

# 対話システムの確認応答がユーザに与える効果の分析

市野貴之 島津明

北陸先端科学技術大学院大学 情報科学研究科

## 1 はじめに

人間同士の対話において、人間は相槌や様々な確認応答を随所に行っていることが観察され、適切な位置に相槌や確認応答を行っていくことで対話は円滑に行われていると思われる。このような確認応答戦略を対話システムに反映し、人間とシステムとの対話を円滑にすることにより、ユーザにとって使いやすいシステムが構築できると考えられる。詳細は割愛するが、人間同士の会議室予約の対話では、キーとなる項目(人数)について始めの段階で直接確認するとか、特定の項目(人数、曜日)にまとめ確認するとかの現象がみられる。本研究では、対話システムが適切に相槌、確認応答を入れてゆき、ユーザにとって使いやすい対話システムを作成することを目的とし、実験システムを作成し、確認応答戦略がユーザにどのような影響を与えるのか実験を行った。これについて以下に報告する。

## 2 関連研究

人間とシステムとの対話に関する確認応答の研究がある。[1]は、対話システムの応答の効果を調査するためにWOZ実験を行っている。電話ショッピングのタスクを用いて、もしユーザ発話にあらかじめ決めておいたキーワードが含まれていたら、そのユーザ発話の終わりに応答を挿入するものであった。この論文では、システムの応答を挿入するタイミングがユーザの満足度に関して重要であるという結果を得ているが、あまりに冗長な応答はユーザの満足度を減らす傾向があるので応答戦略を考える必要があるとしている。[2]は対話システムが打つ相槌がユーザにどのような影響を与えるのかを調べることに焦点をおいている。会議室予約をタスクとし、あらかじめ決めておいた重要語がユーザの発話に含まれていれば、その都度システム応

U: 第一会議室を予約したいのですが。  
S: 第一会議室ですね。[直接確認]  
S: はい。[相槌]  
S: 何月何日に第一会議室ですか。[間接確認]  
S: それでは予約の確認をします。  
5月1日、午前9時から、  
第3会議室でいいですか。[まとめ確認]

表 1: ユーザ発話に対する確認応答の例

答を入れるもの(Immediate条件)とユーザ発話が終わったところでシステム応答を入れる(Orderly条件)ものという2つの条件を設定して実験を行い、Orderly条件がユーザに対して好印象を得る結果をえているが、システム応答として「はい」以外の確認応答については考えられていない。これらの研究では円滑な対話を行う、使いやすいシステムの作成には不十分であると考えられる。[3]では、会議室予約を対象に3つの確認応答戦略による対話と対話時間、システム発話時間、対話ターン数との関係を調べているが、ユーザの評価はなされていない。

## 3 本研究で注目する確認応答

本研究では4種類の確認応答に着目した(図1)。直接確認とは、繰り返し応答を行い確認することである。相槌とは了解を意味する「はい」などの確認である。間接確認とは、次の発話に確認事項を入れて確認を行うものをいう[4]。まとめ確認は、対話の最後にすべての条件に対して確認を行うものを言う。

## 4 実験システム

対話システムの確認応答がユーザにどのような影響を及ぼすのかを調べるため、会議室予約を対象として、様々な確認応答戦略で対話を進めるシステム主導の実

験システムを作成した。システムは、音声認識モジュールと予約モジュールから構成される。音声認識モジュールは、IBMのVoiceclientを用いて、ユーザから受けた音声入力を文字列に変換し、その結果を予約モジュールに送る働きをしている。予約モジュールは、音声認識モジュールから送られた文字列とキーワードマッチングを行い、その結果に応じて、対話システムの応答文を組み立てて、Viavoiceの音声を出力する。

確認応答戦略のパターンとして、予約条件(人数,曜日,時間)のすべてに直接確認するもの,すべてに相槌を打つもの,間接確認するもの,すべてに無応答のもの,部分的に直接確認するもの(2種類)とそれぞれに対して,まとめ確認を行うものの合計12種類で,システム発話の表現が違うもの2種類を用意して,それぞれAとBとし,合計で応答戦略24パターン考えた。表現Aは,ユーザの発話に制約をかけるようにした「人数は何名ですか」,「曜日は何曜日ですか」,「時間は何時からにしますか(何時までにしますか)」で,表現Bは,何も制約をかけない「人数をおっしゃってください」,「曜日をおっしゃってください」,「時間をおっしゃってください(何時までにしますか)」である。

## 5 実験方法

実験はまず,被験者に実験の目的,方法が書かれた実験シートを配布し,内容を読んでもらい,実験目的や実験方法などを明確にしてもらう。この際,被験者に実験シート以外の情報を何も与えなかった。次に被験者に予約をしてもらう人数,曜日,時間などが文章で明記された予約シートを渡し,その予約シートに沿った条件で実験システムと対話して予約を行ってもらった。そして,対話が終了する毎にシステムのユーザビリティを調べる5段階評価のアンケートに答えてもらった。アンケートの項目としては,(1)システムとの対話は円滑に進められたか,(2)システム表現は応答しやすかったか,(3)システムは使いやすかったか,(4)まとめ確認は必要であったか。(1)は各対話戦略が対話の円滑性に与える影響を見るために,(2)はシステム表現がシステム全体のユーザビリティに与える影響を見るために,(3)は被験者が最も使いやすかったものを見るために,(4)は各対話戦略とまとめ確認の相関からその必要性を調べるためにそれぞれ設けられた。

S: 会議室予約システムです。人数は何名ですか。  
 U: 15名です。  
 S: 15名でよろしいですか。  
 U: はい。  
 S: 曜日は何曜日ですか。  
 U: 金曜日で。  
 S: はい。  
 .....  
 S: それでは,予約の確認をします。  
 人数は15名で,曜日は金曜日。  
 時間は5時から6時までで,  
 第3会議室になります。  
 以上でよろしいでしょうか。  
 U: はい。  
 S: 予約は完了しました。

表 3: 収録した対話の例

## 6 実験結果

被験者は,20歳前後の非情報系大学生12名(男性4名,女性8名)で,それぞれ24の応答戦略毎にシステムと対話をして合計288対話を収録した。収録した対話例の一部を図2に示す。

ここで対話戦略の説明をする。対話戦略1は,全予約条件に直接確認するもので,対話戦略2は対話戦略1とまとめ確認の組み合わせたもの。対話戦略3は,人数,曜日に直接確認し,時間に相槌を打つもので,対話戦略4は対話戦略3とまとめ確認を組み合わせたもの。対話戦略5は,人数に直接確認し,曜日,時間には相槌を打つもので,対話戦略6は対話戦略5にまとめ確認を組み合わせたもの。対話戦略7は,全予約条件に相槌を打つもので,対話戦略8は対話戦略7とまとめ確認を組み合わせたもの,対話戦略9は間接確認をするもので,対話戦略10は対話戦略9とまとめ確認を組み合わせたもの。対話戦略11は無応答で対話を進めるもので,対話戦略12は対話戦略11とまとめ確認を組み合わせたものである。

288対話中35対話はタスクを達成できなかった。今回は,タスクが達成できたものを対象に対話システムの確認応答がユーザに与える効果を調べた。実験結果に対する評価の項目として,アンケートの得点,対話ターン数,対話時間,繰り返し発話の回数を用いた。実験の結果をそれぞれ表3,4,5に示す。以下に主な分析結果を示す。

### (1) 対話の円滑性の分析

表 2: 対話の円滑性のデータ

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値	対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1A	3.9	1.2	5	1	4	5	1B	3.5	1.3	5	1	3	3
2A	3.3	1.2	5	2	3	2	2B	4.0	0.8	5	3	4	3
3A	3.2	1.4	5	1	3	5	3B	3.4	1.2	5	2	3	2
4A	4.4	0.8	5	3	5	5	4B	3.8	1.1	5	2	4	5
5A	3.3	1.2	5	1	4	4	5B	3.4	1.4	5	1	3	5
6A	3.8	1.0	5	2	4	3	6B	3.6	1.1	5	2	4	4
7A	3.7	1.1	5	1	4	4	7B	4.0	0.7	5	2	4	4
8A	4.0	0.8	5	3	4	4	8B	4.1	0.9	5	3	4.5	5
9A	3.6	0.9	5	2	4	4	9B	3.3	1.3	5	2	3	2
10A	3.2	1.1	5	2	3	3	10B	3.8	1.0	5	2	4	4
11A	3.8	1.0	5	2	4	4	11B	3.4	1.2	5	2	3	2
12A	3.8	1.4	5	1	4	5	12B	3.8	1.1	5	2	4	4

表 5: 使いやすさのデータ

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値	対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1A	4.3	0.6	5	3	4	4	1B	4.2	0.9	5	2	4	4
2A	3.7	0.9	5	2	4	4	2B	4.1	0.7	5	3	4	4
3A	3.8	1.1	5	2	4	4	3B	3.9	1.0	5	2	4	5
4A	4.4	0.6	5	3	5	5	4B	4.1	0.9	5	2	4	5
5A	4.0	0.7	5	3	4	4	5B	3.7	1.2	5	1	4	3
6A	3.9	1.1	5	2	4	5	6B	3.8	0.9	5	3	3.5	3
7A	3.7	1.1	5	2	4	5	7B	4.2	0.9	5	2	4.5	5
8A	4.0	0.6	5	3	4	4	8B	4.1	0.7	5	3	4	4
9A	4.0	0.8	5	3	4	3	9B	3.4	1.0	5	2	3	3
10A	3.7	1.1	5	2	4	4	10B	4.2	0.7	5	3	4	4
11A	4.1	0.9	5	3	4	5	11B	3.9	1.1	5	2	4	5
12A	4.0	1.1	5	2	4	5	12B	3.9	0.9	5	2	4	4

表 4: まとめ確認に関するデータ

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2A	4.2	1.3	5	1	5	5
4A	4.3	1.2	5	1	5	5
6A	4.3	0.6	5	3	4	4
8A	4.4	0.7	5	3	5	5
10A	4.4	0.6	5	3	5	5
12A	3.9	1.3	5	1	4.5	5
2B	4.0	1.2	5	1	4.5	5
4B	4.6	0.5	5	4	5	5
6B	4.2	0.7	5	3	4	4
8B	4.3	0.8	5	3	4.5	5
10B	4.0	1.4	5	1	4	4
12B	4.5	0.6	5	3	5	5

・対話戦略を区別せずにまとめて見ると対話の円滑性と対話ターン数の関係は、相関係数-0.125であり、強い相関関係がなかった。対話の円滑性と対話時間の関係も相関係数-0.046から相関関係があるとは言えない。

・まとめ確認の有無では、対話戦略 1B と 2B, 3A(B) と 4A(B), 5A(B) と 6A(B), 7A(B) と 8A(B), 9B と 10B, 11A(B) と 12A(B) のほとんどの組み合わせにおいてまとめ確認を行うものの平均点が高かった。対話戦略 1A と 2A, 9A と 10A に関しては、音声認識の問題によ

り繰り返し発話が多く、対話時間が長くかかったために平均点に影響が出たと考えられる。統計的には、3A と 4A の間では有為な差 ( $t = |-2.25| > t(0.05, 17) = 2.1$ ) があったが、他は有為な差はなかった。

・システムの確認応答の違いでは、対話ターンを減らす効果がある間接確認は、どの組み合わせにおいても直接確認より平均点は低い。

・無応答のものは、情報をシステムが確実に処理したか確信できないためか、直接確認を全予約条件に対して行っている方がユーザーに評価されている。

・まとめ確認を行うのであれば、無応答であっても平均点が高い事が対話戦略 2A と 12A の比較からもわかる。

・全予約条件に対して直接確認を行うものが、予約条件に部分的に直接確認を行うもの比べて平均点が高い。これはユーザーが予約したい情報が確実にシステムに伝わっているのか確認したいためであると思われる。

・全予約条件に直接確認、まとめ確認も行う組み合わせが部分的に直接確認し、まとめ確認も行う組み合わせに比べて平均値は低い。

・同じ対話戦略でも、システム発話の表現の違いに

よって評価値が違い、表現が対話の円滑性に影響を与えていると考えられるが、A、B どちらの表現が良いとは一概に言えない。

### (2) まとめ確認の必要性の分析

・同じ対話戦略でも、システム発話の表現の違いによって評価値が違い、表現がまとめ確認の必要性に影響を与えていると考えられるが、A、B どちらの表現が良いとは一概に言えない。

・全予約条件に対して相槌を打つものは、まとめ確認も行う方がユーザの得点が高く、逆に直接確認を行うものは、まとめ確認を行わない方が平均点が高い。

・間接確認に関しても相槌の場合と同様のことが言える。

・全予約条件に対して直接確認を行い、最後にまとめ確認を行う戦略は、他の戦略よりユーザの得点が低い。

・以上については、有為水準 1% で平均点に有為な差は見られなかった。

### (3) 使いやすさの分析

・対話戦略を区別せずにまとめて見ると、使いやすさと対話ターン数の相関は、0.035 とほとんど相関がないと言える。また、使いやすさと対話時間の関係もあまりお互いに影響している関係を持っているとは言えないことが相関係数 (-0.12) からわかる。

・使いやすさとその他の要因の相関を調べるために重回帰分析を行った。目的変数  $Y$  は、使いやすさとし、説明変数  $X$  は、対話の円滑性 ( $x_1$ )、システム表現 ( $x_2$ )、対話ターン数 ( $x_3$ )、対話時間 ( $x_4$ ) とした。標準偏回帰係数により単位の影響を除き、各要因との比較を行った。

$$Y^* = 0.519x_1^* + 0.41x_2^* + 0.252x_3^* - 0.299x_4^* \quad (1)$$

重回帰式の標準偏回帰係数の比較より、対話の円滑性、システムの表現が使いやすさに強い影響を与えている。また、対話時間、対話ターン数の影響はそれらに比べて小さい。

・システム発話の表現の違いで対話戦略毎に平均点が異なり、システム発話の表現が使いやすさに影響があると考えられる。統計的には 2A と 2B の組み合わせのみ ( $t = |-2.27| > t(0.05, 14) = 2.14$ )、システム表現に有為な差が見られた。

・全予約条件に直接確認を行う方がまとめ確認を加えた場合よりも得点が高い、相槌、無応答の場合は、まとめ確認を行なう方が得点が高い。まとめ確認の有

無では、対話戦略 1A と 2A の組み合わせでのみ有為な差が見られた ( $t = 2.58 > t(0.05, 14) = 2.14$ )。

・間接確認を行うものは、平均点が低かった。これは日常的に用いる機会が少ないため、被験者が違和感を感じたためではないかと考えられる。

## 7 おわりに

システムの確認応答がユーザに与える効果を調べるために、会議室予約を対象に対話実験を行い、使いやすさに関して、システムの表現の関わりが大きいこと、対話時間、対話ターン数は対話の円滑性、使いやすさと相関が小さいこと、適度な確認が使いやすさ、対話の円滑性に必要であるという結果を得た。

今後は、例えば、システム発話の表現の種類を増やし、どのような表現がユーザにとって使いやすいものであるかなどの検討を行うことが必要である。

## 参考文献

- [1] Y.Okato, K.Kato, M.Yamamoto, S.Itabashi. "System-User Interaction and Response Strategy in Spoken Dialogue System", ICSLP'98, vol2, pp.495-498, 1998.
- [2] J.Hirasawa, M.Nakano, T.Kawabata, K.Aikawa. "Effects of System Barge-in Responses on User Impressions", Eurospeech'99, pp.1391-1394, 1999.
- [3] A.Shimazu, H.Taguchi, M.Kawamori. "Characteristics of Acceptance Utterances and Their Relations to Dialog Strategies", Proc. of The Third International Workshop on Human-Computer Conversation, pp152-158, 2000.
- [4] 小林, 中島, 新美. 「音声模擬対話における対話制御と快適性について」, 音声言語情報処理 6-2, pp.9-14, 1995.