機械学会論文集, 80(819), 321–331, 2014.
- [6] 横山真男, “人間型ロボットの対話インタフェースにおける発話交替時の非言語情報の制御,” 情報処理学会論文誌, 40(2), 487–496,

2019.

- [7] 齋圭作, 加藤宏脩, 阿部秀尚, “音声対話ロボットとリアルタイム画像提示による調理支援システムの開発,” 情報処理学会第 80 回全国大会講演論文集, 189–190, 2018.
- [8] 塩見昌裕, 大和信夫, 前田武志, 横山智彰, 深津将生, 今川拓郎, 石黒浩, “テーブルトップ型対話ロボットプラットフォーム「Sota (ソータ):Social Talker」の開発,” 第 33 回日本ロボット学会学術講演会予稿集, 3G3–06, 2015.
- [9] 市川熹, 川端良子, 菊池英明, 堀内靖雄, 黒岩眞吾, “母語 (日本語) における話者交替タイミングの獲得,” 言語・音声理解と対話処理研究会 SIG-SLUD-B403-03, 14–19, 2015.
- [10] 朱怡潔, “主導権の交替から見た話題開始の談話標識,” 言語科学論集, 24, 41–50, 2020.