

音声対話インタフェースにおける マルチタスク性の適切な表出方法の検討

INVESTIGATION OF APPROPRIATE REPRESENTATION FOR MULTIPLE TASKS IN SPOKEN DIALOGUE INTERFACE

小中 彩貴^{1*} 李 晃伸¹
Saki Konaka¹ Akinobu Lee¹

¹ 名古屋工業大学
¹ Nagoya Institute of Technology

Abstract: This study aims to improve affableness of spoken dialogue interface. In this paper, we focus on that many spoken dialogue interface have multitask. First, we investigate the relationship between affableness and expression such as icons and agents. In addition, we evaluate the multiple agent interface that set agents for each tasks.

1 はじめに

近年，音声を用いたインタフェースの研究が盛んに行われている．音声認識精度の向上に加え，人工知能分野から音声対話に関する研究が進められているためである．最近では，スマートフォンに搭載されている音声アシスタントなどのクラウドコンピューティングによるサービスにも後押しされ，音声をインタフェースとして利用する機会が増加している．

本研究では，音声を用いたインタフェースの中でも「音声対話インタフェース」に特に注目する．自然な発話による機械との簡単な対話によって単純なタスクを達成し，ユーザが気軽に使用できることを主眼とする．

現状の音声対話インタフェース普及の壁として、「話しかけにくさ」が挙げられ，解消することが音声インタフェース全体の普及に繋がると考えられる．音声対話インタフェースは本質的に広い範囲のタスク処理を暗黙のうちにユーザから期待される傾向がある．そのため，音声対話インタフェースは一般的にマルチタスクであることが多い．しかし，ユーザが期待するあらゆる対話タスクに対応するのは困難であり，この暗黙の期待と実際にできることのギャップが話しかけにくさの要因となっていると考えられる．本研究では，それを「マルチタスク性」とし，音声対話インタフェースとして「マルチタスク性」の適切な表出方法，様式について提案・検討する．

2 音声対話インタフェースにおける マルチタスク性

現在，自然な対話ができる音声対話インタフェースとして，Appleの「Siri」[1]やNTTdocomoの「しゃべってコンシェル」[2]などのスマートフォンに搭載されている音声アシスタントが挙げられる．音声アシスタントとは，デバイスの中にまるで人がいて対話ができるようなソフトウェアである．奈良先端科学技術大学院大学の「たけまるくん」[3]や本学開発のMMDAgent[4]もこれにあたる．音声アシスタントは，話しかけることにより今週の天気や次ぎに来る電車を調べたり，登録されている番号に電話をかけたることができる．

音声アシスタントを始めとした音声対話インタフェースでは，このようにマルチタスクであることが多い．それは，ユーザが音声対話インタフェースにマルチタスクであることを期待するためであり，本研究では，これを「マルチタスク性」と定義する．ユーザは音声対話インタフェースを目の前にしたとき，特にタスク内容が明示されていない場合，ユーザはシステムが何でもできると感じてしまう傾向がある．そのため，システムを開発する際には，ありとあらゆるタスクが実行できるよう実装しなければならない．その一例として，AppleのSiriが挙げられる．ユーザは自由に発話を行うことができるが，際限がなく，システムは多くの応答パターンを持つような無限のマルチタスク性を保障しなければならない．また，このようにユーザの期待と実際の能力にギャップがあるとスムーズなインタラクションは成立しない．

このように，暗黙の期待に合わせて実行可能なタス

*連絡先：名古屋工業大学
愛知県名古屋市昭和区御器所町
E-mail: s-konaka@slp.nitech.ac.jp

クを増やすことには限界があり、適切なタスク範囲をシステムからユーザへ表出することが必要である。本研究では、このマルチタスク性を適切に表示することによって、ユーザが感じる使いにくさの改善を図る。

3 マルチタスク性の表出方法

3.1 従来のタスク表出方法

マルチタスク性をもつ音声対話インタフェースの従来の表出方法として、あらかじめ受理可能な発話内容を明示する方法がある。例えば、受理可能な発話内容を明示する方法の一つとして、立看板やポスターなどにより「”こんにちは”と話してみてください」などの文字情報を示す方法がある。これは公共施設やイベントで利用される設置型の音声対話インタフェースに多く、具体的な発話案の提示はユーザにはっきりと伝えやすく、またシステムや対話内容に変更があった場合にも設定しやすい。

しかし、ユーザにとっては与えられた言葉リストから選択して発生している状態となる。このような語彙レベルの明示はタスクの了解性と引き換えに音声対話インタフェース本来の自発的・対話的な発話行為を阻害してしまう。また、自然言語を扱う音声対話インタフェースは発話可能な語彙が数百語以上であることが多く、それら全てを表示することは難しい。

文字だけではなく、アイコンを表示することでユーザにタスク内容を提示する方法もある。これは GUI で一般的な表出方法であり、音声対話インタフェースにおいても適切である可能性が高いが、このようなイメージによるタスク範囲の表出が音声対話インタフェースにおいて十分であるかは検証する必要がある。

以上のようなタスク範囲を直接システム側から表出する方法の他に、ユーザの発話履歴を表示することで、他ユーザの利用状況を見せる表出方法がある。この表出方法は一定の評価を得ている [5]。他ユーザの発話内容を知ることにより、システムが受理可能なタスクが理解でき、システムへの抵抗感が減少できるとされている。

3.2 提案法

今回、音声対話インタフェースにおいて音声対話のメリットであるユーザの発話行動の自発性とインタフェースとしてのわかりやすさであるタスク表出の明示度を両立する方法として、マルチエージェントによるタスク表示を提案する。

これは、対話可能性を表出するソフトウェアエージェントを複数配置し、エージェントごとに少量のタスク

を分配することによって、マルチタスク性を示し、アイコンや文字情報表示のように明示的にしすぎない表出である。ユーザから見ると、ユーザの対話に応じているのはコンピュータではなく、表示されているエージェントであるといえる。つまり、システムが所有する機能が増えると必然的にエージェントが対応する機能が増えるかのように表出される。全てのユーザが実行可能なタスクを把握しているとは限らないため、タスク量を一定にしたエージェントを複数配置することによって、ユーザがタスク範囲を自然に理解でき話しかけやすく、かつ発話の自由度・自発度をも保つインタフェースとなると考えられる。

4 評価実験


4.1 実験条件

本研究では、提案法に加え、従来の表出方法のうち、アイコンによるタスク表出について評価を行うため、3種類のシステムを作成した。被験者 18 名がシステムを利用し、与えられた課題を音声対話インタフェースをもつタブレット端末を想定したシステムを使って解決する主観評価実験を行った。

4.1.1 実験システム

今回、実験用に 3 種類のシステムを用意し評価実験を行った。本研究室が取り組んでいる音声言語処理を利用し、オープンソースとして公開している MMDAgent と呼ばれるソフトウェアを用いる。モニタとして画面比率 16:9 の 32V の液晶ディスプレイを設置し、入力デバイスはスマートフォンやタブレットなどの端末を想定しているため、マイクなどの設置は行わず、Kinect for Windows センサー [6] を採用した。MMDAgent をベースに双方向音声案内デジタルサイネージとして開発されたキャラクターである SD 版メイちゃん¹を利用した。

各システムの詳細を述べる。まずは、従来の音声対話インタフェースを表す、アイコン表示がなく、一人のエージェントが全てのタスクを担当するシステム (以下、Alone システム) である (図 1)。2 つ目のシステムとして、図 2 のようなエージェントは一人でないながら、アイコン表示を行うシステム (以下、Icon システム) である。従来の音声対話インタフェースを表した Alone システムと比較し、マルチタスク性を適切に表示され、さらに話しかけやすさの改善に繋がっているか調査する。また、図 3 のようなアイコンとマルチエージェン

 Copyright 2009-2015 Nagoya Institute of Technology (MMDAgent Model "SD Mei")

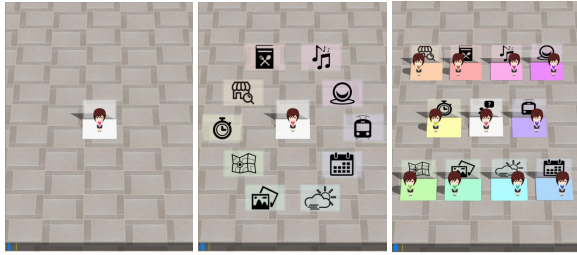


図 1: Alone システム 図 2: Icon システム 図 3: Multi システム (提案法)

トを表示しているシステム (以下, Multi システム) と比較することでエージェントのタスク分量を分けることによってマルチタスク性の表出によりユーザのタスク理解までの負担が軽減できるか調べる。

4.1.2 実験環境

実験環境はプライベートな室内空間を想定する。そのため、被験者のみの静かな部屋で実験を行った。ビデオカメラを設置し、実験の様子を後で把握できるようにした。

4.1.3 事前説明

被験者への事前説明として、表 1 の内容を紙媒体で提示した。また、今回の評価実験では、実験の様子を映像として記録するため、被験者全員に承諾書を読んでもらい署名をもらった。

最初に説明用紙の内容を見せ、表 2 のような実験想定が書かれた紙を実験中被験者の側に置き、見てもらいながら実験を行った。システムへの慣れを正準化するため、被験者によって実験システムの順番を入れ替えて行った。

実験を始める前に、システム紹介として実験を行うシステムを見せ、およそ 1 分間眺めたり、自由に話しかけたりする時間を設けた。音声対話システム自体にあまり触れたことのない被験者には、全く発話しないことがないように、あらかじめ使い方を説明し、実際の実験想定とは関係ないあいさつや占いといった項目を被験者に伝えた。各システム実験後、それぞれ次節の 5 段階評価のアンケートを行った。

4.1.4 評価方法

一つのシステムの実験が終わるごとに 5 段階のアンケート調査を行った。システムに関するアンケート項

表 1: 説明用紙の内容

- ・これから 3 種類の音声対話システムを利用させていただきます。
- ・想定される実験時間は 20 ~ 30 分です。システム紹介、実験、アンケート記入を 1 セットとし、3 セット行います。
- ・実験の様子をビデオカメラを使って収録するため、承諾書を了解した上でサインをお願いします。
- ・対話に躓いたとき、進まなくなったとき、また全く別の話題について話したいときは「ホーム」または「戻る」「やめる」というと次ページのような画面に戻ります。
- ・今、自分が実験想定のような状況にあることを想定し、タブレット端末だと思って話しかけて調べてみてください。
- ・1 システムの実験にかかる時間は 3 分です。3 分間常に話しかけようとしなくても構いません。
- ・話しかけてシステムが応答しなくても気にしないでください。実験想定の内容について全て聞く必要はありません。
- ・実験中、観察者に話しかけたとしても質問などに応えることができません。

目を表 3 に示す。全ての項目について、1:決してそう思わない、2:あまりそう思わない、3:普通、4:ややそう思う、5:とてもそう思うの 5 段階評価を行い、総合評価に関しては 1 が最も低く、5 が最も高いという評価基準である。

また、日常的に音声対話システムを使用している人はコンピュータとの対話に慣れていることを加味しなければいけない。また、日常的にコンピュータで知りたいことを調べているのかによって発話内容を考える時間や苦勞が異なると考え、表 4 にある 2 項目についても実験の最後に回答を得た。

表 2: 実験想定の内容

- ・今日は 12 月 10 日です。
- ・もうすぐ友人の誕生日です。
- ・今週末、別の友人とロフト名古屋にプレゼントを探しに行くことになりました。
- ・現在、名古屋工業大学にいます。おそらく当日の朝も泊まり込みで名古屋工業大学にいます。
- ・名古屋工業大学の最寄り駅は鶴舞駅です。
- ・あなたは毎週土日どちらかでアルバイトをしています。
- ・友人とはロフト最寄りの駅に 11 時に集合します。
- ・今週は週末にかけて天候が悪くなる予想ですので電車で移動しましょう。
- ・あなたはよく寝坊します。遅刻しないようにしましょう。

表 3: アンケート項目 (システム関連)

Q.1	システムを前にして使いたいと思ったか
Q.2	話しかけたいと思ったか
Q.3	使いやすかったか
Q.4	対話に引き込まれたか
Q.5	不快に感じたか
Q.6	話しやすかったか
Q.7	使い方が直感的にわかったか
Q.8	操作に戸惑ったか
Q.9	今後も使いたいと思ったか
Q.10	総合評価

表 4: アンケート項目 (ユーザ関連)

Q.11	音声対話システムを使う方だ
Q.12	知りたいことはスマートフォンで調べる方だ

表 6: システム毎のスコア平均

項目	Alone	Icon	Multi
Q.1	3.00	3.94	3.94
Q.2	2.94	3.67	4.06
Q.3	2.22	3.56	3.56
Q.4	2.78	3.11	3.28
Q.5	2.39	2.00	2.00
Q.6	2.67	3.50	3.50
Q.7	2.44	4.17	4.28
Q.8	3.17	1.94	2.11
Q.9	2.50	3.50	3.72
Q.10	2.50	3.61	3.72

4.2 実験結果

4.2.1 実験用 3 システム間による比較

被験者 18 名による評価実験の結果を示す。今回の「総合評価」を除くシステムに関するアンケート項目は図 5 のように使いやすさに関する項目、話しかけやすさに関する項目に分類できる。

表 6 は全てのアンケートについて各項目の平均を集計したものである。値は有効数字 3 桁とする。表 6 より、Alone システムと Icon システムの比較においてどの項目についても差があることからアイコン表示の有無によって、使いやすさ、話しやすさが改善されていると言える。特に、使いやすさに関する項目は全て 0.5 以上の差が見られ、GUI のみならず音声対話インタフェースにおいてもアイコン表示はマルチタスク性を表出に有効であった。総合評価においても、1.11 の差が見られ、Icon システムの方が好印象であった。

一方で、Icon システムと Multi システムの比較からエージェント数の変化による差はあまり見られなかった。わずかに Q.2「話しかけたいと思ったか」で 0.39、Q.4「対話に引き込まれたか」で 0.17 上昇しており、話

表 5: 質問の分類分け

・使いやすさに関する項目	Q.1 システムを前にして使いたいと思ったか
	Q.3 使いやすかったか
	Q.7 使い方が直感的にわかったか
	Q.8 操作に戸惑ったか
	Q.9 今後も使いたいと思ったか
・話しやすさに関する項目	Q.2 話しかけたいと思ったか
	Q.4 対話に引き込まれたか
	Q.5 不快に感じたか
	Q.6 話しやすかったか

表 7: 被験者自身に関するアンケート 2 項目の人数分布

項目	5 段階評価				
	1	2	3	4	5
Q.11	2	8	2	1	5
Q.12	0	1	3	6	8

しやすさに関する項目においてマルチエージェントによる表示の方が優れていた。Q.8「操作に戸惑ったか」という項目ではマルチエージェント表示よりも従来の単一エージェント表示の方がよいという結果になっている。これは、多くのエージェントが並ぶことによって、逆に誰に話しかけるのか迷いが生じたことが考えられる。マルチエージェントによるマルチタスク性の表出は、エージェント数が多いと混乱を招く可能性もあるが、エージェントの数や配置、大きさなどによってさらに改善する可能性が示唆された。

4.2.2 音声対話システムへの慣れによる差

3 システムの実験を行った後に被験者に答えてもらった 2 つの項目について集計を行った。表 7 のようになり、Q.11「音声対話システムを使う方だ」という項目に関して、1 または 2 を選択した人は 10 人、4 または 5 を選択した人は 6 人いた。そこで、1 または 2 を選んだ人を音声対話システムをあまり使わないグループ A、4 または 5 を選んだ人を音声対話システムをよく使うグループ B に分け、集計を行った。結果を表 8 に示す。ここでは、各グループ毎の 3 システムのスコア平均を示している。

表 8 より、いずれのグループにおいても Alone システムより Icon システムの方が高く、アイコンによるタスク表出は音声対話システムへの慣れによらず効果的であることがわかる。より詳細に比較すると、あまり

表 8: システム毎のスコア平均 (音声対話システムへの慣れによるグループ分け)

項目	Alone		Icon		Multi	
	A	B	A	B	A	B
Q.1	2.90	3.00	4.00	3.67	3.90	3.83
Q.2	3.10	2.83	4.00	3.33	4.30	3.67
Q.3	2.10	2.33	3.70	3.17	3.70	3.33
Q.4	2.90	2.83	3.60	3.33	3.70	3.17
Q.5	2.30	2.67	2.00	2.17	1.70	2.67
Q.6	2.50	2.83	3.60	3.33	3.70	3.17
Q.7	2.00	3.33	4.30	4.17	4.30	4.33
Q.8	3.50	2.67	2.10	1.67	2.10	1.83
Q.9	2.50	2.50	3.70	3.17	3.90	3.17
Q.10	2.40	2.50	3.60	3.50	3.90	3.33

グループ A:Q.11 について 1 または 2 を選択した人
 グループ B:Q.11 について 4 または 5 を選択した人

慣れていないグループ A の方がグループ B より Alone システムでは低く、Icon システムでは高く評価されている。このことから、慣れていない人にこそ適切なタスク表出の必要性が高いことがわかる。

Icon および提案法である Multi システムを比較すると、スコア平均の差はあまり見られないが、アイコンによる表出による効果を阻害せず、さらに、Q.2「話しかけたいと思ったか」において改善が見られた。

さらにグループ A の Multi システム、グループ B の Alone システムを比較するとグループ A の評価が高い。これは、グループ A の Multi システムはインタフェースの改善による効果、グループ B の Alone システムは音声対話システムへの慣れによる効果を指しているといえる。つまり、このことから音声対話インタフェースにおいてはユーザに慣れを促すことよりもインタフェースの改善の方がよりよいことがわかる。ただし、慣れの度合いと適切なインタフェースの関係についてはさらなる実験が必要である。

今回の実験は、エージェント数が少ないほうがすっきりして見えるなどの感受性による個人差もあり、これらの問題を改善することにより、さらにエージェント数による話しかけやすさの検証ができると考えられる。

5 むすび

本研究では、音声コマンドを用いない自然な人間らしい発話でタスク達成ができる音声対話インタフェースに対する話しかけやすさの改善を目標に、多くの音声対話インタフェースがマルチタスクであることに着目した。これは、ユーザが音声対話インタフェースに

ありとあらゆるタスクができると感じてしまい、暗黙のうちに期待してしまうためである。ユーザのシステムへの期待と実際に対話した場合のギャップによって話しかけにくいと考えられ、本研究ではこれを「マルチタスク性」と定義した。音声対話インタフェースにおけるマルチタスク性の適切な表出方法について検討するため、従来の表出方法であるアイコン表示と、今回提案したマルチエージェント表示について評価実験を行った。

アイコン表示においては、ユーザの音声対話システムに対する慣れに関わらず、総合的に高い評価を得られた。GUI で広く使われているアイコン表示はマルチタスク性の表出に有効であった。また、提案したマルチエージェント表示は、大きな差は見られなかったが話しかけやすさに関する項目で良い評価を得られており、個人差があるものの音声対話インタフェースへの抵抗感を減らす一つの方法として有効であると考えられる。マルチタスク性の適切な表出方法であるかという点については明確には判断できず、今後さらに実験を重ねていく必要があるといえる。

参考文献

- [1] Apple - iOS 8 - Siri
<http://www.apple.com/ios/siri>
- [2] 辻野孝輔, 栄藤稔, 磯田佳徳, 飯塚真也, “実サービスにおける音声認識と自然言語インタフェース技術”, 人工知能学会誌, Vol.28, No.1, pp.75-81, 2013.
- [3] 西村竜一, 西原洋平, 鶴見玲典, 李晃伸, 猿渡洋, 鹿野清宏, “実環境研究プラットホームとしての音声情報案内システムの運用”, 電子情報通信学会論文誌, Vol.J87-D2, No.3, pp.789-798, Mar.2004.
- [4] 大浦圭一郎, 山本大介, 内匠逸, 李晃伸, 徳田恵一, “キャンパスの公共空間におけるユーザ参加型双方向音声案内デジタルサイネージシステム”, 人工知能学会誌, Vol.28, No.1, pp.60-67, 2003.
- [5] 飯塚遼, 李晃伸, “ユーザ生成型音声対話システムにおけるクリエイターとユーザの相互刺激によるインセンティブ向上の検討”, 音声言語情報処理研究会 SIG-SLP 第 101 回研究会, (2014.5).
- [6] KINECT for Windows
<http://www.microsoft.com/en-us/kinectforwindows/>