JAIST Repository
https://dspace.jaist.ac.jp/

Title	対話システムの確認応答がユーザに与える効果の分析
Author(s)	市野，貴之
Citation	
Issue Date	2002-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1567
Rights	
Description	Supervisor:島津 明，情報科学研究科，修士

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

修 士 論 文

対話システムの確認応答がユーザに与える効果の分析

指導教官　島津 明 教授

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

市野貴之

2002 年 2 月 15 日

Copyright © 2002 by Takayuki ICHINO

要 旨

人間同士の対話では，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．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く，人間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，先ず，人間同士の対話を対象に相槌，確認応答などがどのように出現しているか分析を行った．次に，確認応答戦略が受け手にどのような影響を与えるか，会議室予約を対象に対話システムによる実験を行い，確認応答戦略の種類による違いを分析した．実験結果の分析より，対話の円滑性，使いやすさを高めるために，システムの表現，システムの確認応答とまとめ確認の組み合わせが重要な要因であること，また，対話時間，対話ターン数はあまり対話の円滑性，使いやすさにあまり影響しないという結果を得た．

目 次

1 はじめに **1**
1.1 研究の背景 1
1.2 関連研究 2
1.3 研究の目的 3
1.4 本論文の構成 3

2 本研究で扱うシステムの確認応答 **4**
2.1 直接確認 4
2.2 間接確認 5
2.3 相槌 5
2.4 まとめ確認 6

3 対話コーパス分析 **7**
3.1 分析の目的 7
3.2 分析に用いた対話データ 7
3.3 分析方法 7
3.4 対話データの内訳 8
3.5 考察 8
3.5.1 直接確認・相槌の考察 8
3.5.2 まとめ確認の考察 9
3.5.3 その他の考察 9

4 実験システム **10**
4.1 本研究の対話タスク 10
4.2 実験システムの構成 11
4.3 音声認識モジュール 12

4.4 予約モジュール 12
4.5 実験システムの (確認) 応答について 13
4.6 実験システムの対話戦略について 13

5 実験 **15**
5.1 予備実験 15
5.1.1 予備実験の目的 16
5.1.2 実験の方法 16
5.1.3 予備実験の結果 17
5.1.4 予備実験の考察 17
5.2 実験 17

6 実験結果 **19**
6.1 分析の評価項目 19
6.1.1 分析の評価項目の説明 20
6.2 実験で得られたデータ 20
6.3 対話の円滑性の分析 21
6.3.1 システム発話の表現の違いによる比較 21
6.3.2 まとめ確認の有無による比較 25
6.3.3 システム応答 (直接，間接確認，相槌) の違いによる比較 27
6.3.4 対話の円滑性と対話ターン数の相関の分析 (対話戦略毎) 31
6.3.5 対話の円滑性と対話ターン数の相関の分析 32
6.3.6 対話の円滑性と対話時間の相関の分析 (対話戦略毎) 32
6.3.7 対話の円滑性と対話時間の相関の分析 33
6.4 まとめ確認の必要性の分析 34
6.4.1 システム発話の表現の違いによる比較 34
6.4.2 システム応答 (直接，間接確認，相槌) の違いによる比較 36
6.5 使いやすさの分析 39
6.5.1 システム発話の表現の違いによる比較 40
6.5.2 まとめ確認の有無による比較 42
6.5.3 システム応答 (直接，間接確認，相槌) の違いによる比較 44
6.5.4 使いやすさと対話ターン数の相関の分析 (対話戦略毎) 49
6.5.5 使いやすさと対話ターン数の相関の分析 49
6.5.6 使いやすさと対話時間の相関の分析 (対話戦略毎) 49

6.5.7 使いやすさと対話時間の相関の分析 50
6.5.8 使いやすさと対話を特徴づける他の要因との相関関係の分析 (対話戦略毎) 50
6.5.9 使いやすさと対話を特徴づける他の要因との相関関係の分析 51
6.6 繰り返し発話と対話時間の相関の分析 (対話戦略毎) 52
6.7 繰り返し発話と対話時間の相関の分析 52

7 おわりに **53**
7.1 まとめ 53
7.2 今後の課題 54

図 目 次

2.1 会議室予約タスクでのシステムの直接確認の例 4
2.2 会議室予約タスクにおける間接確認の例 5
2.3 会議室予約タスクにおける相槌の例 5
2.4 会議室予約タスクにおけるまとめ確認の例 6

4.1 実験システム概略図 11

5.1 配布した予約シートの例 16

6.1 対話の円滑性と対話ターン数の散布図 32
6.2 対話の円滑性と対話時間の散布図 33
6.3 使いやすさと対話ターン数の散布図 49
6.4 使いやすさと対話時間の散布図 50
6.5 対話時間と繰り返し発話の散布図 52

.1 対話の円滑性と対話ターンの散布図 (1A〜12A) 63
.2 対話の円滑性と対話ターンの散布図 (1B〜12B) 66
.3 対話の円滑性と対話ターンの散布図 (1A〜12A) 70
.4 対話の円滑性と対話ターンの散布図 (1B〜12B) 73
.5 使いやすさと対話ターンの散布図 (1A〜12A) 77
.6 使いやすさと対話ターンの散布図 (1B〜12B) 80
.7 使いやすさと対話時間の散布図 (1A〜12A) 84
.8 使いやすさと対話時間の散布図 (1B〜12B) 87
.9 繰り返し発話と対話時間の散布図 (1A〜12A) 91
.10 繰り返し発話と対話時間の散布図 (1B〜12B) 94

表 目 次

3.1 直接確認・相槌の内訳 8
3.2 まとめ確認の種類と内訳 8

6.1 対話の円滑性に関する実験データ 21
6.2 システム発話の表現の違いによる比較 22
6.3 まとめ確認の有無による比較 25
6.4 まとめ確認に関するデータ 34
6.5 システム発話の表現の違いによる比較 34
6.6 使いやすさに関するのデータ 39
6.7 システム発話の表現の違いによる比較 40
6.8 まとめ確認の有無の比較 42

.1 対話の円滑性と対話ターン数の相関係数 67
.2 対話の円滑性と対話時間の相関係数 74
.3 使いやすさと対話ターンの相関係数 81
.4 使いやすさと対話時間の相関係数 88
.5 繰り返し発話と対話時間の相関係数 95

第 1 章

はじめに

1.1 研究の背景

近年の情報化社会の発展にともない，我々の日常生活の中で情報機器は欠かすことのできないものとなってきている．そのような社会の中でユーザが音声を用いて情報をやり取りし，観光案内を行ったり，予約を行ったりする事を目的とした音声対話システムが開発され，用いられるようになってきた．

人間がシステムと音声を用いて対話を行うことが，あたかも人間同士で行う対話のように自然に進めることができれば，音声対話システムは，人間にとってより使いやすいものになると考えられる．

そのため，音声対話システムには，より高精度の音声認識器や対話の主導権の制御方法，状況に応じたシステムの応答の生成が欠かすことのできない要因であると考えられる．

しかし，現状の技術では，言い直し，言い間違いなどの話し言葉特有の入力のために情報機器の音声認識や音声理解の誤りを避けられず，タスクが遂行されないといった問題があり，人間にとって音声対話システムが使いやすいものにはなっていない．

こうした問題を解決するために人間とシステムの対話の分析を行い，対話の特徴を解明する研究 [1,2,3] や，人間と情報機器との対話のやりとりの中から情報機器の誤解を検出，復旧しようとする研究 [4] が行われている．また，効率的に対話を進めるためのシステムの対話戦略制御を考える研究 [5,6,7,8,9]，異なる対話戦略の評価法の研究 [10,11,12,13,14,15,16] もなされてきている．

このように使いやすい音声対話システムを構築するためには，未だに多くの課題が残されている．これらの課題を一つ一つ解決し，または明らかにしていくことが人間にとってより使いやすい音声対話システムにつながると考えられる．

1.2 関連研究

[1] は，音声対話システムを構築する際に考慮することが重要であろう対話システムが打つ相槌のタイミングに注目して，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た対話実験が，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答を入れるもの (Immediate 条件-ユーザの発話と重なる) 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-ユーザの発話と決して重ならない) という 2 つの条件を設定して行われた．

実験の結果，Immediate 条件 (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答を入れるもの) における相槌では，何人かの被験者がシステムに対してあまり好印象を持たないというものとなった．

このことは，Orderly 条件 (ユーザ発話が終わったところでシステム応答を入れる) によるシステムの応答戦略がユーザに対して好印象を得る結果を得ているが，システム応答は「はい」だけであり「はい」以外の確認応答については考えられていない．

[2] では，音声認識エラーや不規則な発話により起こる誤解を防ぐために確認応答戦略の必要性を主張している．そして，L-type(システムがユーザの発話に「はい」と確認して行き，最後にまとめて確認を行う)，E-type(すべてのユーザの発話に明確に確認を行う)，I-type(他の発話によって間接的に確認を行う) といったシステムの確認応答戦略を考え，会議室予約をタスクとする音声対話システムを用いて対話実験を行い，それぞれの確認応答戦略の対話時間，対話ターン数，システムの平均発話長といった要因で分析を行っている．実験の結果，E-type が対話時間は長く，対話ターン数も多いが，システムの平均発話長が他の type と比較して短く，E-type が比較的効果的であると結論づけている．しかし，この研究では，ユーザの評価に関する調査が行われていない．

[3] では，情報機器と音声対話における情報伝達が能率良く，快適であるためには，どのような対話制御戦略を採用するのが適切であるかを検討する目的で情報機器と人間との音声模擬対話収集支援システムを構築し,「観光案内」を対話タスクとして，音声認識性能や対話制御戦略を変化させながら音声対話を収録し，それらを用いて対話の快適性や情報伝達の能率を分析，検討をおこなっている．

しかし音声認識性能や対話制御戦略をどのように変化させたのかは述べられていない．またどの程度タスクが達成されているのかも不明である．

[5] 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．

1.3 研究の目的

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することで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の構築できると考えられる．そこで本研究では，対話システムの確認応答に注目し，対話システムが適切に相槌，確認応答を入れてゆき，対話システムの最良の振る舞いを行えるようにするモデルを明かにすることを目的に，先ず，人間同士の対話を対象に相槌，確認応答などがどのように出現しているか分析を行った．次に，確認応答戦略が受け手 (対話の円滑性，システムの使いやすさ等) にどのような影響を与えるか，会議室予約を対象に対話システムによる実験を行い，確認応答戦略の種類による違いを調べ，それぞれの確認応答戦略の有効性の検証を行う．

1.4 本論文の構成

本論文は，2 章で本研究で着目するシステムの確認応答 (直接確認，間接確認，相槌，まとめ確認) とはどのようなものであるのかを述べる．3 章では，2 章で述べた相槌，確認応答が人間同士の対話では，どのような場合に出現しているのかの分析について述べる．4 章では，対話システムの確認応答がユーザにどのような効果を与えているのかを調べるために作成した実験システムについて述べる．5 章では，対話システムの確認応答がユーザにどのような効果を与えているのかを分析するために行う実験方法について述べる．6 章では，得られた実験結果をどのように分析を行ったのか，またその分析結果はどのようなものであったのかについて述べる．最後に 7 章でまとめと結論，そして今後の課題について述べる．

第 2 章

本研究で扱うシステムの確認応答

本研究では，明らかにされていないと考えられる確認応答の組合わせがユーザに与える効果を調べるため，対話システムの確認応答として主に直接確認，間接確認，相槌，まとめ確認について着目する．

2.1 直接確認

直接確認とはユーザの応答に対して，繰り返し確認を行うものである．以下は会議室予約タスクにおける直接確認の例を示している．

S: 会議室の予約を行います．人数は何名ですか．
U: 13 名です．
S: 13 名ですね．[直接確認]
U: はい．
S: 曜日は何曜日にしますか．
U: 水曜日で
S: 水曜日でよろしいですね [直接確認]
U: .....

図 2.1: 会議室予約タスクでのシステムの直接確認の例

2.2 間接確認

間接確認とは，ユーザの応答に対して直接に確認を行うものではなく，他の発話に確認の項目を含ませ間接的に確認を行うものである．以下に間接確認を行う会議室予約をタスクとする対話例を示す.

S: 会議室の予約を行います，人数は何名ですか.
U: 13 名です.
S: 人数は 13 名で何曜日にしますか．[間接確認]
U: 火曜日で.
S: 火曜日の何時から予約しますか．[間接確認]
U: .....

図 2.2: 会議室予約タスクにおける間接確認の例

2.3 相槌

相槌とは，ユーザの応答に対して，了解の意を示す「はい」等で確認を行うものである．以下に会議室予約をタスクとする対話で相槌を行うものの例を示す.

S: 会議室の予約を行います．人数は何名ですか.
U: 13 名です.
S: はい，[相槌] 曜日は何曜日にしますか.
U: 水曜日で
S: はい，[相槌] 時間は何時からにしますか.
U: .....

図 2.3: 会議室予約タスクにおける相槌の例

2.4 まとめ確認

まとめ確認とは，対話の最後に今までユーザによって埋められた予約条件項目の全てをまとめて確認するものである．以下は会議室予約タスクの対話で各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例を示している．

S: 会議室の予約を行います．人数は何名ですか．
U: 13 名です．
S: 13 名ですね．[直接確認]
U: はい．
S: 曜日は何曜日にしますか．
U: 水曜日で
S: 水曜日でよろしいですね [直接確認]
: ……
S: それでは予約の確認をします．
: 人数が 13 名で，水曜日の 5 時から 8 時までで,
: 第 3 会議室でよろしいですか．[まとめ確認]
U: はい．

図 2.4: 会議室予約タスクにおけるまとめ確認の例

第 3 章

対話コーパス分析

本章では，2 章で述べた本研究で着目する確認応答が実際どのように出現しているのか等を明確にするために，人間同士の対話を対象に行った分析について述べ，その考察も行う．

3.1 分析の目的

本研究で扱う確認応答 (直接確認，間接確認，相槌，まとめ確認) が，どのような場合に出現しているのか，また，円滑な対話を特徴づける要因とは何であるのかを明確にすることを目的に人間同士の対話を対象に対話データの分析をおこなった．

3.2 分析に用いた対話データ

分析に用いた対話データは，以下の通りである．

- セミナー日程予約タスクの対話データ (全 26 対話)
- 9 名がそれぞれシステム役，ユーザ役などの役割を変更して対話を行い作成
- 26 対話中 21 対話が混合主導型，5 対話がシステム主導型

3.3 分析方法

分析は，各予約条件 (人数，日付，曜日，時間，場所) に対して直接確認，間接確認，相槌がどう現れるのかを明確にするため，それぞれの対話ターン数との関係を分析するこ

と，また，まとめ確認が対話中のどこに出現しているのかを明確にするため，その有無と種類の分析を行った．

3.4 対話データの内訳

対話データの内訳は以下のようになった．本研究で分析に用いた対話データの中には，間接確認は現れていなかった．

表 3.1: 直接確認・相槌の内訳

-	直接確認	相槌
人数	22/26(S:22,U:0)	3/26(S:3,U:0)
曜日	22/26(S:22,U:0)	1/26(S:1,U:0)
時間	15/26(S:15,U:0)	10/26(S:10,U:0)
場所	14/26(S:0,U:14)	10/26(S:0,U:10)
日付	12/26(S:0,U:12)	12/26(S:0,U:12)

表 3.2: まとめ確認の種類と内訳

種類	
人数・曜日	2 (U:途中)
人数・場所・時間	1(S)
日付・曜日・場所	2 (S)
日付・曜日・場所・人数	1(S)
曜日・時間	1(S)
時間・曜日・人数	1(S)

3.5 考察

3.5.1 直接確認・相槌の考察

- 直接確認は人数，曜日に関して多く出現していることがわかる

- 場所，日付，曜日に関するものはユーザがあらかじめ決めてはいないので，部分対話ターン数が長くなった
- ユーザ発話に 1 つ以上の予約条件が含まれる場合，始めの条件に対してだけ直接確認をすることがみられた

3.5.2 まとめ確認の考察

- 途中のまとめ確認は人数，曜日が多く確認される
- 人数，曜日は対話中に直接確認された場合でもまとめ確認の項目に含まれる割合が多い (人数 5/8, 曜日 5/8, 場所 3/8，時間 2/8，日付 2/8)

3.5.3 その他の考察

- 本研究で分析に用いた対話データの中には，間接確認は現れていなかった.

次の章では，以上の考察を参考にして作成された対話実験を行うために必要な具体的な実験システムについて述べる.

第 4 章

実験システム

本章では，音声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に，様々な確認応答戦略で対話を進める実験システムの説明を行う．

4.1 本研究の対話タスク

本研究では，会議室の予約を行うことを対話システムのタスクとした．主に対話システムが対話の主導権を持ち，対話システムの質問に対して，ユーザは，予約する人数 (15 名，30 名等)，予約する曜日 (月曜日，金曜日等)，予約時間 (5 時から 6 時まで等) などの予約条件を言っていくものとした．

4.2 実験システムの構成

本研究で用いる実験システムは，主に音声認識モジュールと予約モジュールで構成され(図 1)，対話の主導権は，システムが持つものとした．

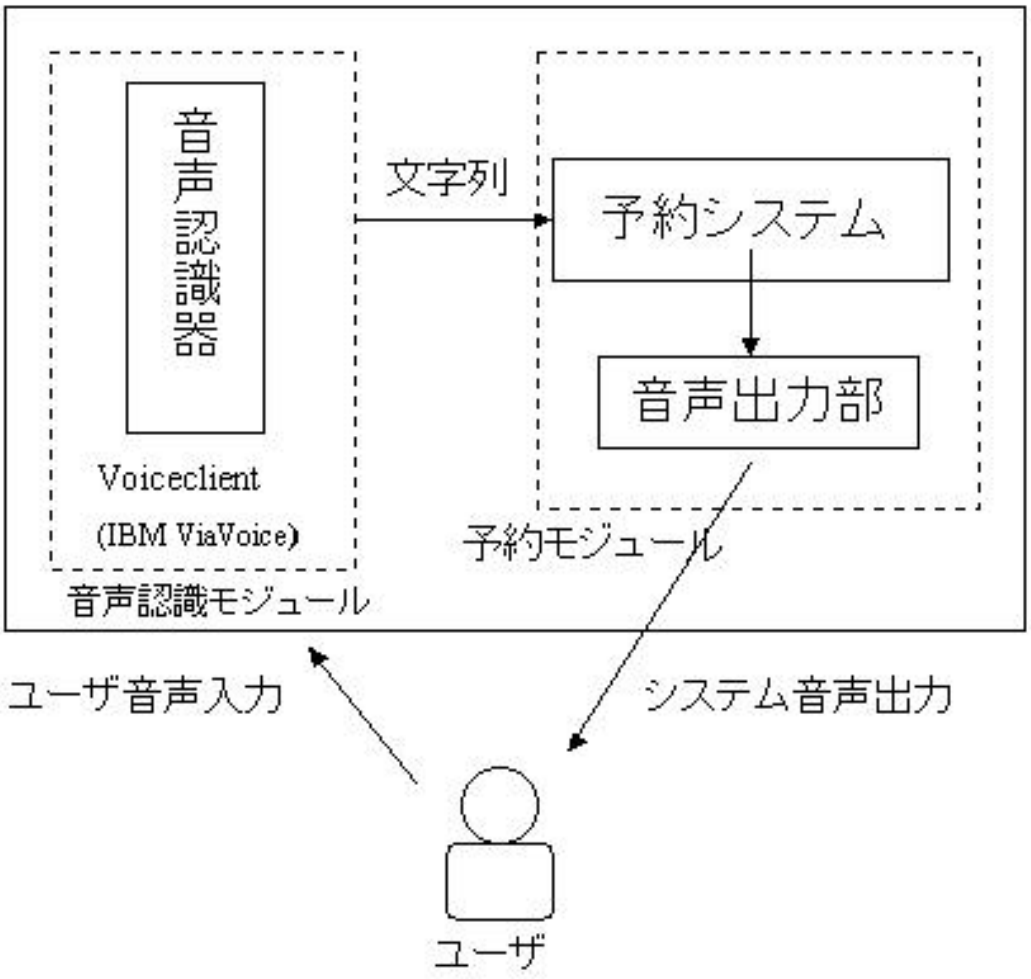


図 4.1: 実験システム概略図

4.3 音声認識モジュール

実験システムの音声認識モジュールは，IBM の Voiceclient を用いている．これは IBM ViaVoice と予約サーバー間の橋渡しを行っているもので，具体的には，ユーザが ViaVoice を使って音声入力を行い，その入力信号を音声認識し，認識した結果 (文字列) を決められたポート番号のネットワーク上にあるコンピュータに送る働きをしている．

Voiceclient には，文法ファイルと辞書ファイルがあり，それらに新たな文法規則や語彙を書き加えるなどの変更することにより，システムが受け付けられる文字列を増やしたり，制限したりすることを可能にしている．

本実験システムでは，文法ファイル，辞書ファイルとも会議室予約タスク用にすべて書き換えを行った．

4.4 予約モジュール

予約モジュールは，常に音声認識モジュールからの入力を待っており，入力がきたと同時に処理を行うようになっている．

予約システムサーバは受理した文字列に対して，文法ファイルと辞書ファイルの参照を行い，キーワードのマッチングをする．そして，そのキーワードマッチングを行った結果応じて，対話システムの応答文を組み立てて，音声 (事前に作成されている wav ファイル) をユーザに返す役割を行っている．さらに，予約モジュールは，ユーザとのやり取り，対話にかかった時間，そして対話ターン数をログファイルとして記憶する役割も行う．

4.5 実験システムの (確認) 応答について

実験システムの応答として，直接確認は,「*でよろしいですか」(ex. 月曜日でよろしいですか等)，間接確認は,「*で*にしますか」(ex.13 名で何曜日にしますか等), 相槌は「はい」とした.

また，実験システム表現は以下のように制約がないものとあるものの 2 種類を人間同士の対話を参考にして考えた.

- システム表現 A
  - 「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)
- システム表現 B
  - 「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)

4.6 実験システムの対話戦略について

人間同士の対話データの確認応答を参考にして，実験システムの対話戦略として，12 の対話戦略を作成した.

- 対話戦略 1
  - 全予約条件に対して，直接確認を行うもの
- 対話戦略 2
  - 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もの
- 対話戦略 3
  - 人数，曜日の予約条件に対して，直接確認をし，時間の条件に対してのみ相槌を打つもの
- 対話戦略 4

  - 人数，曜日の予約条件に対して直接確認をし，時間の条件に対してのみ相槌を打ち，最後にまとめ確認を行うもの

- 対話戦略 5
  - 人数の予約条件に対して直接確認を行い，曜日，時間には相槌を打つもの

- 対話戦略 6
  - 人数の予約条件に対して直接確認を行い，曜日，時間には相槌を打ち，最後にまとめ確認を行うもの

- 対話戦略 7
  - 全予約条件に対して相槌を打つもの

- 対話戦略 8
  - 全予約条件に対して相槌を打ち，最後にまとめ確認を行うもの

- 対話戦略 9
  - 予約条件に対して間接確認を行うもの

- 対話戦略 10
  - 予約条件に対して間接確認を行い，最後にまとめ確認を行うもの

- 対話戦略 11
  - 全予約条件に対して無応答で対話を終了するもの

- 対話戦略 12
  - 全予約条件に対して無応答で最後にまとめ確認を行うもの

第 5 章

実験

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．そこから，対話システムとユーザ間の対話においても，より円滑に対話を進めることができ，また，より使いやすいシステムにするためには対話システムの応答というものが重要な役割をしているのではないかと考えられます．そこで，実際にシステムと対話を行ってもらい，どのようなシステム応答がユーザにどのような効果を与え，その対話にどのように影響するのかを調べることを目的として対話実験を行った．分析をするための実験を行う前に少ない被験者で予備実験を行った．本節では，予備実験，具体的な実験方法を述べる．

5.1 予備実験

対話実験を行うにあたり実験方法の設定について調べることを目的として情報系の大学院生 5 名で予備実験を行った．実験では，全予約条件に対して，直接確認を行う対話戦略 1，全予約条件に直接確認をし，最後にまとめ確認を行う対話戦略 2，全予約条件に対して,「はい」という相槌を打つ対話戦略 7，全予約条件に対して,「はい」という相槌を打ち，最後にまとめ確認を行う対話戦略 8，全予約条件に対して，無応答で対話を進める対話戦略 11，そして，全予約条件に無応答で対話を進め，最後にまとめ確認を行う対話戦略 12 の 6 種類の対話戦略を用いた．

5.1.1 予備実験の目的

予備実験では，本実験に用いるようなデータを取ることが目的ではなく，細かい実験のやり方，実験の環境について調べること，ユーザの表現の調査を目的としている．具体的には以下のことに注目した．

- 実験方法の確認
- ユーザ入力の表現
- 実験の環境の確認
- 音声認識について

5.1.2 実験の方法

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 (AppendixH) 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，被験者には，予約をしてもらう人数，曜日，時間などを明記した予約シートを渡し，その予約シートに明記され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．実際に配布された予約シートを下図に示す．

-予約シート-
人数は 20 名で曜日は金曜日に
時間は 5 時から 6 時までという
条件で予約を行ってください．

図 5.1: 配布した予約シートの例

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各アンケート項目の目的を以下に示す．

- システムとの対話は円滑に進められたか．
  - 対話戦略と対話の円滑性の相関を見る

- システム発話の表現は応答しやすかったか.
  - システム発話の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
- システムは使いやすかったか.
  - どのシステムが使いやすかったのかを見る
- まとめ確認は必要であったか.
  - 対話戦略とまとめ確認の相関を見て必要性を見る

5.1.3 予備実験の結果

ユーザ入力の表現として「1 時から 2 時まで」と「1 時から」と「2 時まで」といった表現をうまく処理ができないことがあった．また，設置式のマイクのため雑音を拾ってしまい，極端に音声認識率が低くなってしまうことがあった．システム発話の音声に関しては，聞きとりにくいといった問題もなかった.

5.1.4 予備実験の考察

アンケート結果には反映されてはいなかったが，対話戦略の順番をランダムに行わなかったために，被験者の学習効果から実験の後半になるにつれてアンケートに影響を及ぼす可能性があるので，対話戦略をランダムにして実験を行うこととした．システムが処理できなかった表現に関しては，文法と辞書を新たに加えることで処理できるようにした．また，本実験では，システムと被験者の対話を録音し，被験者の表現を収集し，新たな表現を処理できるシステムの作成に役立てるようにする．音声認識に関しては，ユーザがヘッドホンマイクを使用することで音声認識の問題を解決することとした.

5.2 実験

実験の方法は，基本的に予備実験の方法と変わらないが，予備実験で得られた問題点を修正した．また，予備実験の時には含まれていなかった予約条件に対して，間接確認を行う対話戦略 9，予約条件に対して，間接確認を行って最後にまとめ確認を行う対話戦略 10，予約条件に対して，部分的に直接確認を行う対話戦略 3，5，そして，対話戦略 3，5 のそれぞれにまとめ確認を行うようにした対話戦略 4，6 の新たに 6 種類の対話戦略を加

え，システム表現 A とさらに制約を設定したシステム表現 B をパラメータとして追加し，合計 24 パターン (12 対話戦略，システム表現 A, B) を 20 歳前後の非情報系大学生 12 名 (男性 4 名，女性 8 名) の被験者で実験を行った．

第 6 章

実験結果

本章では，実験で得られたデータをどのように分析をしたのか，また，その結果がどのようなものであったのかを述べる．分析は，実験で得られたデータを対象に平均点の比較，検定，重回帰分析などをおこなった．

6.1 分析の評価項目

本研究では，実験結果に対する評価の項目として以下の項目を用いた．

- アンケート
- 対話ターン数
- 対話時間
- 繰り返し発話の回数

6.1.1 分析の評価項目の説明

- アンケート
  - 被験者の主観的意見として 5 段階評価によるアンケートを 1 対話が終わる毎に被験者に記入してもらった．アンケートの内容は，システム発話の表現は応答しやすかったか，対話が円滑に進められたか，システムは使いやすかったか，などである．
- 対話ターン数
  - システムと被験者が対話を達成させるまでに要した対話の全ターン数を数えた．これは被験者が，対話を円滑に感じる，またはシステムを使いやすいと感じることに対して，対話ターン数が影響しているのかという対話の円滑性，使いやすさと対話ターン数との相関関係を調べるために数えた．
- 対話時間
  - タスクが達成されるまでにかかった対話時間を計測した．理由は，被験者が，対話を円滑に感じる，またはシステムを使いやすいと感じることに対して，対話時間が影響しているのかという対話の円滑性，使いやすさと対話時間との相関関係を調べるために測定した．
- 繰り返し発話の回数
  - 被験者がシステムと音声で対話を行う際に 1 回では認識されない場合に被験者がシステムに発話が認識されるまで繰り返して行う回数を数えた．理由は，対話時間と繰り返し発話の相関関係を調べるためである．

6.2 実験で得られたデータ

被験者は，20 歳前後の非情報系大学生 12 名 (男 4 女 8) で，それぞれ 24 パターンのシステムと対話をしてもらい合計 288 対話を収録した．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かったので，タスクを達成できなかった 35 対話を除いた 253 対話を対象に分析を行った．

6.3 対話の円滑性の分析

実験から得られた対話の円滑性に関するデータは次の通りである．幾つかの対話戦略とシステム発話の表現の違い，まとめ確認の有無の間での比較を示す．

表 6.1: 対話の円滑性に関する実験データ

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1A	3.917	1.24	5	1	4	5
2A	3.333	1.225	5	2	3	2
3A	3.25	1.485	5	1	3	5
4A	4.364	0.809	5	3	5	5
5A	3.333	1.225	5	1	4	4
6A	3.8	1.033	5	2	4	3
7A	3.667	1.155	5	1	4	4
8A	4	0.853	5	3	4	4
9A	3.636	0.924	5	2	4	4
10A	3.273	1.104	5	2	3	3
11A	3.833	1.03	5	2	4	4
12A	3.833	1.403	5	1	4	5
1B	3.556	1.333	5	1	3	3
2B	4	0.853	5	3	4	3
3B	3.455	1.214	5	2	3	2
4B	3.818	1.168	5	2	4	5
5B	3.455	1.44	5	1	3	5
6B	3.625	1.188	5	2	4	4
7B	4.083	0.793	5	2	4	4
8B	4.167	0.983	5	3	4.5	5
9B	3.364	1.362	5	2	3	2
10B	3.857	1.069	5	2	4	4
11B	3.4	1.265	5	2	3	2
12B	3.833	1.115	5	2	4	4

6.3.1 システム発話の表現の違いによる比較

表 6.2: システム発話の表現の違いによる比較

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1A	3.917	1.24	5	1	4	5
1B	3.556	1.333	5	1	3	3

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2A	3.333	1.225	5	2	3	2
2B	4	0.853	5	3	4	3

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
3A	3.25	1.485	5	1	3	5
3B	3.455	1.214	5	2	3	2

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
4A	4.364	0.809	5	3	5	5
4B	3.818	1.168	5	2	4	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
5A	3.333	1.225	5	1	4	4
5B	3.455	1.44	5	1	3	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
6A	3.8	1.033	5	2	4	3
6B	3.625	1.188	5	2	4	4

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
7A	3.667	1.155	5	1	4	4
7B	4.083	0.793	5	2	4	4

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
8A	4	0.853	5	3	4	4
8B	4.167	0.983	5	3	4.5	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
9A	3.636	0.924	5	2	4	4
9B	3.364	1.362	5	2	3	2

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
10A	3.273	1.104	5	2	3	3
10B	3.857	1.069	5	2	4	4

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
11A	3.833	1.03	5	2	4	4
11B	3.4	1.265	5	2	3	2

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
12A	3.833	1.403	5	1	4	5
12B	3.833	1.115	5	2	4	4

システム表現の違いによる平均点の比較を行った結果，システムの表現の平均点は，システムの確認応答の違いによって変化しているため，全体として対話の円滑性に影響を与えていると考えられるが，システム発話の表現 A，B のどちらがよいか一概に言えない．

最もシステム表現の違いによって違いがみられたのは，全ての予約条件 (人数，曜日，時間) に対して直接確認をして，最後にまとめ確認を行う対話戦略 2 で対話の円滑性を高めるには，システム表現 B を用いることが好まれると考えられる．それに対して，全ての予約条件に対して，無応答で対話を進めていき，最後にまとめ確認を行う対話戦略 12 に関しては，システム表現の違いによる影響は見られなかった．しかし，制約をかけたシステム表現 A と制約のないシステム表現 B のどちらがより対話の円滑性に影響を与える要因であるかは，対話戦略との組合わせで異ることがわかる．ここから対話の円滑性は，システムの表現の違いといったある 1 つのパラメータだけでは，決定することができないことがわかる．

次に，システム表現のみが異る対話戦略の組合わせで，平均点の違いの有為性を t 検定を行うことで調べた．検定に用いた帰無仮説は，“システム表現の違いで平均値に差はない” とした．

検定結果はそれぞれ，対話戦略 1A と 1B では，$t = 0.63 < t(0.05, 17) = 2.10$，対話戦略 2A と 2B では，$t = |-1.39| < t(0.05, 14) = 2.14$，対話戦略 3A と 3B では，$t = |-0.36| < t(0.05, 21) = 2.07$，対話戦略 4A と 4B では，$t = 1.27 < t(0.05, 20) = 2.08$，対話戦略 5A と 5B では，$t = |-0.20| < t(0.05, 18) = 2.1$，対話戦略 6A と 6B では，$t = 0.32 < t(0.05, 14) = 2.14$，対話戦略 7A と 7B では，$t = |-1.03| < t(0.05, 19) = 2.09$，対話戦略 8A と 8B では，$t = |-0.35| < t(0.05, 9) = 2.26$，対話戦略 9A と 9B では，$t = 0.54 < t(0.05, 20) = 2.08$，対話戦略 10A と 10B では，$t = |-1.11| < t(0.05, 13) = 2.16$，対話戦略 11A と 11B では，$t = 0.86 < t(0.05, 17) = 2.1$，対話戦略 12A と 12B では，$t = 0 < t(0.05, 22) = 2.07$ となった．

t 検定の結果，すべての組合わせにおいて有為水準 5%で帰無仮説は棄却されず，システム表現の違いは，対話の円滑性に必ずしも影響するとは言えないことがいえた．

6.3.2 まとめ確認の有無による比較

対話の円滑性にまとめ確認の有無が影響しているのかを調べるため，まとめ確認の有無で組合わせで分析を行った.

表 6.3: まとめ確認の有無による比較

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1A	3.917	1.24	5	1	4	5
2A	3.333	1.225	5	2	3	2
対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
3A	3.25	1.485	5	1	3	5
4A	4.364	0.809	5	3	5	5
対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
5A	3.333	1.225	5	1	4	4
6A	3.8	1.033	5	2	4	3
対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
7A	3.667	1.155	5	1	4	4
8A	4	0.853	5	3	4	4
対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
9A	3.636	0.924	5	2	4	4
10A	3.273	1.104	5	2	3	3
対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
11A	3.833	1.03	5	2	4	4
12A	3.833	1.403	5	1	4	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1B	3.556	1.333	5	1	3	3
2B	4	0.853	5	3	4	3

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
3B	3.455	1.214	5	2	3	2
4B	3.818	1.168	5	2	4	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
5B	3.455	1.44	5	1	3	5
6B	3.625	1.188	5	2	4	4

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
7B	4.083	0.793	5	2	4	4
8B	4.167	0.983	5	3	4.5	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
9B	3.364	1.362	5	2	3	2
10B	3.857	1.069	5	2	4	4

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
11B	3.4	1.265	5	2	3	2
12B	3.833	1.115	5	2	4	4

対話戦略 1A と 2A の組み合わせと対話戦略 9A と 10A の組み合わせ以外の組合わせでは，まとめ確認を行う対話戦略において平均点が高かった．しかし，対話戦略 1A と 2A，対話戦略 9A と 10A のそれぞれの組み合わせに関して，まとめ確認を行わない対話戦略の方が得点が高かった理由は，被験者が音声認識の問題のために，繰り返し発話を行うことで対話時間が長くなってしまったためと考えられる．よって，まとめ確認を行う対話戦略の方が対話の円滑性を高める効果があると考えられる．

次に，まとめ確認の有無による組合わせで，まとめ確認の有無が対話の円滑性に影響を及ぼすのかを検証するため t 検定を行った．検定に用いた帰無仮説は，”まとめ確認の有無で平均値に差はない” とした．

検定結果はそれぞれ，対話戦略 1A と 2A では，$t = 1.07 < t(0.05, 18) = 2.1$ 対話戦略 3A と 4A では，$t = |-2.25| > t(0.05, 17) = 2.1$，対話戦略 5A と 6A では，$t =$

$|-0.89| < t(0.05, 16) = 2.11$，対話戦略 7A と 8A では，$t = |-0.8| < t(0.05, 22) = 2.07$，対話戦略 9A と 10A では，$t = 0.83 < t(0.05, 19) = 2.09$，対話戦略 11A と 12A では，$t = 0 < t(0.05, 22) = 2.07$，対話戦略 1B と 2B では，$t = |-0.87| < t(0.05, 13) = 2.16$，対話戦略 3B と 4B では，$t = |-0.71| < t(0.05, 20) = 2.08$，対話戦略 5B と 6B では，$t = |-0.28| < t(0.05, 17) = 2.1$，対話戦略 7B と 8B では，$t = |-0.18| < t(0.05, 8) = 2.3$，対話戦略 9B と 10B では，$t = |-0.85| < t(0.05, 15) = 2.13$，対話戦略 11B と 12B では，$t = |-0.84| < t(0.05, 18) = 2.1$ となった．

t 検定の結果，3A と 4A の組合わせにおいて有為水準 5%で帰無仮説は棄却され，それ以外の組合わせでは帰無仮説は棄却されなかった．よって 3A の 4A 組み合わせでのみ，まとめ確認が対話の円滑性に影響を及ぼすといえ，それ以外の組合わせでは，対話の円滑性に必ずしも影響を及ぼすとはいえないという結果が得られた．

6.3.3 システム応答 (直接，間接確認，相槌) の違いによる比較

システムの確認応答が対話の円滑性に影響を及ぼすか調べるため，異なる確認応答の組合わせで平均点の比較を行った．さらに，それぞれの有為性を t 検定を行うことで調べた．検定の帰無仮説として，”システム応答の違いで平均値に差はない” を用いた．

- 全予約条件に対して，直接確認を行うものと相槌を打つもの

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1A	3.917	1.24	5	1	4	5
7A	3.667	1.155	5	1	4	4

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1B	3.556	1.333	5	1	3	3
7B	4.083	0.793	5	2	4	4

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2A	3.333	1.225	5	2	3	2
8A	4	0.853	5	3	4	4

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2B	4	0.853	5	3	4	4
8B	4.167	0.983	5	3	4.5	5

平均点の比較を行うと，全予約条件に対して直接確認を行う対話戦略 1A は，全予約条件に対して相槌を行う対話戦略 7A よりも得点が高い．対話戦略 1B と 7B の比較では，1B の方が得点が低くなっているが，これは 1 名の被験者だけが音声認識の問題から対話を円滑に進めることができず，非常に低い得点をつけたためであると考えられる．これらの比較から全予約条件に対して相槌を打つものよりは，直接確認を行うものの方がユーザに評価されていることが読み取れる．

上記と同様の組み合わせに対して，検定を行った．検定結果は，対話戦略 1A と 7A では，$t = 0.51 < t(0.05, 22) = 2.07$，対話戦略 1B と 7B では，$t = |-1.05| < t(0.05, 12) = 2.17$ 対話戦略 2A と 8A では，$t = |-1.39| < t(0.05, 14) = 2.14$，対話戦略 2B と 8B では，$t = |-0.35| < t(0.05, 9) = 2.26$ となり，これらの組み合わせにおいて有為水準 5%で帰無仮説は棄却されなかった．したがって全予約条件に直接確認をするものと相槌を打つもの違いが必ずしも対話の円滑性に影響を及ぼすとはいえないことがいえた．

- 全予約条件に対して，直接確認を行うものと間接確認を行うもの

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1A	3.917	1.24	5	1	4	5
9A	3.636	0.924	5	2	4	4

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1B	3.556	1.333	5	1	3	3
9B	3.364	1.362	5	2	3	2

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2A	3.333	1.225	5	2	3	2
10A	3.273	1.104	5	2	3	3

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2B	4	0.853	5	3	4	3
10B	3.857	1.069	5	2	4	4

平均点の比較より，全予約条件に対して直接確認を行う対話戦略と間接確認を行う対話戦略のどの組合わせにおいても直接確認を行うものの方が平均点が高い．間接確認は，対話のターン数などを減らす効果的なシステムの確認応答であるが対話の円滑性に関しては，高い評価を得ていない事がわかる．

上記と同様の組み合わせに対して検定を行った．検定の結果，対話戦略 1A と 9A では，$t = 0.61 < t(0.05, 20) = 2.08$, 対話戦略 1B と 9B では，$t = 0.31 < t(0.05, 17) = 2.10$

対話戦略 2A と 10A では，$t = |-0.12| < t(0.05, 16) = 2.11$, 対話戦略 2B と 10B では，$t = 0.30 < t(0.05, 10) = 2.28$ であった．

これらの組み合わせにおいても有為水準 5%で帰無仮説は棄却されなかった．したがって，全予約条件に直接確認を行うものと間接確認をするものとの間に有意性は見られなかった．よってこの場合でも，システム応答の違いが対話の円滑性に影響を及ぼすとは必ずしも言えない事がわかった．

- 全予約条件に対して，直接確認を行うものと無応答のもの

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1A	3.917	1.24	5	1	4	5
11A	3.833	1.03	5	2	4	4

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1B	3.556	1.333	5	1	3	3
11B	3.4	1.265	5	2	3	2

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2A	3.333	1.225	5	2	3	2
12A	3.833	1.403	5	1	4	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2B	4	0.853	5	3	4	3
12B	3.833	1.115	5	2	4	2

平均点の比較より，全予約条件に対して無応答のものは，システムが確実に入力した情報を得ているか確信できないためか対話ターン数が増えたとしても，確実に確認が行える直接確認を全予約条件に対して行っているものがユーザに評価されている．

また，最後にまとめ確認を行うのであれば，全予約条件に対して無応答であっても評価されることが，対話戦略 2A と 12A の比較からもわかる．しかし，対話戦略 2B と 12B の場合は，これとは異なる結果が出ている．これはシステムの確認応答戦略に無応答が含まれる場合は，システム表現が影響している可能性があるのではないかと考えられる．ここから全予約条件に直接確認し，最後にまとめ確認をするという組み合わせは，あまりユーザに好まれないことが考えられる．

さらに，上記と同様の組み合わせに対して検定を行った．対話戦略 1A と 11A では，$t = 0.17 < t(0.05, 22) = 2.07$，対話戦略 1B と 11B では，$t = 0.26 < t(0.05, 17) = 2.1$

対話戦略 2A と 12A では，$t = |-0.86| < t(0.05, 18) = 2.1$，対話戦略 2B と 12B では，$t = 0.41 < t(0.05, 21) = 2.07$ という検定結果となった．

これらの組み合わせでも帰無仮説は有為水準 5%で棄却されなかった．したがって全ての条件に直接確認をするものと無応答のものとの間にも有意性は見られず，システムの応答が対話の円滑性に影響を及ぼすとは必ずしも言えない事がわかった．

- 全予約条件に対して，直接確認を行うものとある予約条件に対して部分的に直接確認を行うもの

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1A	3.917	1.24	5	1	4	5
3A	3.25	1.485	5	1	3	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1A	3.917	1.24	5	1	4	5
5A	3.333	1.225	5	1	4	4

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1B	3.556	1.333	5	1	3	3
3B	3.455	1.214	5	2	3	2

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1B	3.556	1.333	5	1	3	3
5B	3.455	1.44	5	1	3	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2A	3.333	1.225	5	2	3	2
4A	4.364	0.809	5	3	5	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2B	4	0.853	5	3	4	3
4B	3.818	1.168	5	2	4	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2A	3.333	1.225	5	2	3	2
6A	3.8	1.033	5	2	4	3

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2B	4	0.853	5	3	4	3
6B	3.625	1.188	5	2	4	4

平均点の比較より，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に比べて平均点が高い．これはユーザが予約したい情報を確実にシステムに伝わっているのかを確認したいと考えているためであると思われる．

また，最後にまとめ確認を行う場合，部分的に直接確認を行うだけでもユーザには評価されると考えられる．しかし，対話戦略 2B と 4B，対話戦略 2B と 6B の比較に関しては，この考えに反する．ここで詳しくデータを調べてみると，対話戦略 4B でも，対話戦略 6B でも 1 名の被験者が音声認識の問題から，対話をうまく進めることができずに他と比べて低い点数をつけたためであると考えられる．

さらに，上記の組合わせに対して検定を行った．検定結果は，それぞれ対話戦略 1A と 3A では，$t = 1.19 < t(0.05, 22) = 2.07$，対話戦略 1A と 5A では，$t = 1.07 < t(0.05, 18) = 2.1$，対話戦略 1B と 3B では，$t = 0.17 < t(0.05, 16) = 2.11$，対話戦略 1B と 5B では，$t = 0.16 < t(0.05, 18) = 2.1$，対話戦略 2A と 4A では，$t = |-2.166| < t(0.05, 13) = 2.160$，対話戦略 2A と 6A では，$t = |-0.89| < t(0.05, 16) = 2.11$，対話戦略 2B と 4B では，$t = 0.42 < t(0.05, 18) = 2.1$，対話戦略 2B と 6B では，$t = 0.77 < t(0.05, 12) = 2.17$ となった．対話戦略 2A と 4A の組合わせにおいて有為水準 5%で帰無仮説を棄却した．それ以外のの組合わせにおいて有為水準 5%で帰無仮説は棄却されなかった．したがって，対話戦略 2A と 4A の組合わせにおいては有為性が見られ，それ以外のシステムの応答戦略は対話の円滑性には影響するとは必ずしもいえない事が言える．

6.3.4 対話の円滑性と対話ターン数の相関の分析 (対話戦略毎)

対話の円滑性と対話ターン数の相関を調べるため，散布図を作成した (AppendixB)．得られた散布図より，対話の円滑性と対話ターン数の相関は，対話ターン数が多いほど，得点が下がるといったことはいえないことが読み取れる．例えば，対話戦略 2B では，対話ターン数はほぼ同じだが，得点が全て異なるものが見られる．このように対話の円滑性と対話ターン数にも相関関係がみられない事がわかる．

6.3.5 対話の円滑性と対話ターン数の相関の分析

対話戦略に関わらず，全体的に対話の円滑性と対話ターン数との相関関係を調べるために散布図を作成した．対話の円滑性と対話ターン数の相関を示す散布図は以下のようになった．

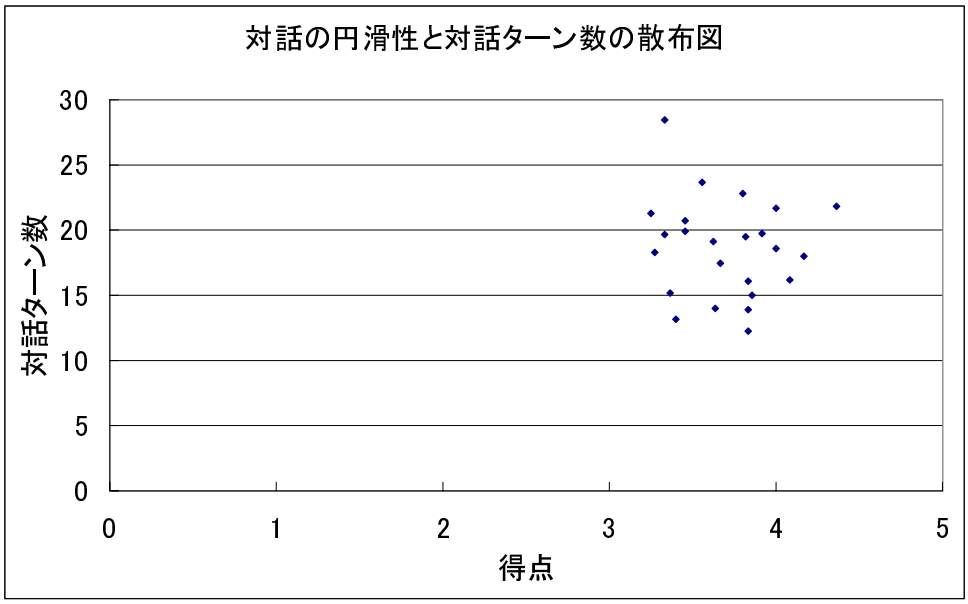


図 6.1: 対話の円滑性と対話ターン数の散布図

対話ターン数の多い少ないに関わらず，得点はバラバラで強い影響を受けているとは考えにくい結果となった．相関係数は，-0.125 であり，ここからも対話の円滑性と対話ターン数との間には強い相関関係がないことがわかる．

6.3.6 対話の円滑性と対話時間の相関の分析 (対話戦略毎)

対話の円滑性と対話時間の相関を調べるため，散布図を作成した (AppendixC)．得られた散布図から，タスクの達成までの対話時間が非常にかかっている場合でも得点が高いもの (対話戦略 11B，12B 等) があったり，対話時間が短くても得点が低いもの (対話戦略 2A，7B 等) があったりとあまり対話の円滑性と対話時間には強い相関関係が見られないことがわかる．

6.3.7　対話の円滑性と対話時間の相関の分析

対話戦略に関わらず，全体的に対話の円滑性と対話時間との相関関係を調べるために散布図を作成した．対話の円滑性と対話時間の関連を表わす散布図は以下のようになった．

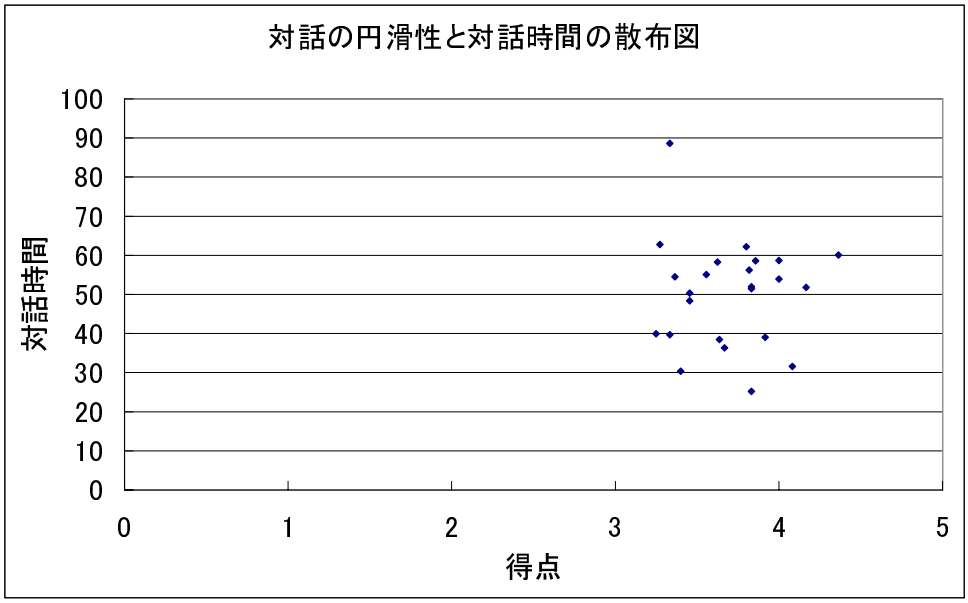


図 6.2: 対話の円滑性と対話時間の散布図

対話の円滑性と対話時間の関係も対話ターン数とのものと同様に対話時間が長くなれば，得点も増えるといったような相関関係など対話時間に応じて得点が影響を受けるといった特徴が見られないことが読み取れる．また，相関係数も-0.046 であることからもあまり相関関係があるとは言えない．

6.4 まとめ確認の必要性の分析

ユーザがまとめ確認が必要であると考えるシステムの確認応答との組合わせはどれかを調べた．また，システム表現の違い，システム応答の違いの比較からまとめ確認の必要性を調べた．実験から得られたまとめ確認に関するデータは次の通りである．

表 6.4: 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	4.222	1.301	5	1	5	5
4A	4.363	1.206	5	1	5	5
6A	4.3	0.674	5	3	4	4
8A	4.416	0.792	5	3	5	5
10A	4.454	0.687	5	3	5	5
12A	3.916	1.378	5	1	4.5	5
2B	4	1.279	5	1	4.5	5
4B	4.636	0.504	5	4	5	5
6B	4.25	0.707	5	3	4	4
8B	4.333	0.816	5	3	4.5	5
10B	4	1.414	5	1	4	4
12B	4.5	0.674	5	3	5	5

6.4.1 システム発話の表現の違いによる比較

表 6.5: システム発話の表現の違いによる比較

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2A	4.222	1.301	5	1	5	5
2B	4	1.279	5	1	4.5	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
4A	4.363	1.206	5	1	5	5
4B	4.636	0.504	5	4	5	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
6A	4.3	0.674	5	3	4	4
6B	4.25	0.707	5	3	4	4

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
8A	4.416	0.792	5	3	5	5
8B	4.333	0.816	5	3	4.5	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
10A	4.454	0.687	5	3	5	5
10B	4	1.414	5	1	4	4

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
12A	3.916	1.378	5	1	4.5	5
12B	4.5	0.674	5	3	5	5

平均点の比較から，システム発話ん表現の違いによって平均点はまばらで，システム発話の表現の制約の違いでまとめ確認の必要性に影響を与えている結果となったが，どちらの表現が良いとは一概に言えない．

システム発話の表現の違いがまとめ確認に影響をしているのかを調べるために検定を行った．帰無仮説は”システム表現の違いによって平均値に差はない”とした．検定結果は，それぞれ対話戦略 2A と 2B では，$t = 0.39 < t(0.05, 17) = 2.1$，対話戦略 4A と 4B では，$t = |-0.69| < t(0.05, 13) = 2.16$，対話戦略 6A と 6B では，$t = 0.152 < t(0.05, 15) = 2.13$，対話戦略 8A と 8B では，$t = 0.20 < t(0.05, 10) = 2.22$，対話戦略 10A と 10B では，$t = 0.792 < t(0.05, 8) = 2.30$，対話戦略 12A と 12B では，$t = |-1.31| < t(0.05, 16) = 2.11$ となり，分析結果から全ての組合わせにおいて，有為水準 5%で帰無仮説を棄却しなかった．したがって，まとめ確認の必要性はシステムの表現の違いに影響を受けるとは必ずしもいえない事が検定結果からも言える．

6.4.2 システム応答 (直接，間接確認，相槌) の違いによる比較

- 全予約条件に直接確認をするものと相槌を打つもの

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2A	4.222	1.301	5	1	5	5
8A	4.416	0.792	5	3	5	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2B	4	1.279	5	1	4.5	5
8B	4.333	0.816	5	3	4.5	5

平均点の比較より，全予約条件に対して相槌を行うものに関して，最後にまとめ確認を行う方が全予約条件に対して直接確認を行うものに比べてユーザに評価されている．このことから，まとめ確認を最後に行う場合には，全予約条件に対して直接確認を行うものは好まれないと考えられる．

システム応答の違いがまとめ確認に影響をしているのかを調べるために検定を行った．帰無仮説は" システム応答の違いによって平均値に差はない" とした．分析結果はそれぞれ対話戦略 2A と 8A では，$t = |-0.39| < t(0.05, 12) = 2.17$，対話戦略 2B と 8B では，$t = |-0.67| < t(0.05, 15) = 2.13$ となった．

分析結果から全ての組合わせにおいて，有為水準 5%で帰無仮説を棄却しなかった．したがって，まとめ確認の必要性はシステムの応答の違いに影響を受けるとは必ずしもいえない事がわかった．

- 全予約条件に直接確認をするものと間接確認をするもの

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2A	4.222	1.301	5	1	5	5
10A	4.454	0.687	5	3	5	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2B	4	1.279	5	1	4.5	5
10B	4	1.414	5	1	4	4

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うものと最後にまとめ確認という組み合わせは，ユーザに評価されないことが平均点の比較からわかる．

システム応答の違いがまとめ確認に影響をしているのかを調べるために検定を行った．帰無仮説は" システム応答の違いによって平均値に差はない" とした．分析結果はそれぞれ対話戦略 2A と 10A では，$t = |-0.81| < t(0.05, 10) = 2.22$，対話戦略 2B と 10B では，$t = 0 < t(0.05, 12) = 2.17$ であった．分析結果から全ての組合わせにおいて，有為水準 5%で帰無仮説を棄却しなかった．したがって，まとめ確認の必要性はシステムの応答の違いに影響を受けるとは必ずしもいえない事がわかった．

- 全予約条件に直接確認をするものと無応答のもの

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2A	4.222	1.301	5	1	5	5
12A	3.916	1.378	5	1	4.5	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2B	4	1.279	5	1	4.5	5
12B	4.5	0.674	5	3	5	5

対話戦略 2B と 12B の比較では，予想通り，全予約条件に対して無応答で対話を進め，最後にまとめ確認を行う組合わせがユーザに支持される結果となっている．一方，対話戦略 2A と 12A の比較では，予想と反している．これはシステムが無応答にも関わらず，音声認識の問題のため，対話時間が長くかかってしまう被験者が多く見られたため評価が低くなってしまったという事がデータからわかった．

システム応答の違いがまとめ確認に影響をしているのかを調べるために検定を行った．帰無仮説は" システム応答の違いによって平均値に差はない" とした．分析結果はそれぞれ対話戦略 2A と 12A では，$t = 0.51 < t(0.05, 18) = 2.1$，対話戦略 2B と 12B では，$t = |-1.19| < t(0.05, 17) = 2.1$ であった．分析結果から全ての組合わせにおいて，有為水準 5%で帰無仮説を棄却しなかった．したがって，まとめ確認の必要性はシステムの応答の違いに影響を受けるとは必ずしもいえない事がわかった．

- 全予約条件に直接確認をするものとある条件に対してだけ部分的に直接確認をするもの

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2A	4.222	1.301	5	1	5	5
4A	4.364	0.809	5	3	5	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2B	4	1.279	5	1	4.5	5
4B	3.818	1.168	5	2	4	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2A	4.222	1.301	5	1	5	5
6A	3.8	1.033	5	2	4	3

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2B	4	1.279	5	1	4.5	5
6B	3.625	1.188	5	2	4	4

平均点の比較より，どの比較においても全予約条件に部分的に直接確認を行い，最後にまとめ確認を行う組合わせが全予約条件に対して直接確認を行うものよりも得点が高い．このことからも全予約条件に対して直接確認し，最後にまとめ確認を行うものの組み合わせがあまり評価されないことがわかる．

システム応答の違いがまとめ確認に影響をしているのかを調べるために検定を行った．帰無仮説は”システム応答の違いによって平均値に差はない”とした．分析結果は，それぞれ対話戦略 2A と 4A では，$t = |-0.24| < t(0.05, 17) = 2.1$，対話戦略 2B と 4B では，$t = |-1.59| < t(0.05, 15) = 2.13$，対話戦略 2A と 6A では，$t = |-0.16| < t(0.05, 12) = 2.17$，対話戦略 2B と 6B では，$t = |-0.56| < t(0.05, 18) = 2.1$ となった．

分析結果から全ての組合わせにおいて，有為水準 5%で帰無仮説を棄却しなかった．したがって，まとめ確認の必要性はシステムの応答の違いに影響を受けるとは必ずしもいえない事がわかった．

6.5 使いやすさの分析

システムの使いやすさがシステム発話の表現の違い，まとめ確認の有無，システム確認応答の違い等によって影響を受けているのかどうかを調べるため，まず，平均点での比較を行い，次にそれぞれの特徴に対して検定をおこなった．さらに，対話時間，対話ターン数との相関を調べ，最後にすべての特徴を用いて重回帰分析をおこない，どの特徴が使いやすさに最も影響を与えているのかを調べた．対話実験から得られたシステムの使いやすさに関するデータは次の通りである．

表 6.6: 使いやすさに関するのデータ

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1A	4.333	0.651	5	3	4	4
2A	3.778	0.972	5	2	4	4
3A	3.833	1.115	5	2	4	4
4A	4.455	0.688	5	3	5	5
5A	4	0.707	5	3	4	4
6A	3.9	1.101	5	2	4	5
7A	3.75	1.138	5	2	4	5
8A	4.083	0.699	5	3	4	4
9A	4	0.894	5	3	4	3
10A	3.727	1.104	5	2	4	4
11A	4.083	0.9	5	3	4	5
12A	4	1.128	5	2	4	5
1B	4.222	0.972	5	2	4	4
2B	4.167	0.718	5	3	4	4
3B	3.909	1.044	5	2	4	5
4B	4.182	0.982	5	2	4	5
5B	3.727	1.272	5	1	4	3
6B	3.875	0.991	5	3	3.5	3
7B	4.25	0.965	5	2	4.5	5
8B	4.167	0.753	5	3	4	4
9B	3.455	1.036	5	2	3	3
10B	4.286	0.756	5	3	4	4
11B	3.9	1.101	5	2	4	5
12B	3.917	0.996	5	2	4	4

6.5.1 システム発話の表現の違いによる比較

表 6.7: システム発話の表現の違いによる比較

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1A	4.333	0.651	5	3	4	4
2B	4.167	0.718	5	3	4	4

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2A	3.778	0.972	5	2	4	4
2B	4.167	0.718	5	3	4	4

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
3A	3.833	1.115	5	2	4	4
3B	3.909	1.044	5	2	4	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
4A	4.455	0.688	5	3	5	5
4B	4.182	0.982	5	2	4	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
5A	4	0.707	5	3	4	4
5B	3.727	1.272	5	1	4	3

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
6A	3.9	1.101	5	2	4	5
6B	3.875	0.991	5	3	3.5	3

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
7A	3.75	1.138	5	2	4	5
7B	4.25	0.965	5	2	4.5	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
8A	4.083	0.699	5	3	4	4
8B	4.167	0.753	5	3	4	4

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
9A	4	0.894	5	3	4	3
9B	3.455	1.036	5	2	3	3

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
10A	3.727	1.104	5	2	4	4
10B	4.286	0.756	5	3	4	4

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
11A	4.083	0.9	5	3	4	5
11B	3.9	1.101	5	2	4	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
12A	4	1.128	5	2	4	5
12B	3.917	0.996	5	2	4	4

平均点の比較より，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．しかし，制約をかけたシステム表現 A と制約のないシステム表現 B のどちらがより対話の円滑性に影響を与える要因であるかは，対話戦略との組合わせで異なり，一概にどちらが良いとは言えない．

次に，システムの表現が使いやすさに影響を与えているのかを見るため，t 検定を行った．帰無仮説は，” システム表現の違いで平均値に差はない” とした．

検定結果は，それぞれ対話戦略 1A と 1B では，$t = 0.29 < t(0.05, 13) = 2.16$，対話戦略 2A と 2B では，$t = |-2.27| > t(0.05, 14) = 2.14$，対話戦略 3A と 3B では，$t = |-0.16| < t(0.05, 21) = 2.07$，対話戦略 4A と 4B では，$t = 0.75 < t(0.05, 20) = 2.08$，対話戦略 5A と 5B では，$t = 0.60 < t(0.05, 16) = 2.11$，対話戦略 6A と 6B では，$t = 0.05 < t(0.05, 16) = 2.11$，対話戦略 7A と 7B では，$t = |-1.16| < t(0.05, 22) = 2.07$，対話戦略 8A と 8B では，$t = |-0.22| < t(0.05, 9) = 2.26$，対話戦略 9A と 9B では，$t = 1.32 < t(0.05, 20) = 2.08$，対話戦略 10A と 10B では，$t = |-1.27| < t(0.05, 16) = 2.11$，対話戦略 11A と 11B では，$t = 0.42 < t(0.05, 17) = 2.1$，対話戦略 12A と 12B では，

$t = 0.19 < t(0.05, 22) = 2.07$ であった．

検定結果より，対話戦略 2A と 2B の組み合わせに対してのみ帰無仮説を棄却し，それ以外の組合わせでは棄却されなかった．したがって，対話戦略 2A と 2B の組み合わせではシステム発話の表現が影響しているといえ，それ以外では必ずしも影響しているとは言えないということがわかった．

6.5.2 まとめ確認の有無による比較

表 6.8: まとめ確認の有無の比較

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1A	4.333	0.651	5	3	4	4
2A	3.778	0.972	5	2	4	4
対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
3A	3.833	1.115	5	2	4	4
4A	4.455	0.688	5	3	5	5
対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
5A	4	0.707	5	3	4	4
6A	3.9	1.101	5	2	4	5
対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
7A	3.75	1.138	5	2	4	5
8A	4.083	0.699	5	3	4	4
対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
9A	4	0.894	5	3	4	3
10A	3.727	1.104	5	2	4	4
対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
11A	4.083	0.9	5	3	4	5
12A	4	1.128	5	2	4	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1B	4.222	0.972	5	2	4	4
2B	4.167	0.718	5	3	4	4

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
3B	3.909	1.044	5	2	4	5
4B	4.182	0.982	5	2	4	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
5B	3.727	1.272	5	1	4	3
6B	3.875	0.991	5	3	3.5	3

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
7B	4.25	0.965	5	2	4.5	5
8B	4.167	0.753	5	3	4	4

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
9B	3.455	1.036	5	2	3	3
10B	4.286	0.756	5	3	4	4

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
11B	3.9	1.101	5	2	4	5
12B	3.917	0.996	5	2	4	4

まとめ確認の有無で平均点の比較を行った．全予約条件に直接確認を行う対話戦略 1A の方がまとめ確認を加えた対話戦略 2A よりも得点が高い．その一方で，全予約条件に相槌，無応答の対話戦略の場合では，まとめ確認を行なう対話戦略の方が得点が高い．ここから全条件に対して直接確認を行う場合には，まとめ確認を行わない方がユーザに評価されていることがわかる．

次にまとめ確認の有無が使いやすさに影響をしているのかを統計的に調べるため t 検定を行った．帰無仮説として，” まとめ確認の有無で平均値に差がない” とした．

検定結果は，それぞれ対話戦略 1A と 2A では，$t = 2.58 > t(0.05, 14) = 2.14$，対話戦略 3A と 4A では，$t = |-1.62| < t(0.05, 19) = 2.09$，対話戦略 5A と 6A では，$t = 0.23 < t(0.05, 15) = 2.13$，対話戦略 7A と 8A では，$t = |-0.87| < t(0.05, 22) = 2.07$，対話戦略 9A と 10A では，$t = 0.63 < t(0.05, 20) = 2.08$，対話戦略 11A と 12A では，

$t = 0.2 < t(0.05, 22) = 2.07$，対話戦略 1B と 2B では，$t = 0.144 < t(0.05, 14) = 2.14$，対話戦略 3B と 4B では，$t = |-0.63| < t(0.05, 20) = 2.08$，対話戦略 5B と 6B では，$t = |-0.28| < t(0.05, 13) = 2.1$，対話戦略 7B と 8B では，$t = 0.2 < t(0.05, 13) = 2.16$，対話戦略 9B と 10B では，$t = |-1.96| < t(0.05, 16) = 2.11$，対話戦略 11B と 12B では，$t = |-0.03| < t(0.05, 18) = 2.1$ となった．

検定結果より，対話戦略 1A と 2A の組み合わせにおいて，有為水準 5%で帰無仮説を棄却し，それ以外の組合わせでは棄却されなかった．よって，1A と 2A 以外の組合わせでは必ずしもまとめ確認の有無が影響しているとは言えない事が言えた．

6.5.3 システム応答 (直接，間接確認，相槌) の違いによる比較

システム応答の違いが使いやすさに影響しているのかを見るために平均点の比較を行った．また，それらが有意な差を持つのかと調べるために，t 検定をおこなった．帰無仮説は，” システム応答の違いで平均値に差はない” とした．

- 各条件すべてに直接確認するものと相槌を打つものの比較

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1A	4.333	0.651	5	3	4	4
7A	3.75	1.138	5	2	4	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1B	4.222	0.972	5	2	4	4
7B	4.25	0.965	5	2	4.5	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2A	3.7	0.9	5	2	4	4
8A	4.0	0.6	5	3	4	4

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2B	4.1	0.7	5	3	4	4
8B	4.1	0.7	5	3	4	4

平均点の比較より，全予約条件に直接確認を行うものが，全予約条件に相槌を打つものに比べて得点が高い．対話戦略 1B と 7B の比較では，相槌を行う方が評価値が高いとい

う結果が出ているが，これは音声認識の問題で音声の認識が悪く，繰り返し発話が多かったため，1 名の被験者が低い得点をつけたからであると考えられる．

t 検定結果は，対話戦略 1A と 7A では，$t = 1.54 < t(0.05, 22) = 2.07$，対話戦略 1B と 7B では，$t = |-0.06| < t(0.05, 17) = 2.1$，対話戦略 2A と 8A では，$t = |-0.81| < t(0.05, 13) = 2.16$，対話戦略 2B と 8B では，$t = 0 < t(0.05, 10) = 2.22$ となった．検定結果より，仮説は有為水準 5%で棄却されなかった．よって，直接確認と相槌の違いが使いやすさに影響しているとは必ずしもいえないことがいえる．

- 各条件すべてに直接確認するものと間接確認するものの比較

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1A	4.333	0.651	5	3	4	4
9A	4	0.894	5	3	4	3

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1B	4.222	0.972	5	2	4	4
9B	3.455	1.036	5	2	3	3

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2A	3.7	0.9	5	2	4	4
10A	3.7	1.1	5	2	4	4

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2B	4.1	0.7	5	3	4	4
10B	4.2	0.7	5	3	4	4

平均点の比較より，全予約条件に直接確認を行う対話戦略では，間接確認をおこなう応答戦略よりも平均点が高い．これは対話の効率を高める効果がある間接確認だが，日常的に用いる機会がすくないため被験者が違和感を感じたためであると考えられる．

t 検定結果は，対話戦略 1A と 9A では，$t = 1.01 < t(0.05, 18) = 2.1$，対話戦略 1B と 9B では，$t = 1.7 < t(0.05, 18) = 2.1$，対話戦略 2A と 10A では，$t = |-0.27| < t(0.05, 17) = 2.1$，対話戦略 2B と 10B では，$t = |-0.33| < t(0.05, 12) = 2.17$ であった．

検定結果より，これらの組み合わせにおいても有為水準 5%で仮説は棄却されなかった．よってシステムの確認応答が使いやすさに影響しているとは必ずしも言えないことがわかった．

- 各条件すべてに直接確認するものと無応答のものの比較

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1A	4.333	0.651	5	3	4	4
11A	4.083	0.9	5	3	4	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1B	4.222	0.972	5	2	4	4
11B	3.9	1.101	5	2	4	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2A	3.7	0.9	5	2	4	4
12A	4.0	1.1	5	2	4	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2B	4.1	0.7	5	3	4	4
12B	3.9	0.9	5	2	4	4

平均点の比較により，予想通り，無応答で対話を進めていくものが，全予約条件に直接確認を行うものに比べて得点が高かったことがわかる．これはユーザが予約したい情報を確実にシステムに伝わっているのか確認したいと考えているためであると思われる．

t 検定結果は，対話戦略 1A と 11A では，$t = 0.77 < t(0.05, 22) = 2.07$，対話戦略 1B と 11B では，$t = 0.67 < t(0.05, 17) = 2.1$，対話戦略 2A と 12A では，$t = |-0.48| < t(0.05, 19) = 2.09$，対話戦略 2B と 12B では，$t = 0.7 < t(0.05, 20) = 2.08$ となった．これらの組み合わせでも仮説は有為水準 5%で棄却されなかった．したがって，システムの確認応答は使いやすさに必ずしも影響するとは言えない．

- 各条件すべてに直接確認するものとある予約条件に対して部分的に直接確認をするのものの比較

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1A	4.333	0.651	5	3	4	4
3A	3.833	1.115	5	2	4	4

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1A	4.333	0.651	5	3	4	4
5A	4	0.707	5	3	4	4

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1B	4.222	0.972	5	2	4	4
3B	3.909	1.044	5	2	4	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
1B	4.222	0.972	5	2	4	4
5B	3.727	1.272	5	1	4	3

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2A	3.7	0.9	5	2	4	4
4A	4.4	0.6	5	3	5	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2A	3.7	0.9	5	2	4	4
6A	3.9	1.1	5	2	4	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2B	4.1	0.7	5	3	4	4
4B	4.1	0.9	5	2	4	5

対話戦略	平均値	標準偏差	最高	最低	中央値	最頻値
2B	4.1	0.7	5	3	4	4
6B	3.8	0.9	5	3	3.5	3

全予約条件に対して直接確認を行うものと予約条件の一部に直接確認を行うものとの平均点の比較を行った．全ての条件に対して直接確認を行う方が一部に直接確認をするものに比べて平均点が高いことがわかる．また，それぞれの戦略に対してまとめ確認を加えた場合には，得点は反対の結果となり，全予約条件に直接確認を行うものとまとめ確認をする組み合わせはあまりユーザに評価されない戦略であることがわかる．

t 検定結果はそれぞれ対話戦略 1A と 3A では，$t = |-0.60| < t(0.05, 22) = 2.07$，対話戦略 1A と 5A では，$t = 1.10 < t(0.05, 17) = 2.1$，対話戦略 1B と 3B では，$t = 0.69 < t(0.05, 18) = 2.1$，対話戦略 1B と 5B では，$t = 0.98 < t(0.05, 18) = 2.1$

対話戦略 2A と 4A では，$t = |-1.75| < t(0.05, 14) = 2.14$，対話戦略 2A と 6A では，$t = |-0.25| < t(0.05, 17) = 2.1$ 対話戦略 2B と 4B では，$t = |-0.04| < t(0.05, 18) = 2.1$，対話戦略 2B と 6B では，$t = 0.71 < t(0.05, 12) = 2.17$ であった．検定結果より，全ての組合わせで有為水準 5%で仮説は棄却されなかった．したがって，システム応答の違いは使いやすさに影響するとは必ずしも言えないことがいえた．

6.5.4 使いやすさと対話ターン数の相関の分析 (対話戦略毎)

使いやすさと対話ターン数の相関関係を調べるために散布図を作成 (AppendixD) した．散布図より，対話ターン数が多いために得点が下がってしまったものもあるが，それと比較して対話ターン数が少ないにもかかわらず，得点が低いものもあるなど，この場合も対話時間との関係と同様に強い相関を持っているとは考えられない事が読み取れる．

6.5.5 使いやすさと対話ターン数の相関の分析

使いやすさと対話ターン数の相関を表わす散布図は以下のようになった．

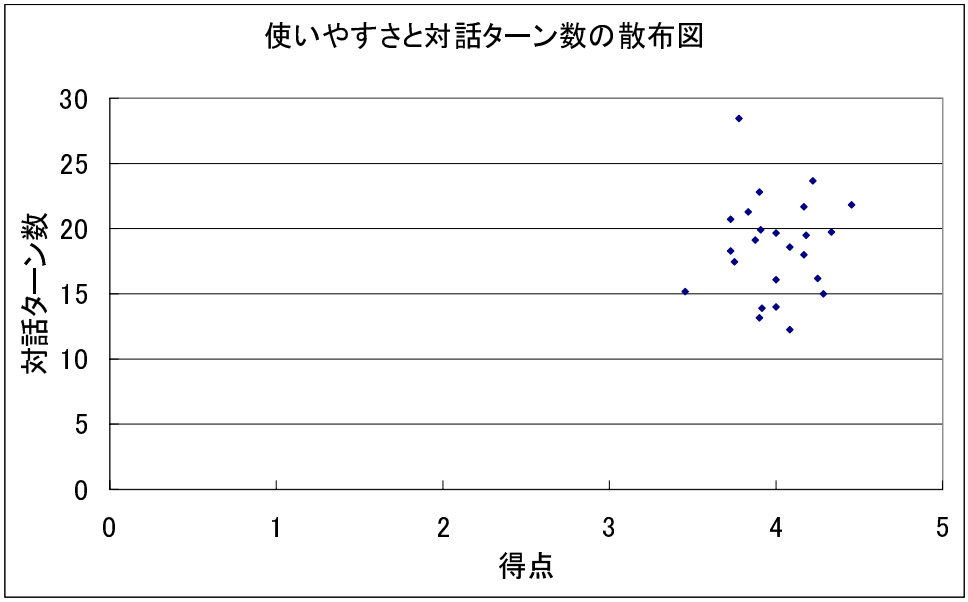


図 6.3: 使いやすさと対話ターン数の散布図

散布図から，使いやすさと対話ターン数との関係はあまりはっきりと関連を見ることはできない．しかし，使いやすさと対話ターン数の相関は他の関係とは異なり，相関係数から (0.035)，正の相関があると考えられるが，そのつながりはほとんど相関がないに等しいものであることがわかる．

6.5.6 使いやすさと対話時間の相関の分析 (対話戦略毎)

使いやすさと対話時間の相関関係を調べるために散布図を作成 (AppendixE) した．散布図より，全体的に対話時間が長いために得点が低くなったり，高くなったりと強い相関

をしているものはない事が読み取れる．この組合わせの関係も対話時間の長短だけで得点に強い影響があるとは考えられないということが読み取れる．

6.5.7 使いやすさと対話時間の相関の分析

使いやすさと対話時間の相関を表わす散布図は以下のようになった．

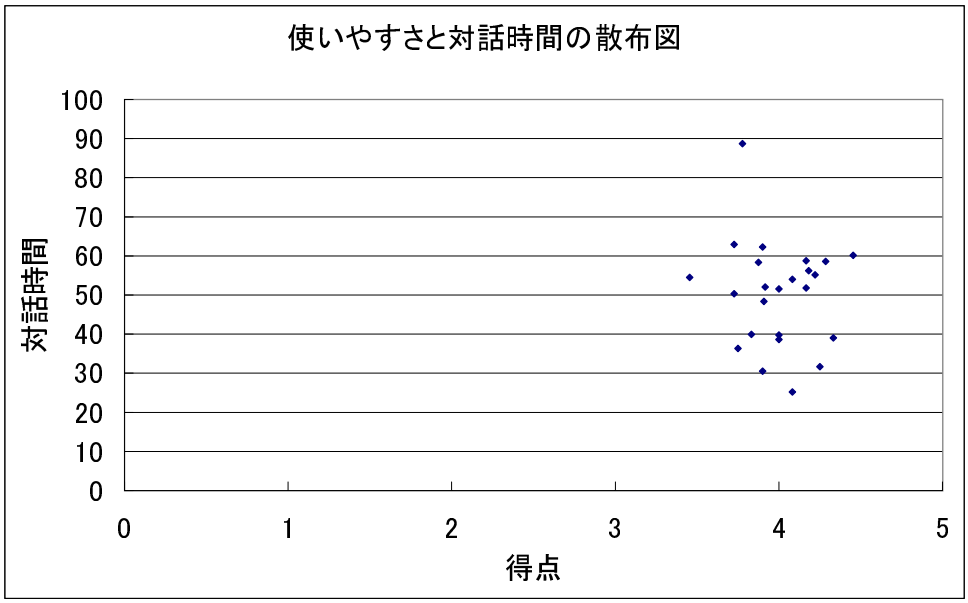


図 6.4: 使いやすさと対話時間の散布図

散布図から，使いやすさと対話時間の関係もあまりお互いを影響している関係を持っているとは言えないことがわかる．使いやすさと対話時間の関係は負の相関関係であるが，相関係数は，-0.12 とあまり強い相関関係を持ってはいないことがいえる．

6.5.8 使いやすさと対話を特徴づける他の要因との相関関係の分析 (対話戦略毎)

各対話戦略毎で重回帰分析を行った．(AppendixF) 重回帰分析に用いた目的変数 Y は，使いやすさ (5 段階評価アンケートで得たもの) 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) として別紙の重回帰式を求め，単位の影響を取るため標準偏回帰係数も求めた．得られた重回帰式から，対話戦略間で異るが，繰り返し発話は比較的，使いやすさに強く影響するとは考えにくい． また，システ

ム表現が使いやすさに影響を与える場合が多い，全体の重回帰式で強い影響を与える要因であった対話の円滑性は，1A, 8A, 9B, 12B を除いて影響は有効ではないということが読み取れる．

6.5.9 使いやすさと対話を特徴づける他の要因との相関関係の分析

前節と同様に，重回帰分析に用いた目的変数 Y は，使いやすさ (5 段階評価アンケートで得たもの) 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) として以下のような重回帰式を求めた．

得られた重回帰式 $Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4$ では，説明変数間で単位が異なるため，単位の影響を受けるので単純に偏回帰係数の大小比較して決めることはできない．そのため単位の影響を除くため，標準偏回帰係数を求めて得られた正規化した式 $Y^* = b_1x_1^* + b_2x_2^* + b_3x_3^* + b_4x_4^*$ を求めて各要因との比較を行った．

$$Y = 0.942 + 0.396x_1 + 0.383x_2 + 0.015x_3 - 0.005x_4 \tag{6.1}$$

$$Y^* = 0.519x_1^* + 0.41x_2^* + 0.252x_3^* - 0.299x_4^* \tag{6.2}$$

得られた重回帰式の標準偏回帰係数の比較を行うことで次のことがわかった．

- 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．
- 対話時間，対話ターン数も使いやすさに影響を及ぼしているがその強さは，対話の円滑性，システム表現よりは低い．

6.6 繰り返し発話と対話時間の相関の分析 (対話戦略毎)

繰り返し発話と対話時間の相関関係を調べるために散布図を作成 (AppendixG) した．散布図より，予想通り，全体的に繰り返し発話と対話時間の相関は比較的強いものが多い事が読み取れる．これはシステム発話の表現やシステムの確認応答戦略毎に相関係数が異るため，これらの要因の影響がある可能性も考えられる．

6.7 繰り返し発話と対話時間の相関の分析

対話時間と繰り返し発話の相関を表わす散布図は以下のようになった．

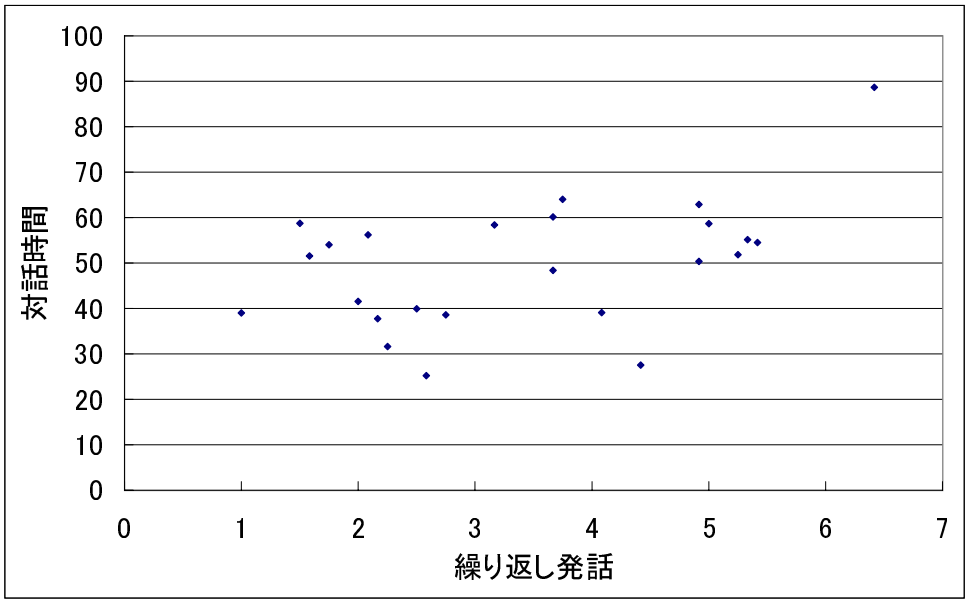


図 6.5: 対話時間と繰り返し発話の散布図

散布図から，繰り返し発話が増えると，対話時間が長くなる傾向がある事が読み取れる．相関係数 (0.463) からもこのことがわかる．

第 7 章

おわりに

7.1 まとめ

対話システムが適切に相槌，確認応答を入れてゆく，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，使いやすさに影響を与えると考えられる様々な要因の中から，まだ明らかにされていないと思われる対話システムの 4 種類の確認応答に着目し，先ず，人間同士の対話を対象に相槌，確認応答などがどのように出現しているのか特徴の分析を行った．次に，システムの確認応答戦略を検討し，実験システムを作成した．そして，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対話実験を行い，分析を行った結果以下のことが得られた．

- 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こと
- 対話の円滑性と対話ターン数，対話の円滑性と対話時間の相関はあまり強く影響している要因ではないこと
- システムの確認応答の違いで，対話ターンを減らす効果的なシステムの確認応答である間接確認が得点が低いことや，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に評価される
- 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うこと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は

言えない．

- 確認応答の違いの比較より，全予約条件に対して直接確認をするものとまとめ確認の組合わせは，冗長に対話を長引かせてしまうためか，ユーザからあまり評価されていない．
- 使いやすさと対話ターン数の相関，また，使いやすさと対話時間の相関の両方とも，あまり強い相関を持っているとは言えない
- 重回帰分析の結果より，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，対話ターン数より繰り返し発話は，使いやすさにあまり影響しない
- 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く，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ことから，適度な確認が使いやすさに必要であること
- 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めではないかと考えられる．

7.2 今後の課題

今後の課題としては，本研究では明らかにならなかったどのようなシステム発話の表現がユーザに影響を与えるのか見るために，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の検討を行うことが必要である．また，割り込み発話などのシステムの確認応答のタイミングについても考慮し検討を行うことが必要である．また，対話の主導権についても検討を行う必要があると考えられる．

謝辞

本研究を進めるにあたり，熱心な御指導をして下さった島津 明 教授に心から感謝致します．また，白井清昭 助教授，望月源 助手には適切な御助言を頂き，深く感謝致します．修士研究中間審査の折には、東条 敏 教授から貴重な御助言を頂き，深く感謝致します．

システム実装に関して議論し，貴重なアドバイスをして頂きました東京工業大学大学院精密工学研究所 藤澤 瑞樹さん，本学 落水研究室 岸 三樹夫さんに感謝致します．日頃から有益な議論をし，貴重な意見を頂いた本学 修了生 植松 秀樹さん，橋本 力さんに感謝致します．

さらに，対話実験に協力頂いた北陸大学の皆様，並びに数々の貴重な御意見，御討論，アドバイスをいただきました島津・白井研究室の皆様に心から御礼を申し上げます．

最後に本研究を行うにあたり，励まし，また経済的援助をして頂いた両親に深く感謝致します．

2002 年 2 月 15 日

市野 貴之

参考文献

[1] J.Hirasawa, M.Nakano, T.Kawabata, K.Aikawa. “Effects of System Barge-in Responses on User Impressions”, Eurospeech’99, pp.1391-1394, 1999.

[2] 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.

[3] 小林 豊，中島大輔，新美康永.「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.

[4] J.Hirasawa, N.Miyazaki, M.Nakano, A.Kiyoshi. “New Feature Parameters For Detecting Misunderstandings in a Spoken Dialogue System”, Proc.of 6th International Conference of Spoken Language Processing (ICSLP2000), vol2, pp154-157, 2000.

[5] Y. Okato, K.Kato, M.Yamamoto, S.Itabashi. “System-User Interaction and Response Strategy in Spoken Dialogue System”, Proc. of International Conference of Spoken Language Processing’98(ICSLP’98),vol2, pp.495-498, 1998.

[6] 堂坂 浩二，中野 幹生，宮崎 昇，安田 宜仁，相川 清明.「制限知識下における効率的対話制御」, 自然言語処理学会 第 6 回年次大会発表論文集. ,pp.340-343, 2000.

[7] Stephan M. Marcus, Deborah W. Brown, Randy G. et al, “Prompt Constrained Natural Language - Evolving the Text Generation of Telephony Servises”, Proc. of International Conference of Spoken Language Processing’96(ICSLP’96), pp.857-860, 1996.

[8] Gie Bouwnman, Joris Hulstijn. “Dialogue Strategy Redesign with Reliability Measures” , Proc.of the First International Conference on Language Resouses & Evaluation, 1998.

[9] Sophie Rosset, Samir Bennacef, Lori Lamel. “Design Strategies For Spoken Language Dialogue Systems”, Proc.of Eurospeech’99, 1999.

[10] L.Devillers, H. Bonneau-Maynard, “Evaluation of Dialogue Strategies for a Tourist Information Retreval System”, Proc. of International Conference of Spoken Language Processing’98(ICSLP’98), pp1187-1190, 1998.

[11] M.A. Walker, D.Litman, C.Kamm, and A.Abella. “PARADASE: A general framework of for evaluating spoken dialogue agents”, Proc.of the 35th Annual Meeting of the ACL, pp.271-280, 1997.

[12] D.Litman, S.Pan, M.A.Walker. “Evaluating Response Strategies in a Web-Based Spoken Dialogue Agent” , Proc. of 36th Annual Meeting of the ACL/COLING’98, pp780-787, 1998.

[13] M.A.Walker. “Experimentally Evaluating Communicative Strategies: The Effect of the Task” AAAI’94, 1994.

[14] M.A.Walker, D.Hindle, J.Fromer, G.Di Fabbrizio and C.Mestel. “Evaluating Competing Agent Strategies for a Voice Email Agent”, Proc. of the 5th European Conference on Speech Communication and Technology(EUROSPEECH’97), pp22-25, 1997.

[15] M.A.Walker, D.Litman, C.Kamm and A.Abella. “Evaluating Spoken Dialog Agents with PARADISE: Two case studies”, Computer Speech and Language,vol12, pp317-348, 1998.

[16] Ronnie W. Smith, “An Evaluation of Strategies for Selective Utterance Verification for Natural Language Dialog”, Proc.of the Fifth Conference on Applied Natural Language Processing, pp41-48,1996.

[17] 鳥居泰彦.「はじめての統計学」，日本経済新聞社，1998.

[18] 菅民郎.「アンケートデータの分析」，現代数学社，1998.

[19] 郡山彬，和泉澤正隆.「統計・確率のしくみ」, 日本実業出版社，1998.

[20] 田村博. 「ヒューマンインターフェース」，オーム社，1998.

[21] Wlliam M. Newman, Michel G. Lamming，北島宗雄 監訳.
「インタラクティブシステムデザイン」, 株式会社ピアソンエデュケーション，1999.

AppendixA

実験シート

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．そこから，対話システムとユーザ間の対話においても，より円滑に対話を進めることができ，また，より使いやすいシステムにするためには対話システムの応答というものが重要な役割をしているのではないかと考えられます．そこで，実際にシステムと対話を行ってもらい，どのようなシステム応答がユーザにどのような効果を与え，その対話にどのように影響するのかを調べるために今回，会議室予約をタスクとする対話実験システムの作成をしました．

実験目的

今回の対話実験では，ユーザの応答の表現の調査とユーザの音声入力に対するシステムの応答がユーザに与える効果，システムのタスク達成率，対話効率などを調査することを目的とする．

対話実験の方法について

1. 状況の設定：
   あなたは会社の企画会議のために会議室の予約をしなくてはいけません．そこで音声会議室予約システムを使って予約します．

   そのシステムは，システムの指示にしたがって対話を行い，予約を完了するもので，あなたは人数，曜日，時間を音声で入力しなければなりません．

2. 準備：

   - 音声認識システムに ViaVoice を用いています．この音声認識率を上げるためにクイックエンロールを行って，自分の声の登録をしてください．(この際の細かい設定はこちらで行います)

3. 対話実験のやり方：(実験を始めるための設定はこちらで行います)

   - はじめに予約してもらう人数，曜日，時間などを明記した予約シートをお渡しするので，それに応じてマイクを通した音声のみでシステムと対話を行い予約してください．
     (音声認識が成功した時,「Accepted:文字列」と表示され，システムは処理を行います.「Rejected:文字列」の時は音声認識エラーなのでシステムは何も応答してこないので再度，予約条件の入力をしてください．)
   - 条件を変えて 6 ないし 7 回対話システムと対話を行って頂きます．

   - 音声認識がうまく行かない場合は，タスク途中でも辞めて頂いて結構です．
     その際にタスクのどこで止まってしまったか明確にしてください．
     (例えば，人数に対する音声認識がうまく行かなかったなど)

4. ユーザ満足度に関するアンケート:

   - 最後にユーザ満足度に関する 5 段階評価のアンケートに答えてください．

ユーザ満足度に関するアンケート (reservel*.pl)

システムの使いやすや，対話の円滑さに関するアンケートに答えてください．評価項目は以下の 3 つでそれぞれ 5 段階評価を行ってください．評価'1' が最低で数字が高くなるにつれて良くなり，評価'5' が最高とします．

1. システム発話の表現は話しやすかったか.
   (話しにくい)1—2—3—4—5(話しやすい)

2. システムとの対話は円滑に進められたか.
   (ぎこちなかった)1—2—3—4—5(円滑であった)

3. システム全体の使いやすさ.
   (使いにくい)1—2—3—4—5(使いやすい)

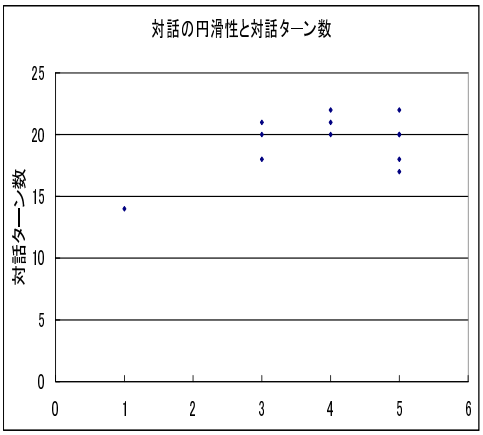
4. システムが最後に行うまとめ確認は必要か.
   (いらない)1—2—3—4—5(いる)

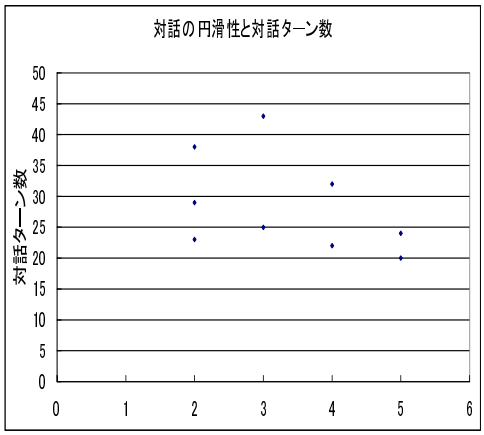
その他のコメント

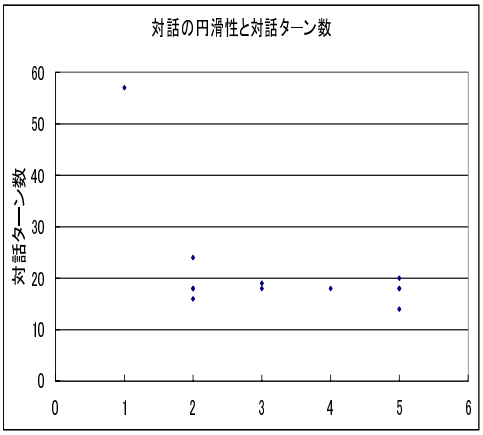
音声認識率やその他なんらかの問題によってタスクを達成できないで終わった場合，どこで困難が起ったのか，など何でもコメントを書いてください．

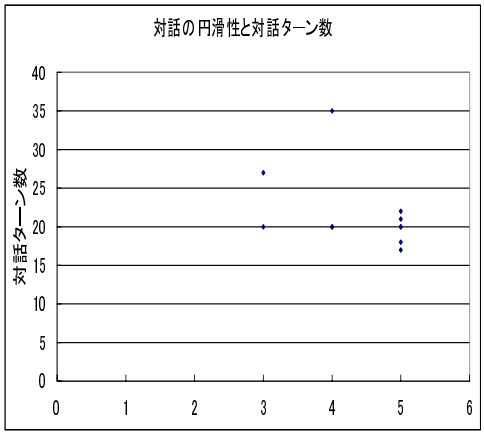
AppendixB

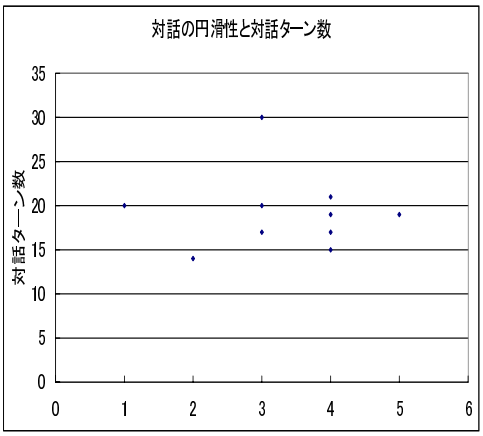
対話の円滑性と対話ターンの散布図 1(対話戦略毎)

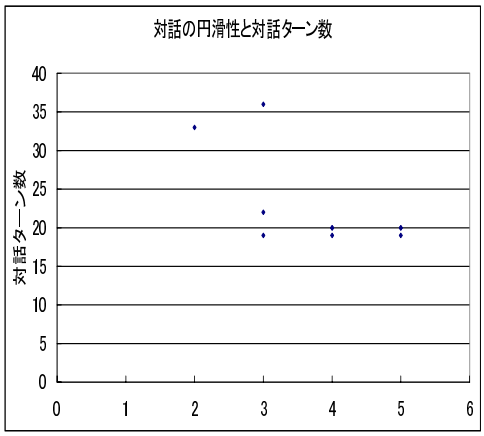


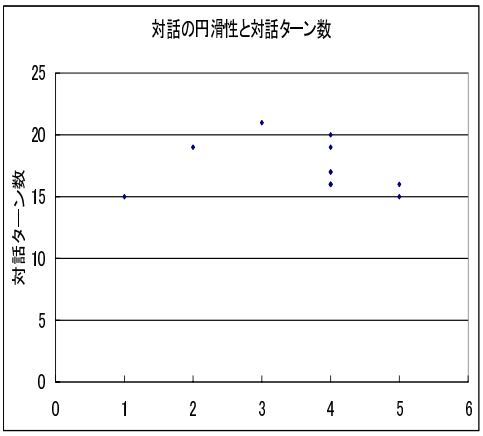


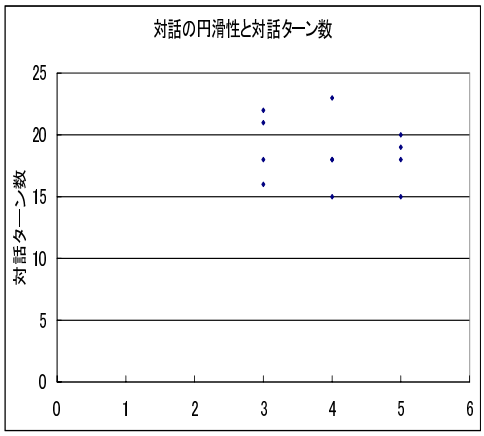





対話の円滑性と対話ターン数
対話ターン数
35
30
25
20
15
10
5
0
0
1
2
3
4
5
6


対話の円滑性と対話ターン数
対話ターン数
40
35
30
25
20
15
10
5
0
0
1
2
3
4
5
6


対話の円滑性と対話ターン数
対話ターン数
25
20
15
10
5
0
0
1
2
3
4
5
6


対話の円滑性と対話ターン数
対話ターン数
25
20
15
10
5
0
0
1
2
3
4
5
6

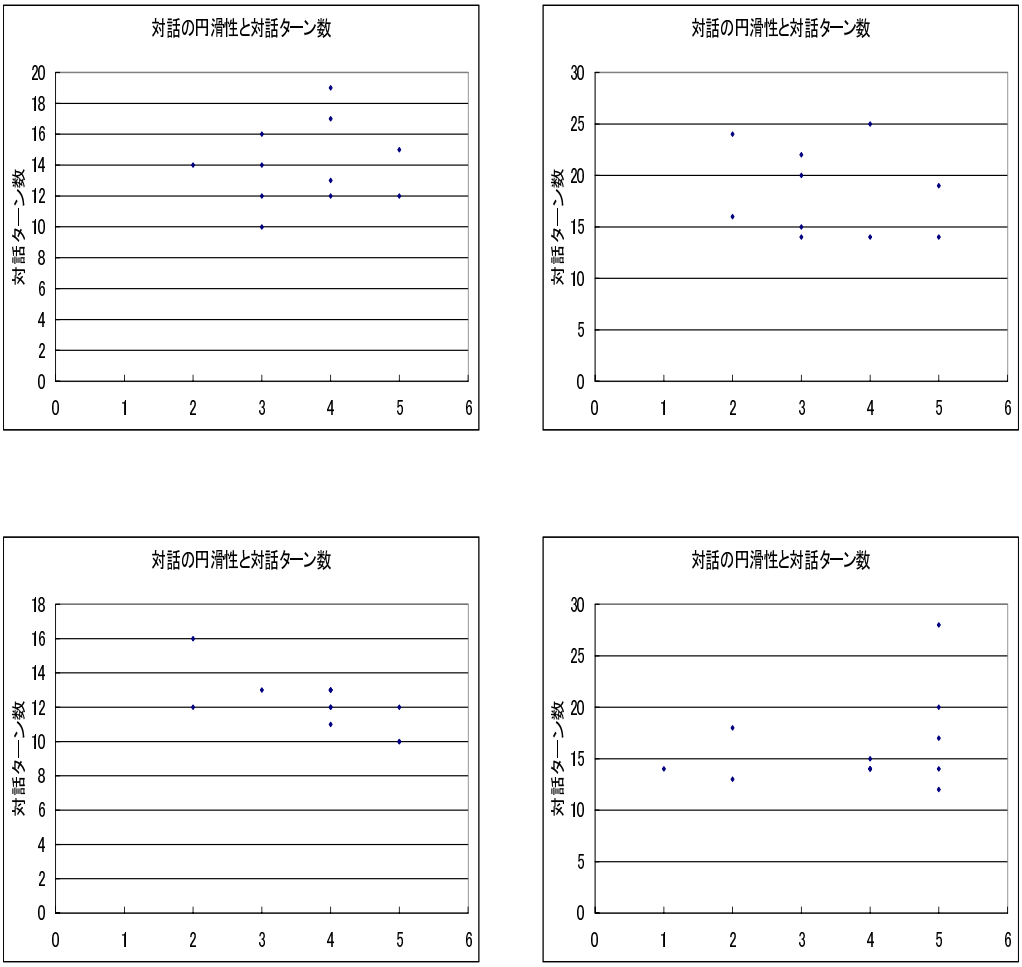
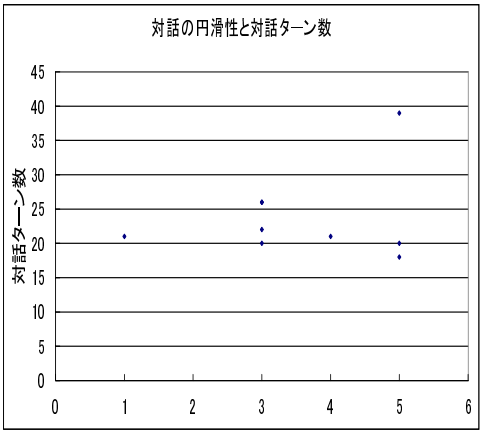
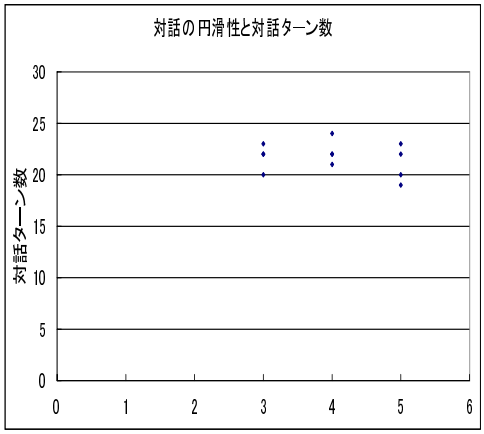
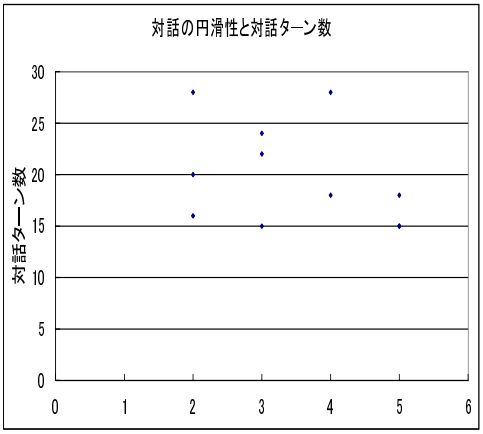


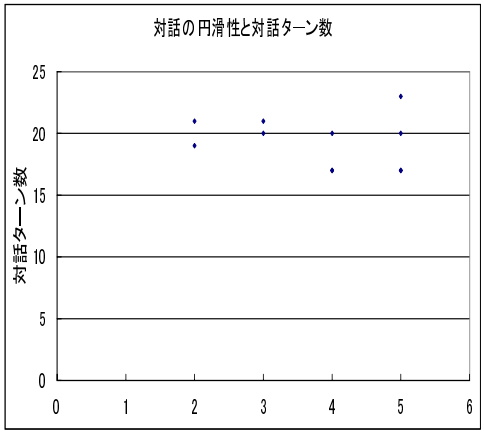
図 .1: 対話の円滑性と対話ターンの散布図 (1A〜12A)

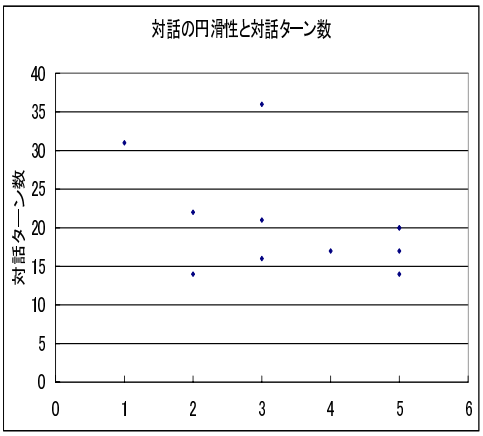
対話の円滑性と対話ターンの散布図 2(対話戦略毎)

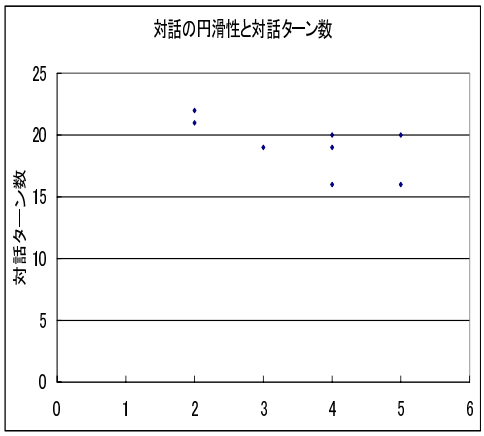


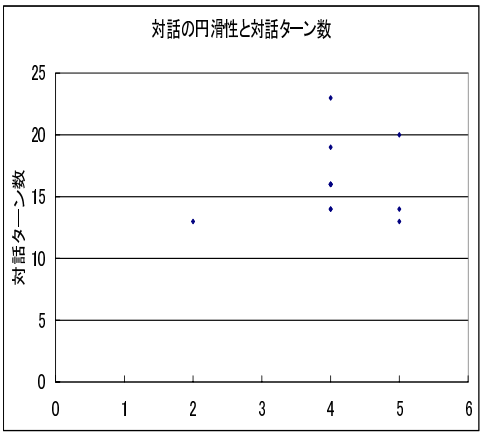


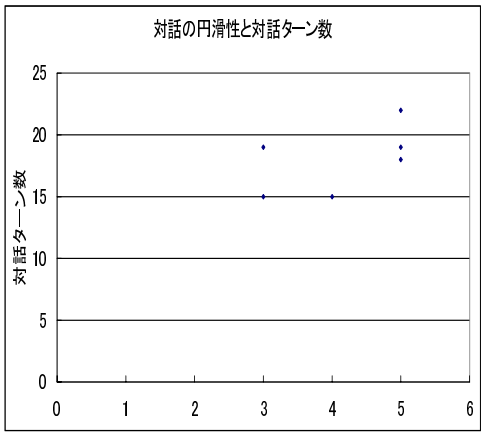





対話の円滑性と対話ターン数
対話ターン数
0
5
10
15
20
25
30
35
40
0
1
2
3
4
5
6


対話の円滑性と対話ターン数
対話ターン数
0
5
10
15
20
25
0
1
2
3
4
5
6


対話の円滑性と対話ターン数
対話ターン数
0
5
10
15
20
25
0
1
2
3
4
5
6


対話の円滑性と対話ターン数
対話ターン数
0
5
10
15
20
25
0
1
2
3
4
5
6

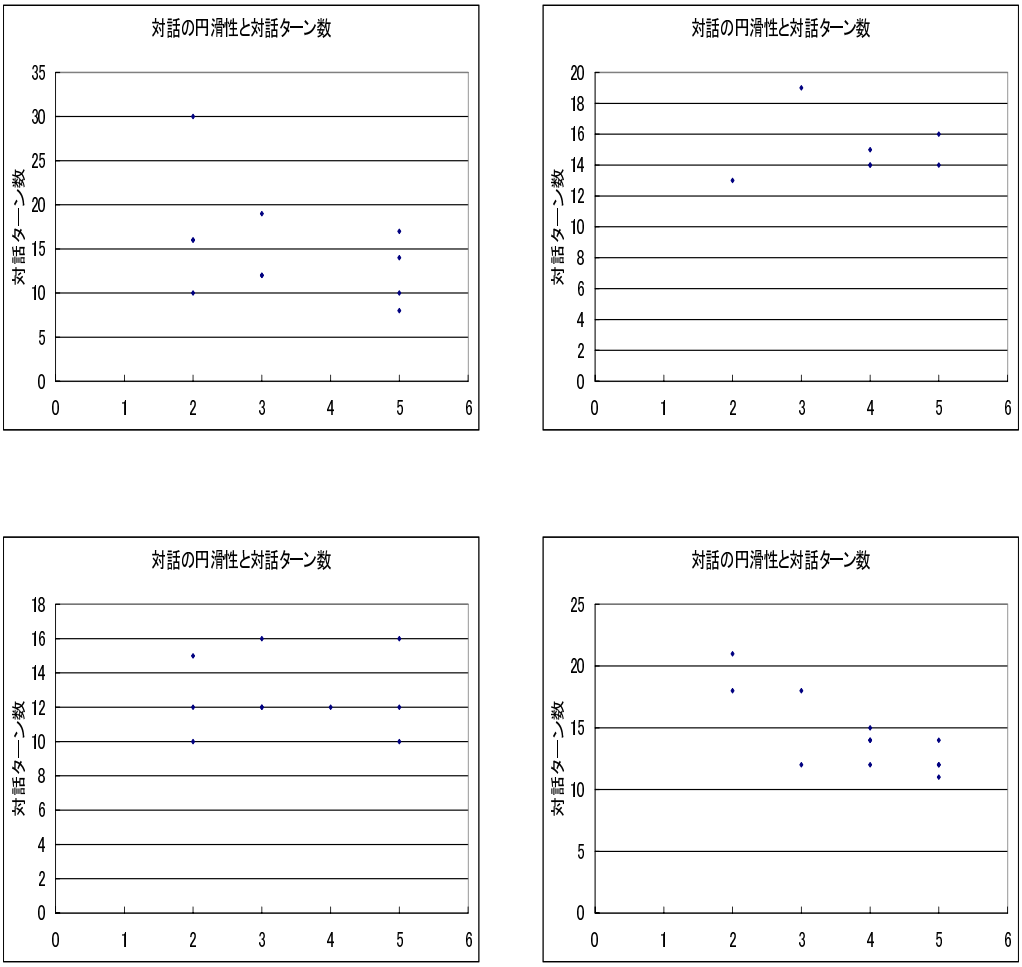


図 .2: 対話の円滑性と対話ターンの散布図 (1B〜12B)

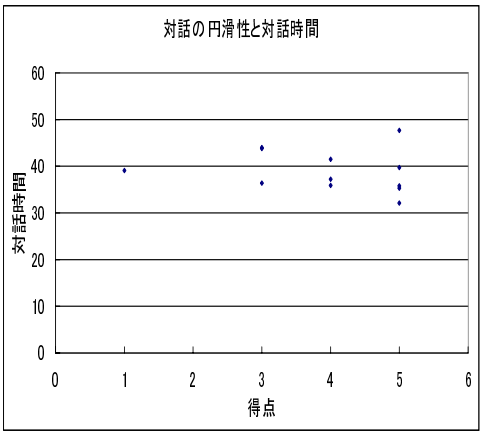
対話の円滑性と対話ターンの相関係数 (対話戦略毎)

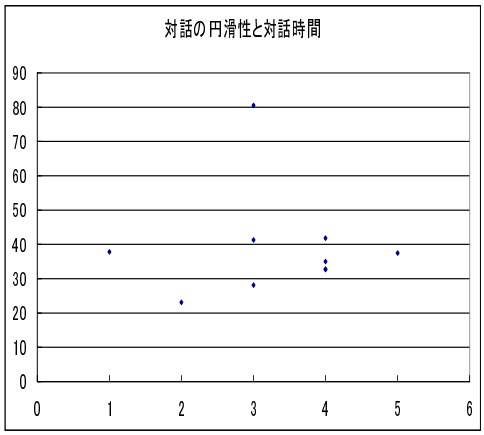
対話戦略	相関係数	対話戦略	相関係数
1A	0.165	1B	0.216
2A	-0.434	2B	-0.223
3A	-0.517	3B	-0.357
4A	-0.397	4B	-0.227
5A	0.209	5B	-0426
6A	-0.659	6B	-0.645
7A	-0.644	7B	0.180
8A	-0.207	8B	0.53
9A	0.124	9B	-0.366
10A	-0.197	10B	0
11A	-0.688	11B	0.127
12A	0.283	12B	-0.897

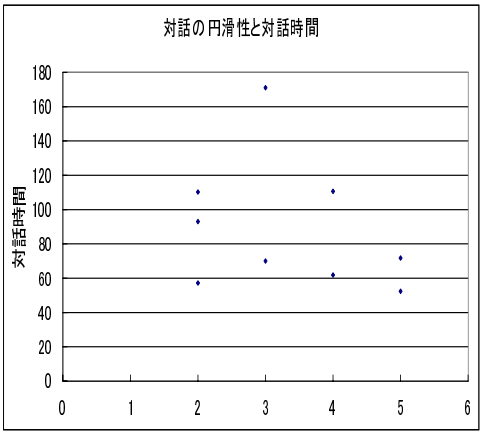
表 .1: 対話の円滑性と対話ターン数の相関係数

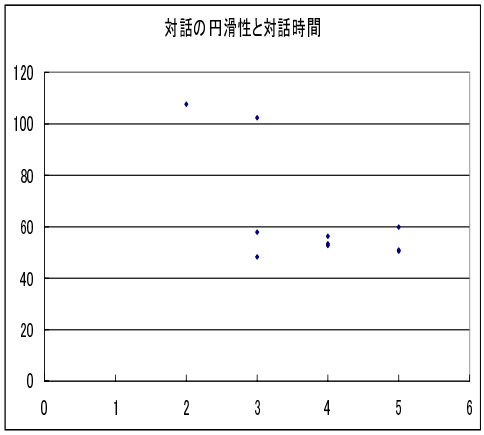
AppendixC

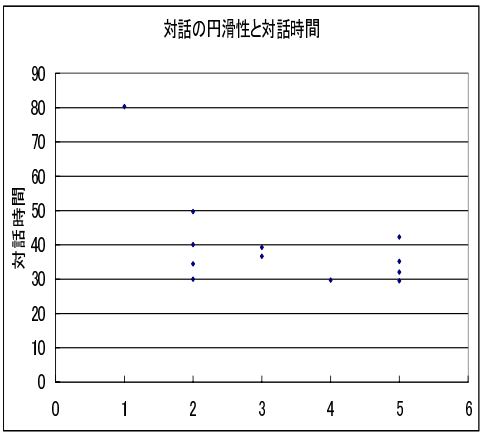
対話の円滑性と対話時間の散布図 1(対話戦略毎)

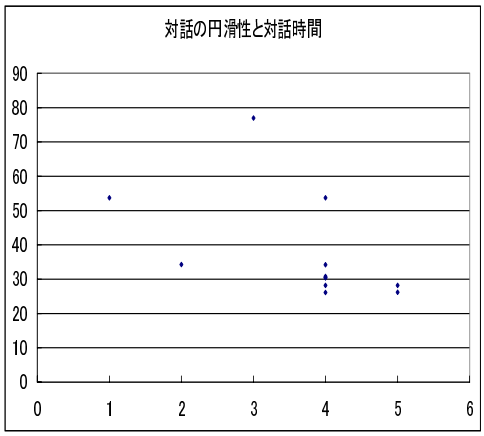


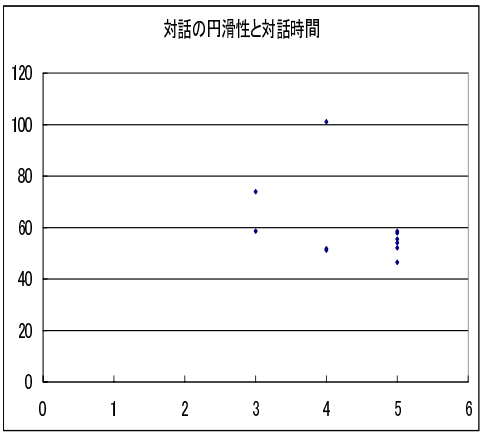


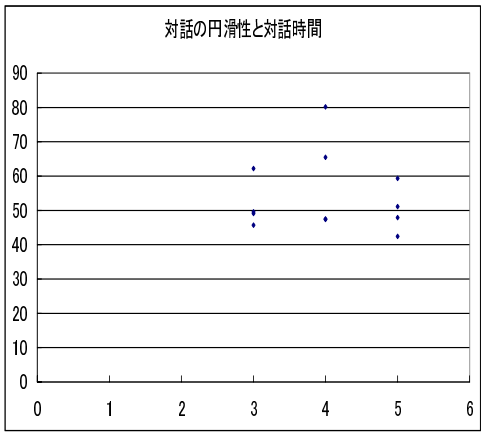





対話の円滑性と対話時間
対話時間
90
80
70
60
50
40
30
20
10
0
0
1
2
3
4
5
6


対話の円滑性と対話時間
90
80
70
60
50
40
30
20
10
0
0
1
2
3
4
5
6


対話の円滑性と対話時間
120
100
80
60
40
20
0
0
1
2
3
4
5
6


対話の円滑性と対話時間
90
80
70
60
50
40
30
20
10
0
0
1
2
3
4
5
6

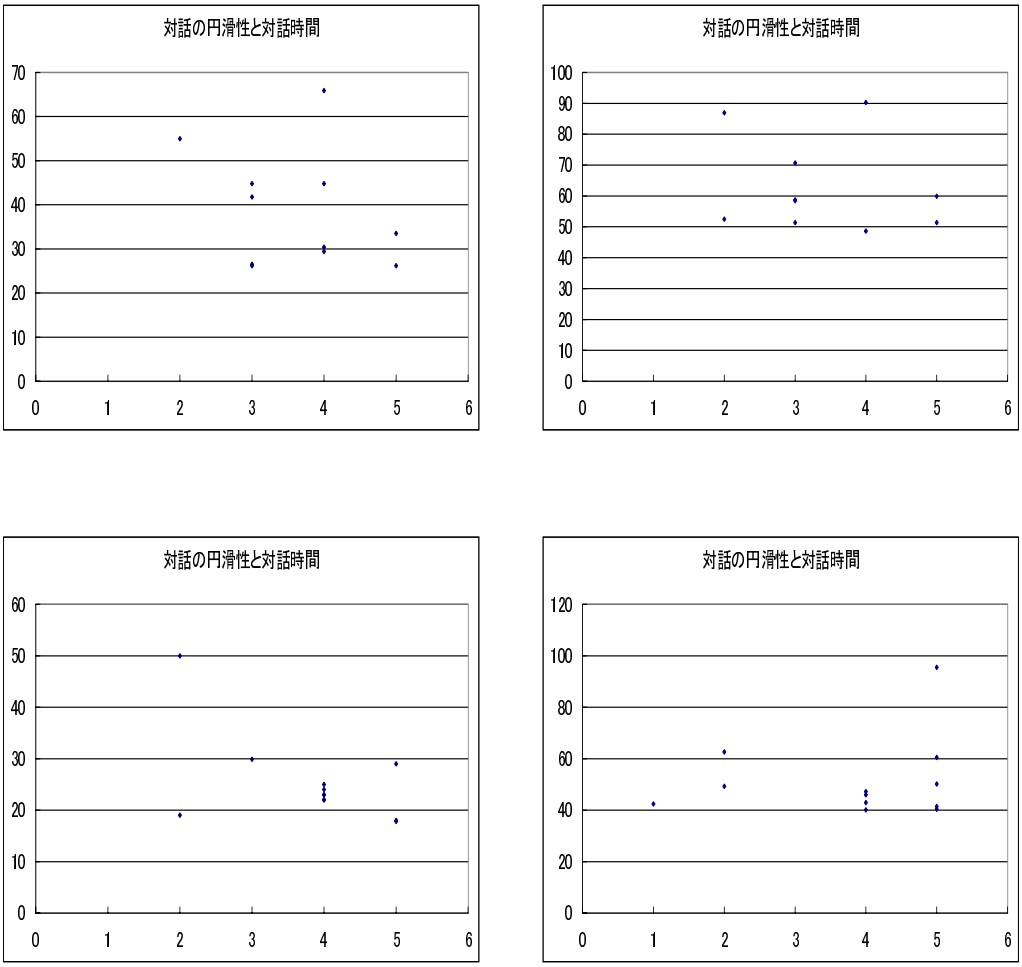
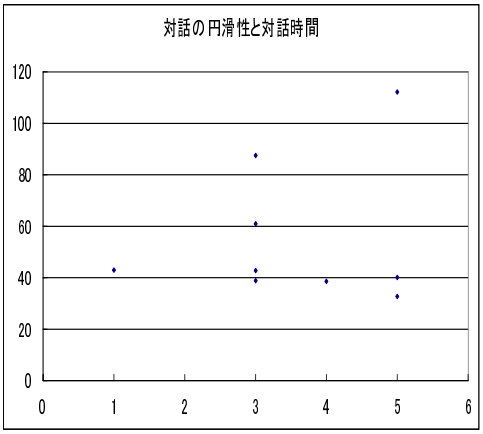
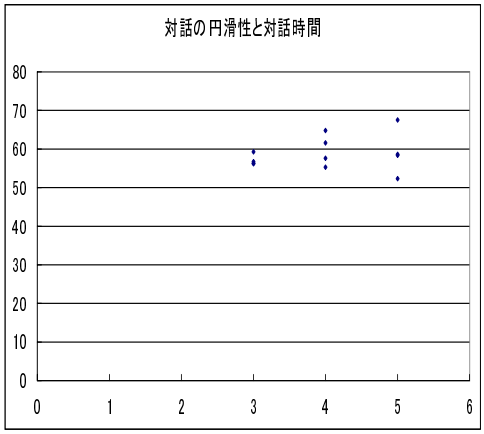
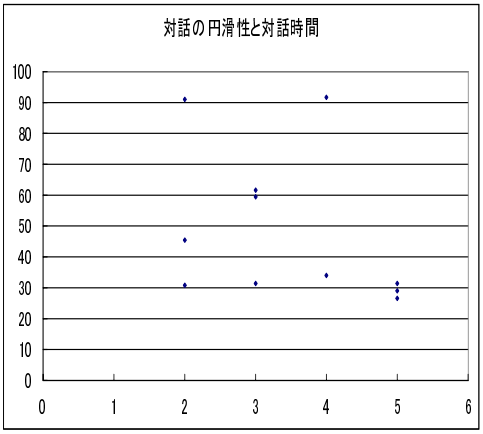


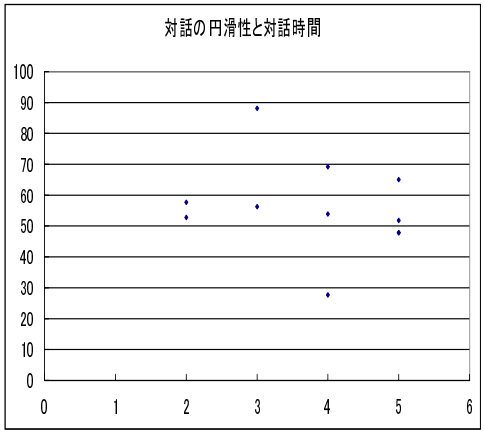
図 .3: 対話の円滑性と対話ターンの散布図 (1A〜12A)

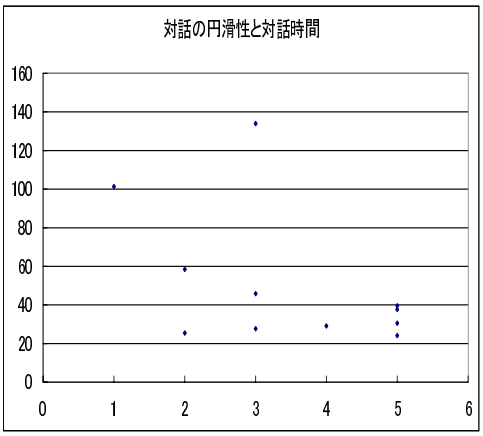
対話の円滑性と対話時間の散布図 2(対話戦略毎)

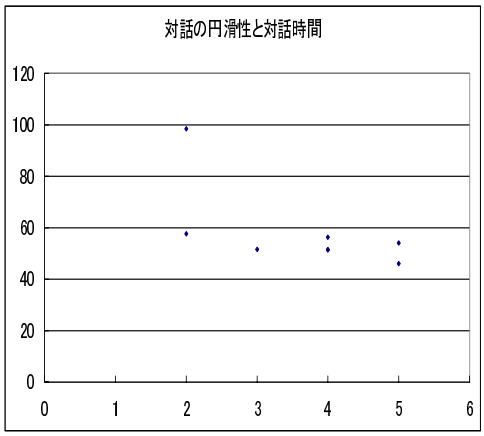


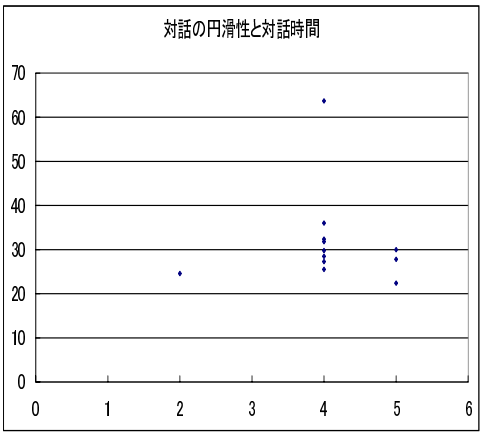


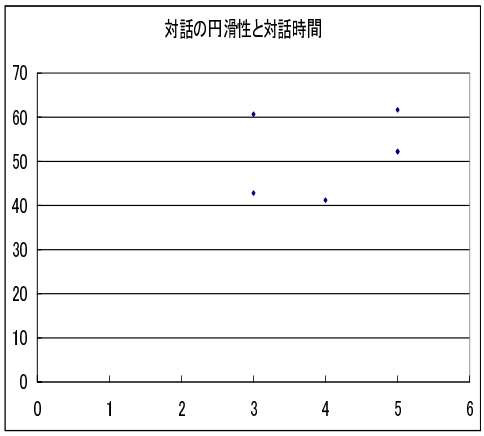





対話の円滑性と対話時間
160
140
120
100
80
60
40
20
0
0
1
2
3
4
5
6


対話の円滑性と対話時間
120
100
80
60
40
20
0
0
1
2
3
4
5
6


対話の円滑性と対話時間
70
60
50
40
30
20
10
0
0
1
2
3
4
5
6


対話の円滑性と対話時間
70
60
50
40
30
20
10
0
0
1
2
3
4
5
6

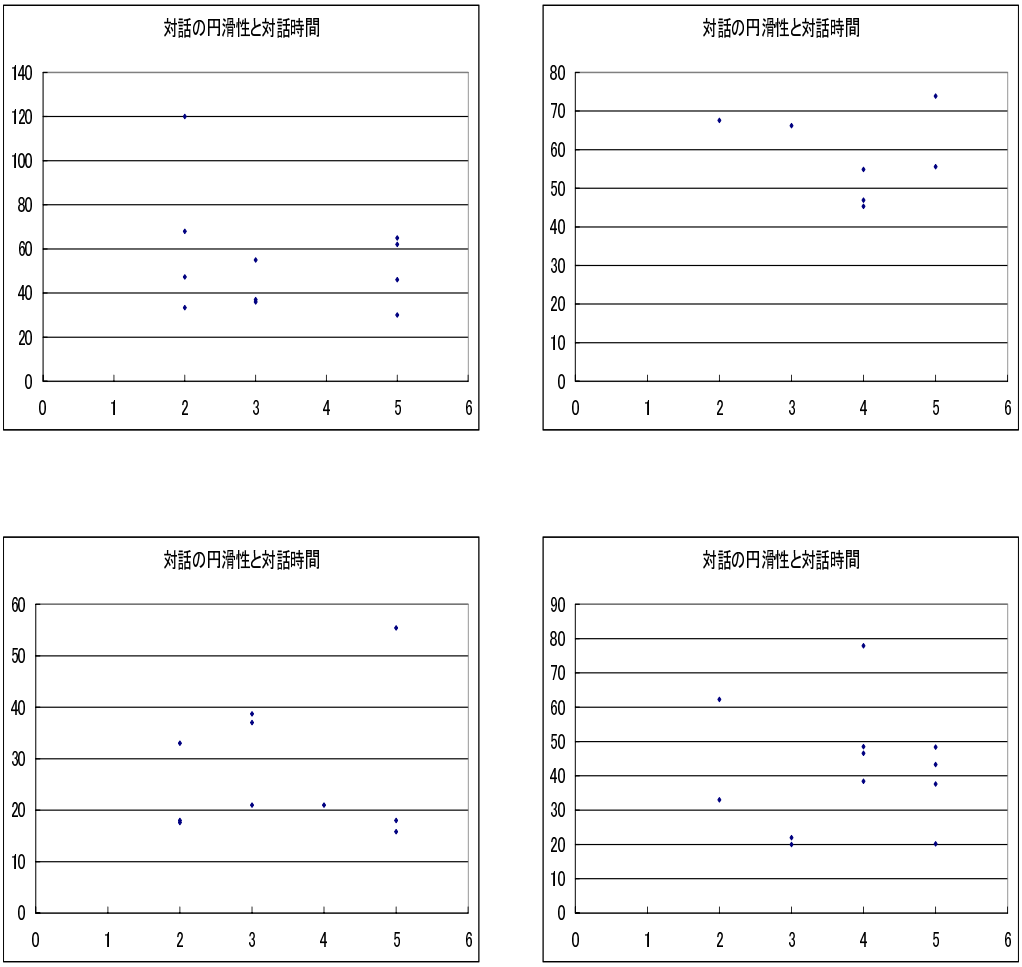


図 .4: 対話の円滑性と対話ターンの散布図 (1B〜12B)

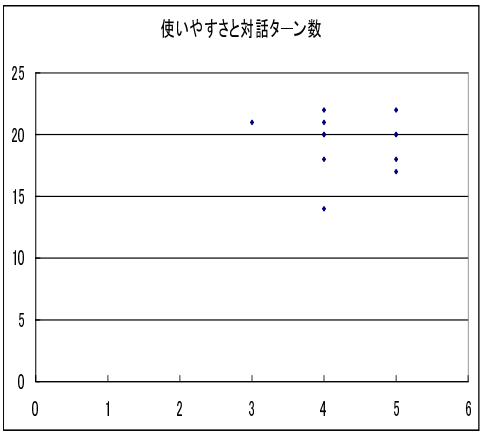
対話の円滑性と対話時間の相関係数 (対話戦略毎)

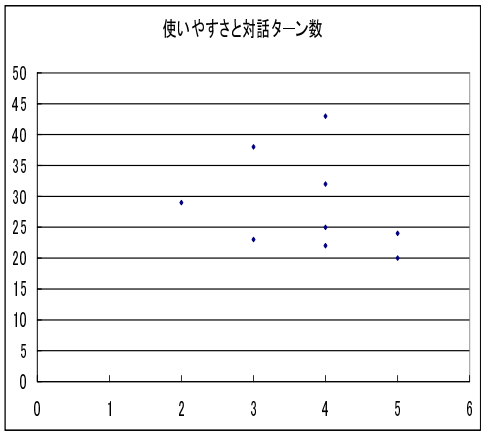
対話戦略	相関係数	対話戦略	相関係数
1A	-0.193	1B	0.162
2A	-0.285	2B	0.211
3A	-0.553	3B	-0.37
4A	-0.39	4B	-0.23
5A	0.283	5B	-0.48
6A	-0.659	6B	-0.641
7A	-0.425	7B	-0.006
8A	-0.058	8B	0.257
9A	-0.289	9B	-0.233
10A	-0.203	10B	-0.19
11A	-0.535	11B	0.139
12A	0.165	12B	-0.849

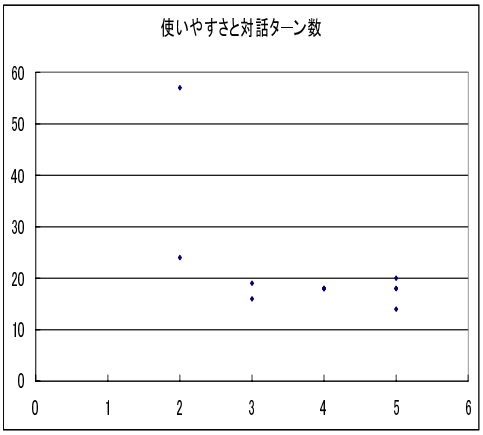
表 .2: 対話の円滑性と対話時間の相関係数

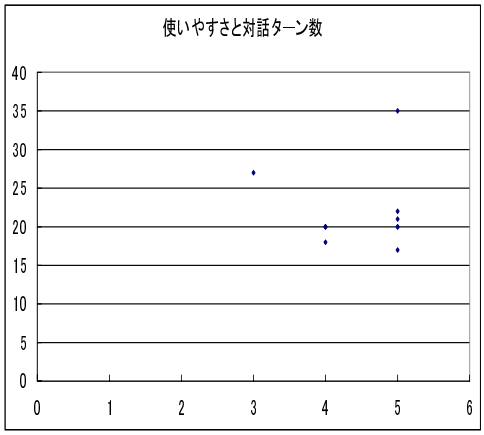
AppendixD

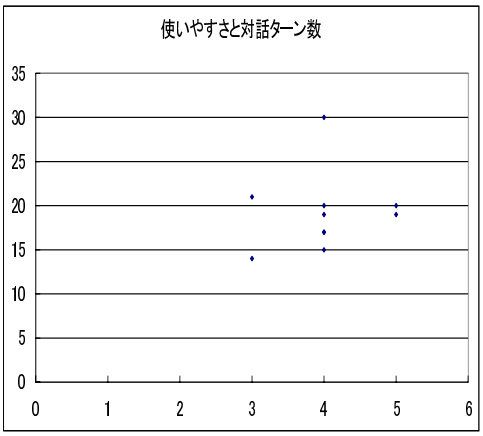
使いやすさと対話ターンの散布図 1(対話戦略毎)

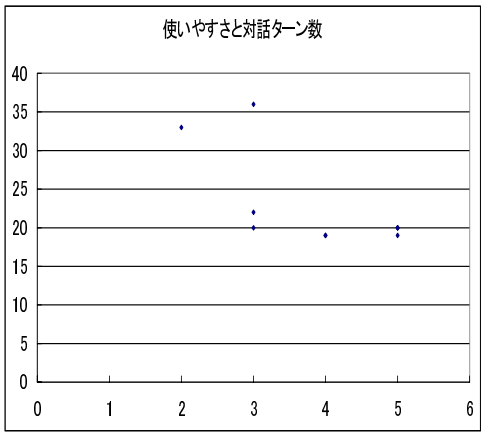


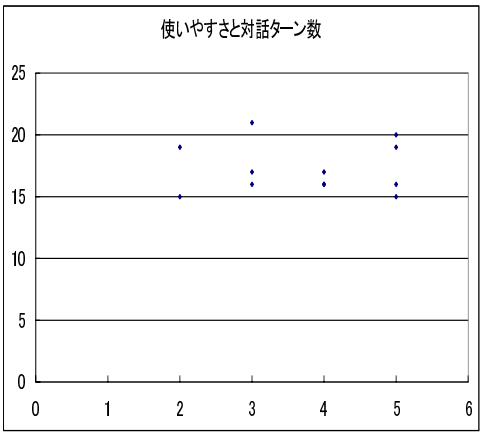


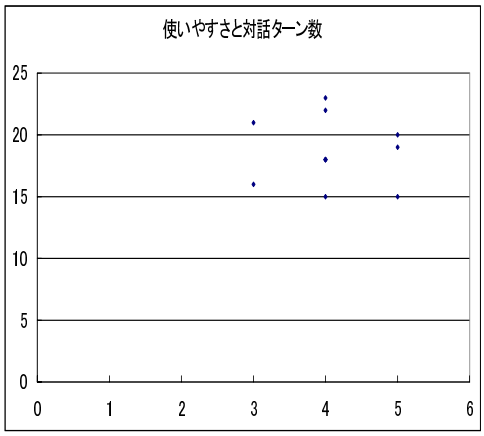





使いやすさと対話ターン数
35
30
25
20
15
10
5
0
0 1 2 3 4 5 6


使いやすさと対話ターン数
40
35
30
25
20
15
10
5
0
0 1 2 3 4 5 6


使いやすさと対話ターン数
25
20
15
10
5
0
0 1 2 3 4 5 6


使いやすさと対話ターン数
25
20
15
10
5
0
0 1 2 3 4 5 6

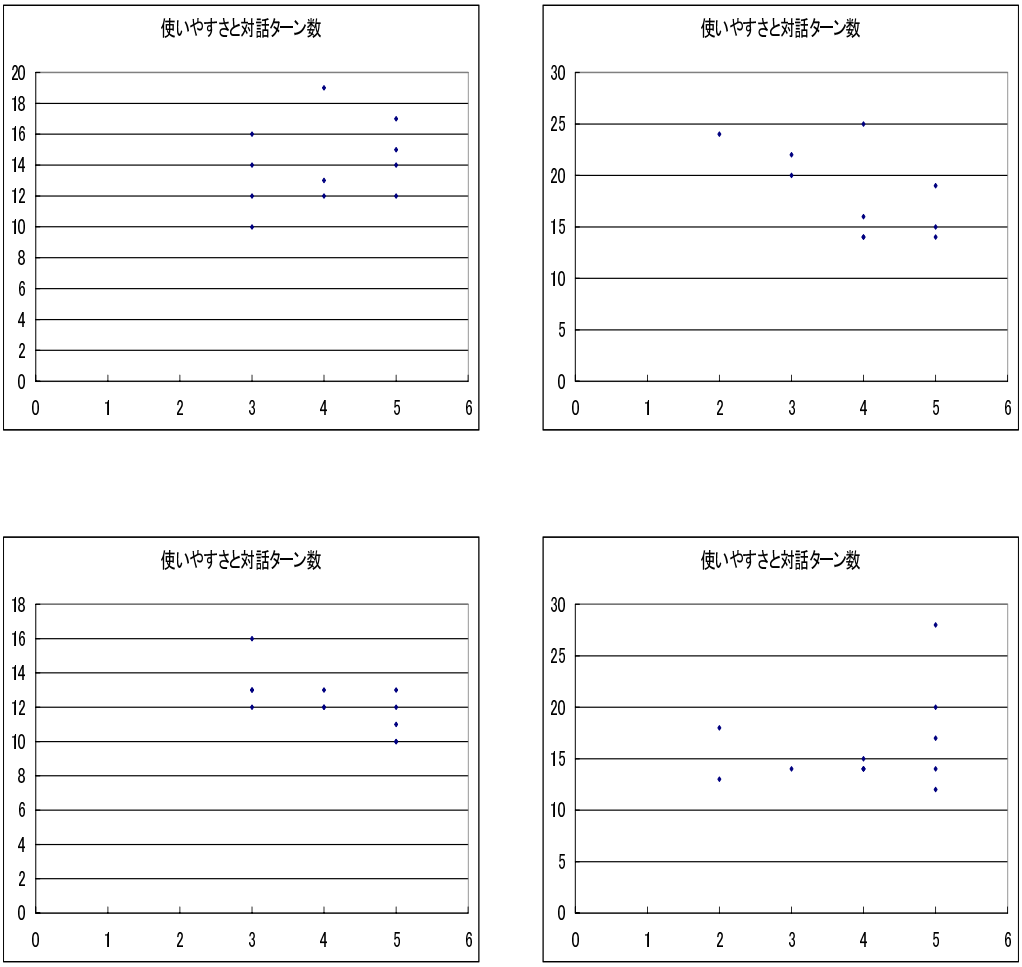
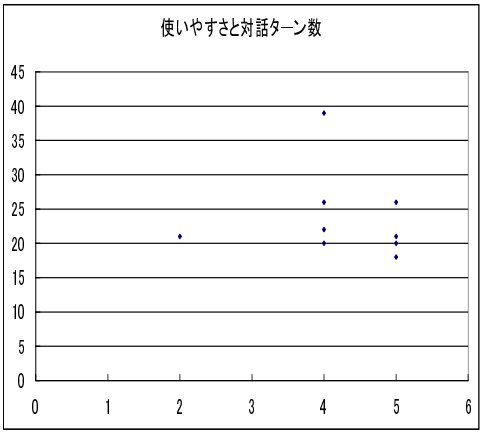
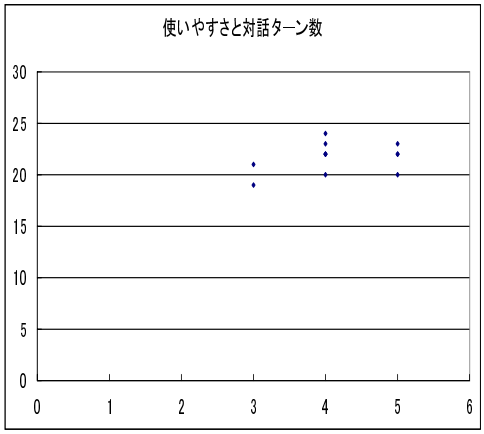
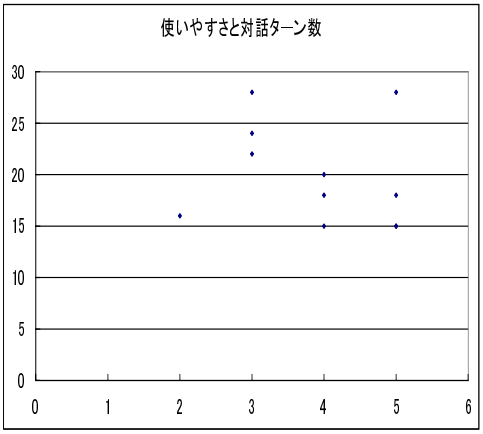


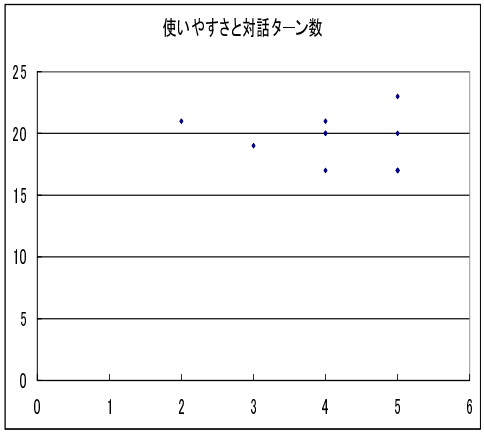
図 .5: 使いやすさと対話ターンの散布図 (1A〜12A)

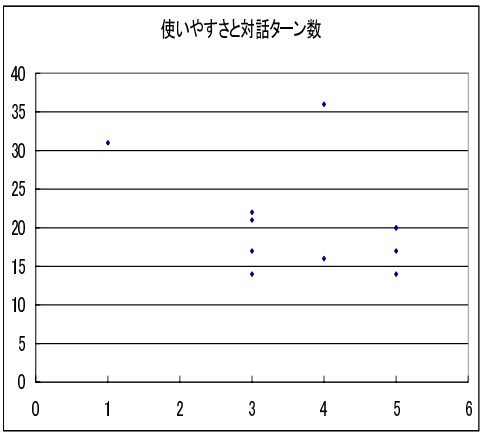
使いやすさと対話ターンの散布図 2(対話戦略毎)

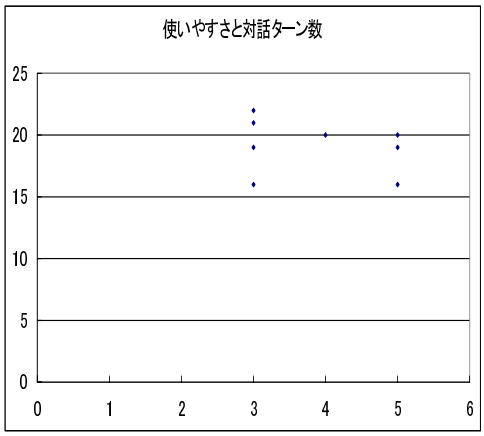


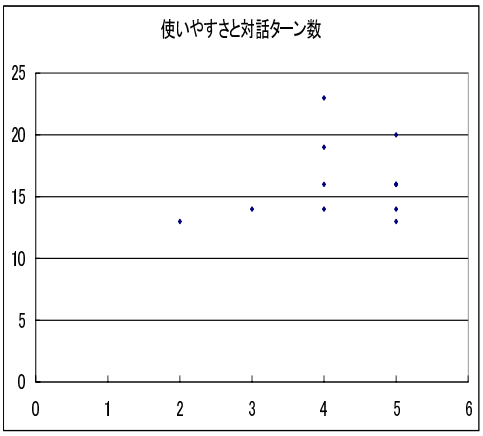


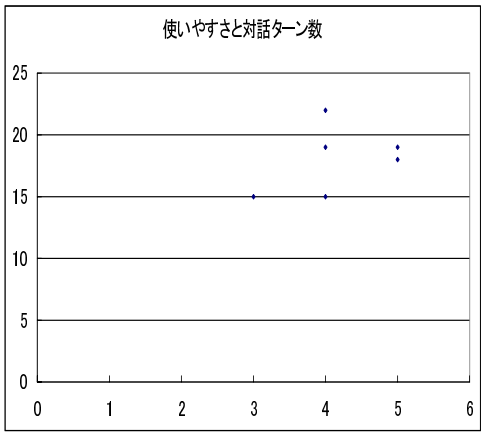





使いやすさと対話ターン数
40
35
30
25
20
15
10
5
0
0
1
2
3
4
5
6


使いやすさと対話ターン数
25
20
15
10
5
0
0
1
2
3
4
5
6


使いやすさと対話ターン数
25
20
15
10
5
0
0
1
2
3
4
5
6


使いやすさと対話ターン数
25
20
15
10
5
0
0
1
2
3
4
5
6

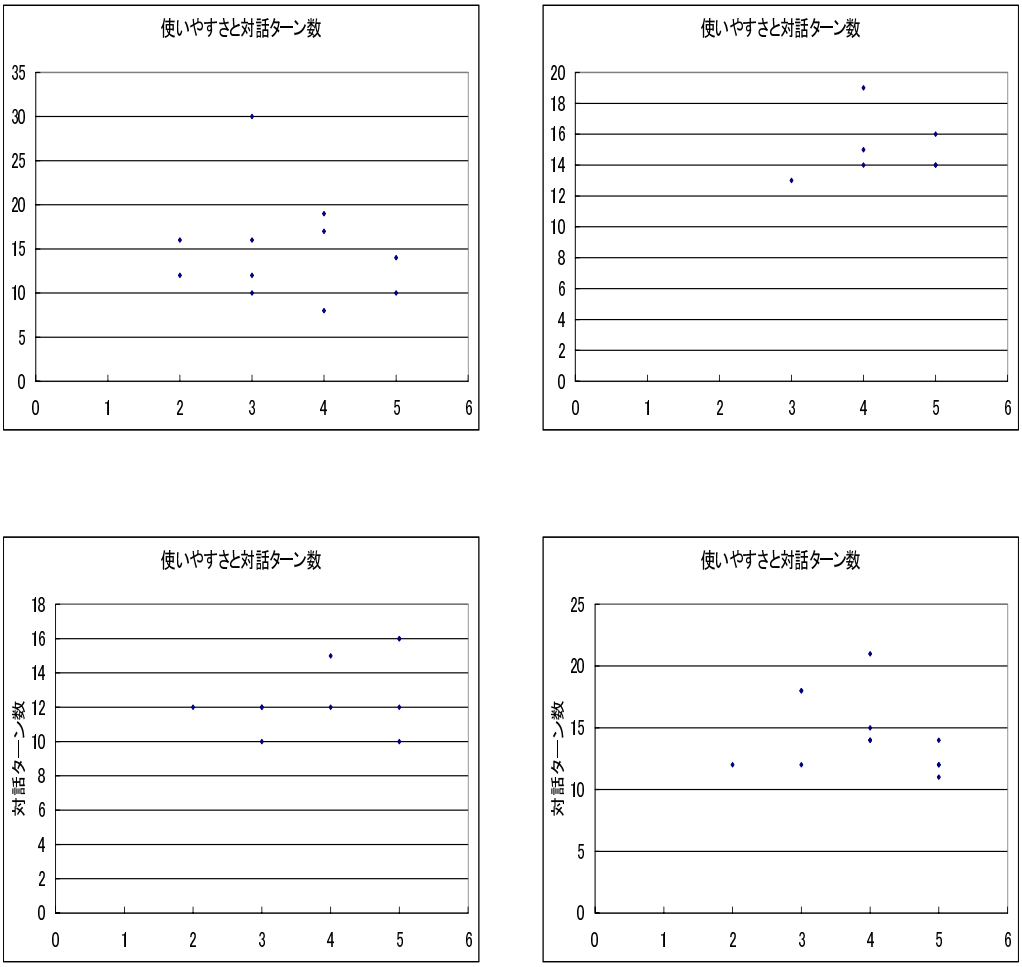


図 .6: 使いやすさと対話ターンの散布図 (1B〜12B)

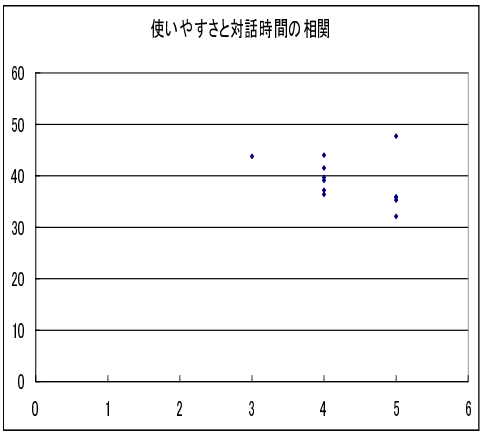
使いやすさと対話ターンの相関係数 (対話戦略毎)

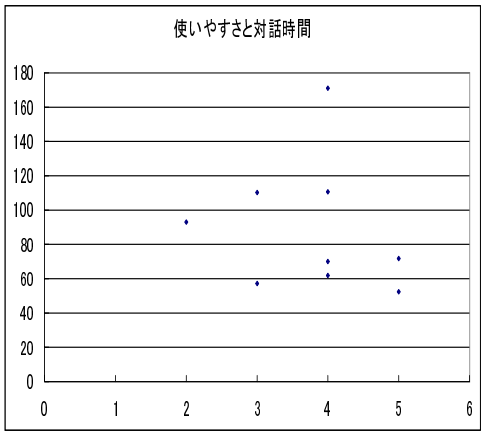
対話戦略	相関係数	対話戦略	相関係数
1A	-0.10	1B	-0.087
2A	-0.297	2B	0.323
3A	-0.599	3B	-0.194
4A	-0.031	4B	-0.275
5A	0.150	5B	-0.382
6A	-0.679	6B	-0.257
7A	-0.009	7B	0.199
8A	-0.083	8B	0.396
9A	0.257	9B	-0.168
10A	-0.616	10B	0.11
11A	-0.645	11B	0.396
12A	0.276	12B	-0.253

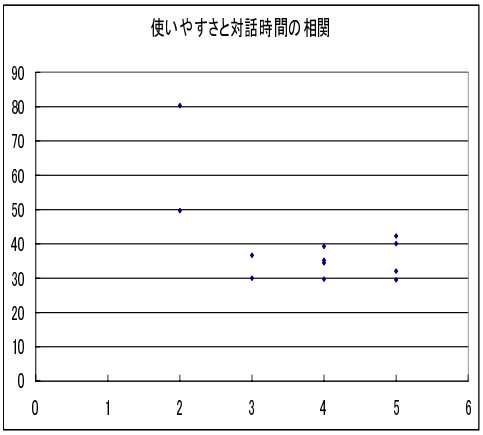
表 .3: 使いやすさと対話ターンの相関係数

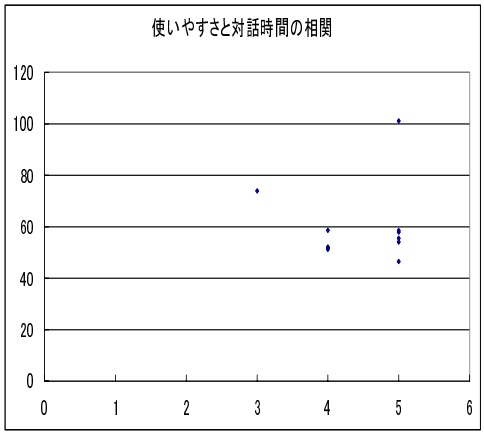
AppendixE

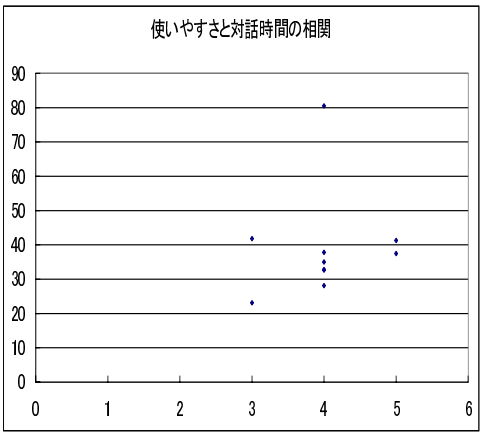
使いやすさと対話時間の散布図 1(対話戦略毎)

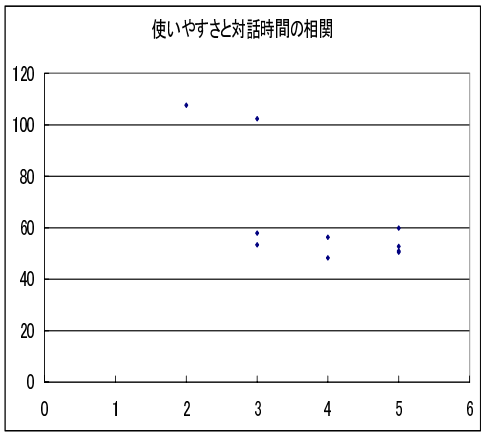


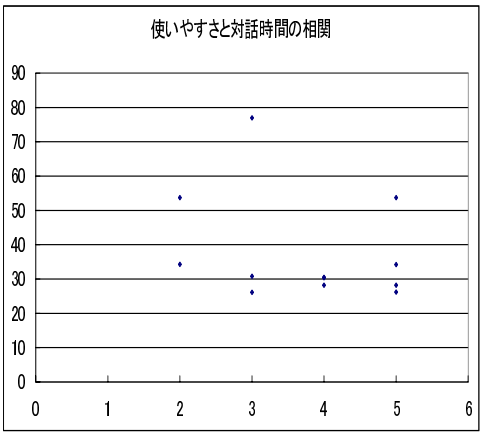


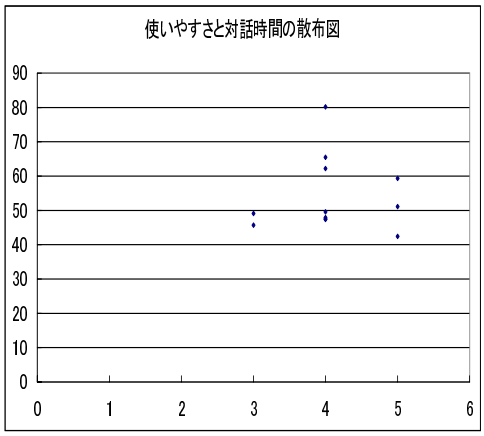





使いやすさと対話時間の相関
90
80
70
60
50
40
30
20
10
0
0
1
2
3
4
5
6


使いやすさと対話時間の相関
120
100
80
60
40
20
0
0
1
2
3
4
5
6


使いやすさと対話時間の相関
90
80
70
60
50
40
30
20
10
0
0
1
2
3
4
5
6


使いやすさと対話時間の散布図
90
80
70
60
50
40
30
20
10
0
0
1
2
3
4
5
6

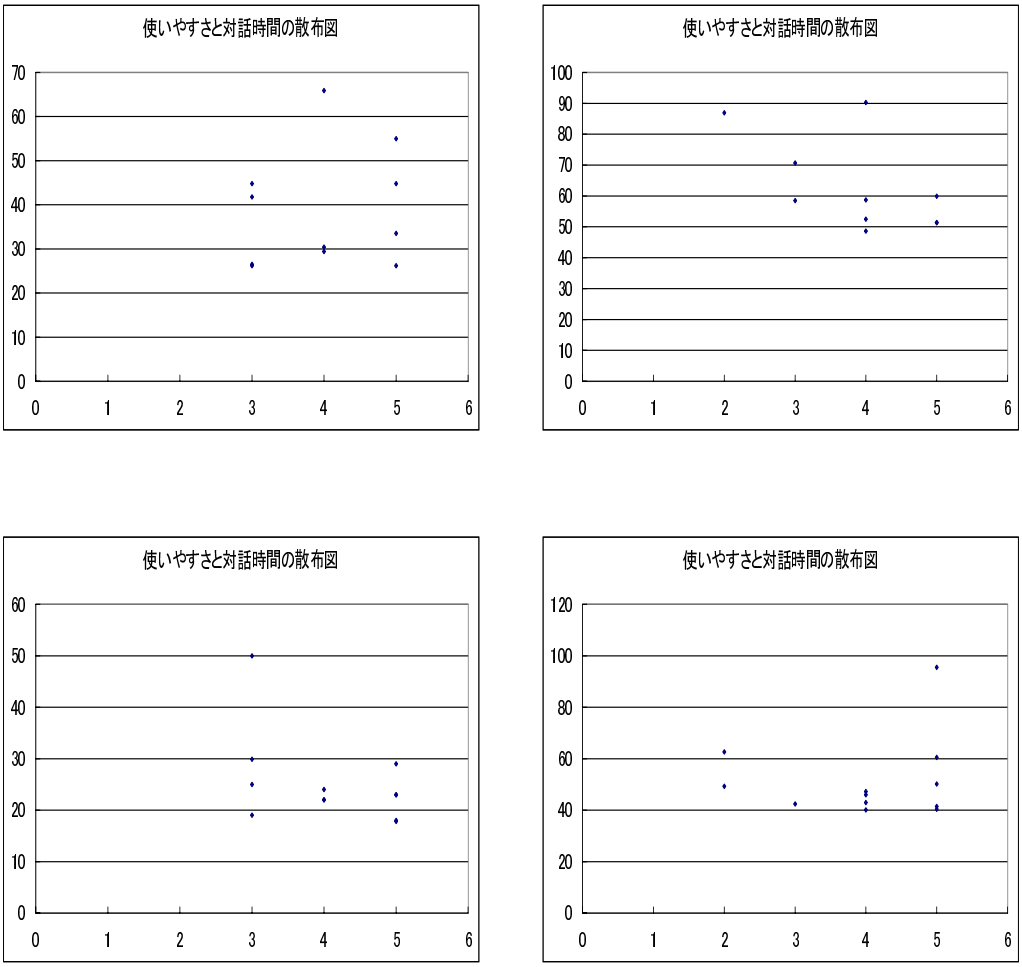
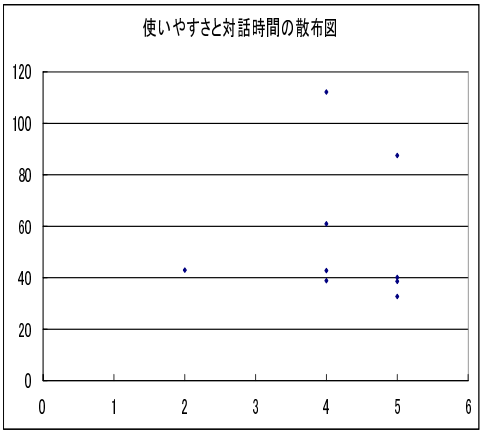
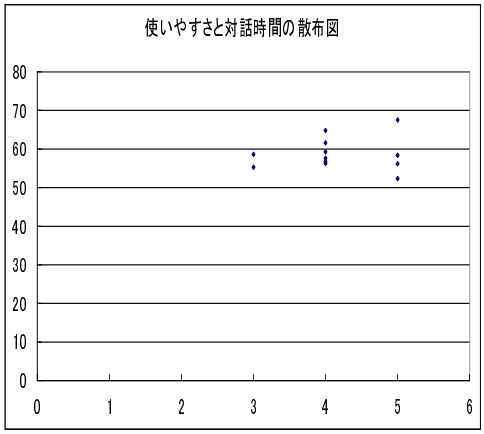
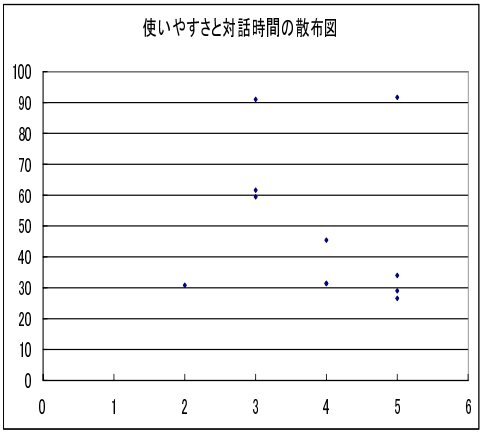


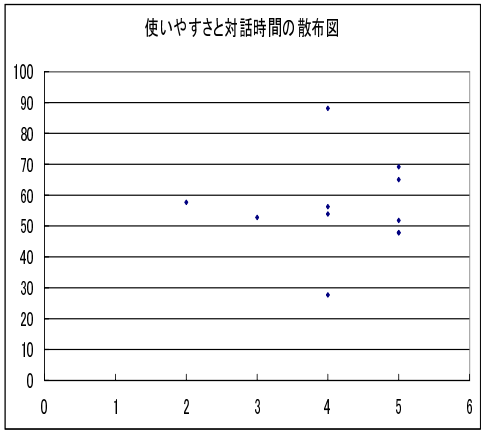
図 .7: 使いやすさと対話時間の散布図 (1A〜12A)

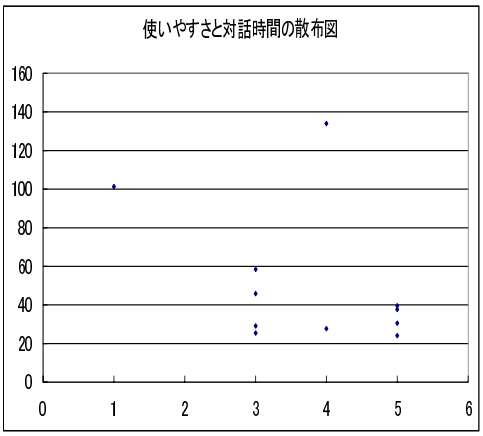
使いやすさと対話時間の散布図 2(対話戦略毎)

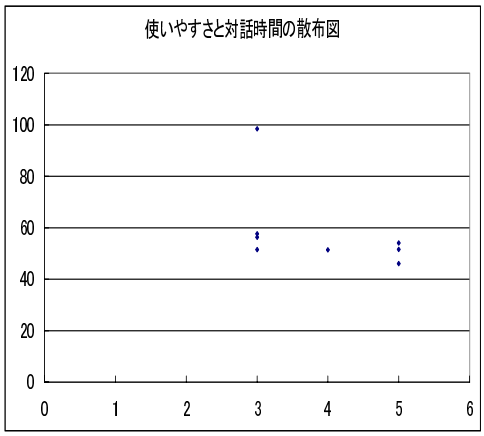


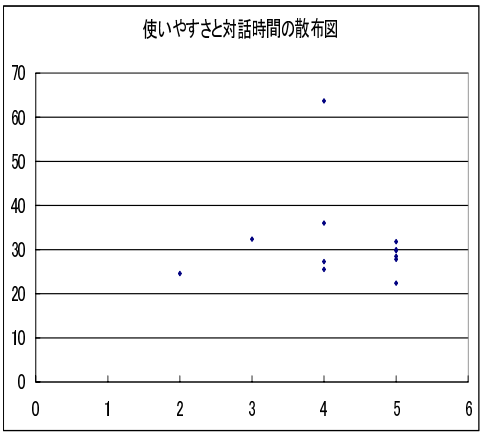


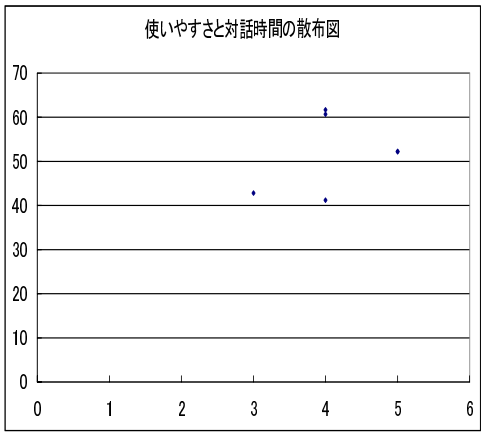





使いやすさと対話時間の散布図
160
140
120
100
80
60
40
20
0
0 1 2 3 4 5 6


使いやすさと対話時間の散布図
120
100
80
60
40
20
0
0 1 2 3 4 5 6


使いやすさと対話時間の散布図
70
60
50
40
30
20
10
0
0 1 2 3 4 5 6


使いやすさと対話時間の散布図
70
60
50
40
30
20
10
0
0 1 2 3 4 5 6

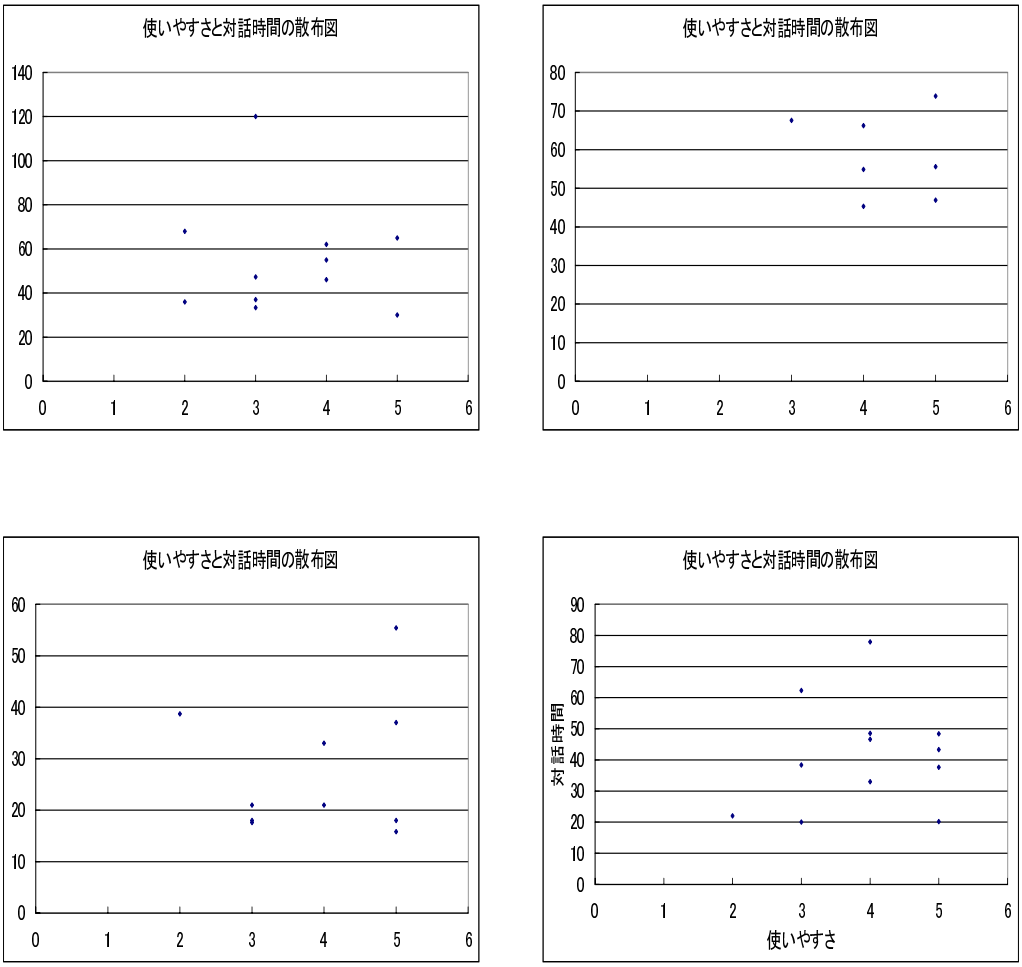


図 .8: 使いやすさと対話時間の散布図 (1B〜12B)

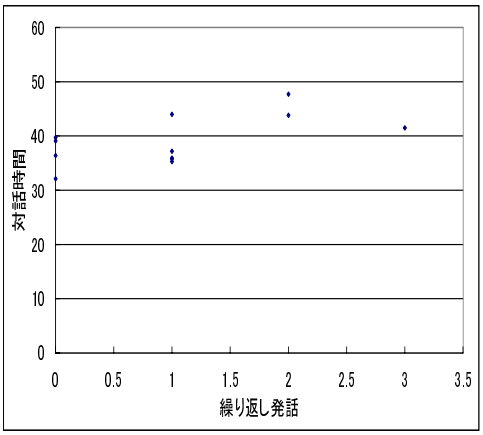
使いやすさと対話時間の相関係数 (対話戦略毎)

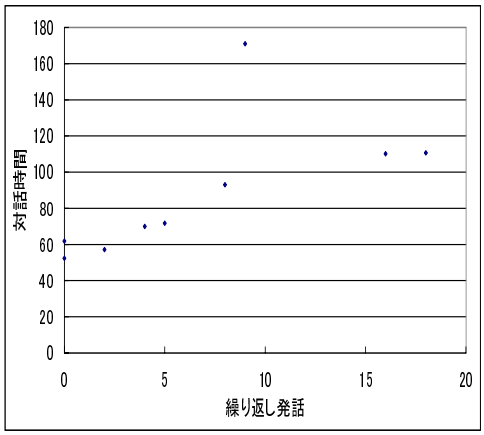
対話戦略	相関係数	対話戦略	相関係数
1A	-0.40	1B	0.013
2A	-0.176	2B	0.092
3A	-0.597	3B	-0.172
4A	-0.008	4B	0.007
5A	0.148	5B	-0.398
6A	-0.694	6B	-0.467
7A	-0.277	7B	-0.054
8A	0.050	8B	0.304
9A	0.171	9B	-0.089
10A	-0.578	10B	-0.171
11A	-0.445	11B	0.153
12A	0.111	12B	0.116

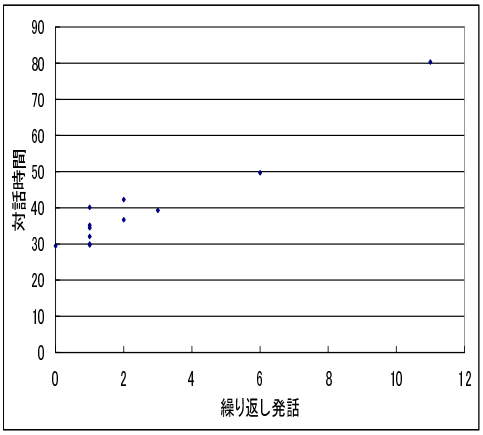
表 .4: 使いやすさと対話時間の相関係数

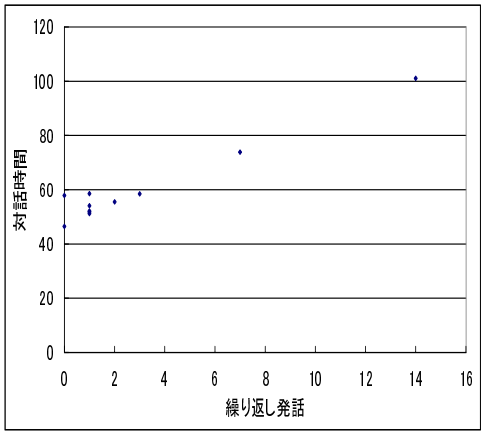
AppendixF

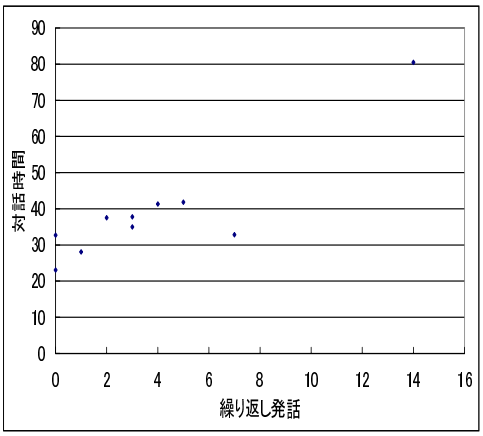
繰り返し発話と対話時間の散布図 1(対話戦略毎)

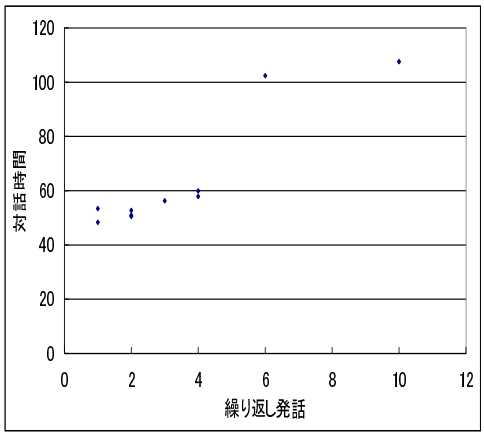


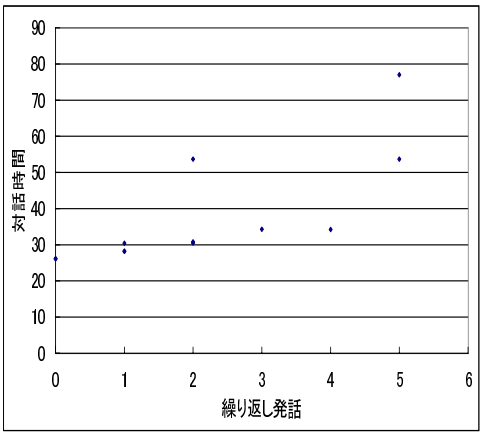


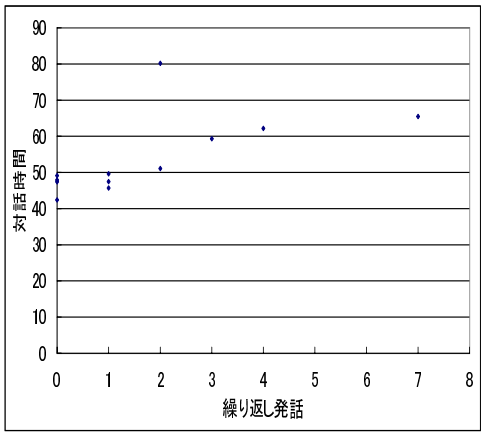





対話時間
繰り返し発話


対話時間
繰り返し発話


対話時間
繰り返し発話


対話時間
繰り返し発話

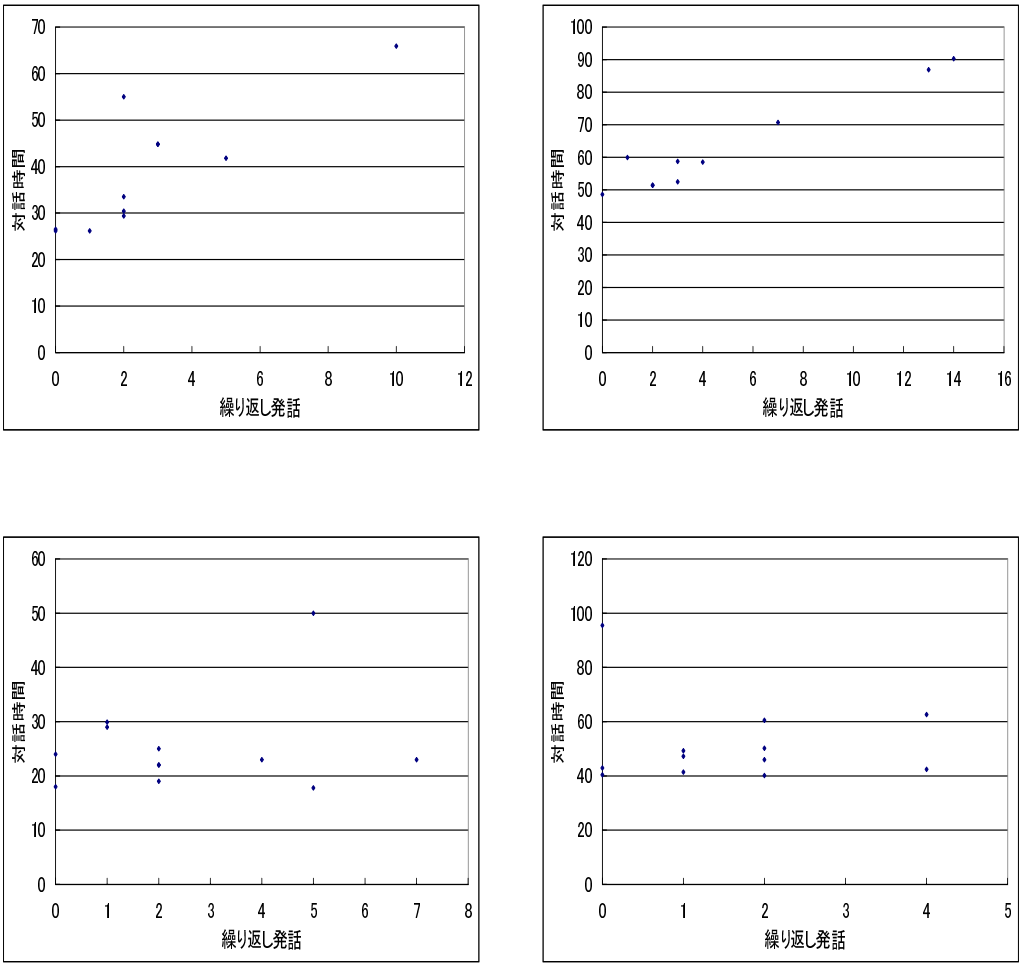
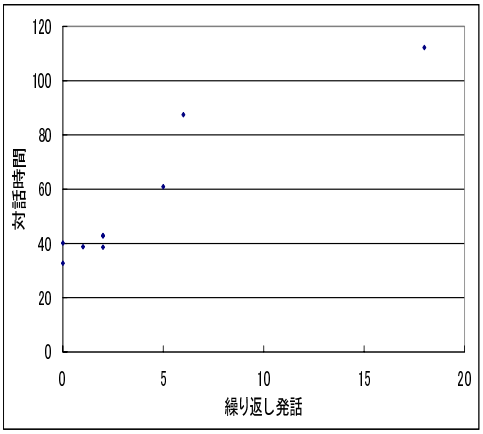
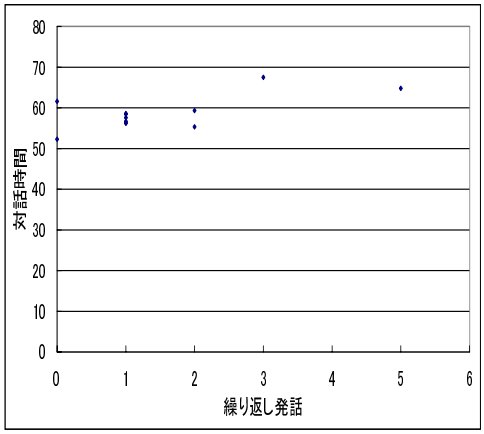
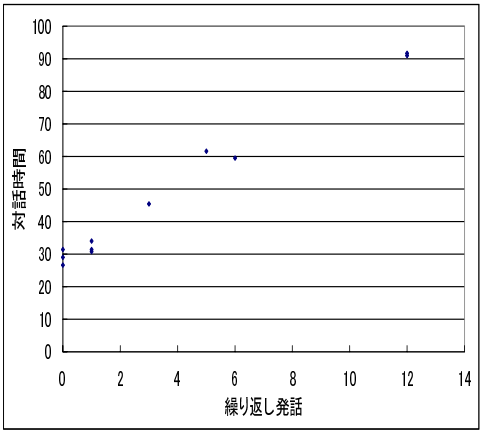


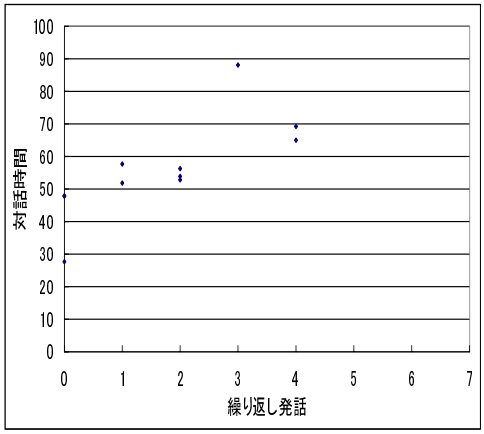
図 .9: 繰り返し発話と対話時間の散布図 (1A〜12A)

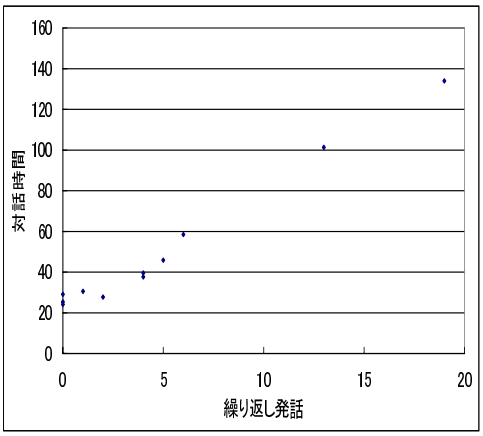
繰り返し発話と対話時間の散布図 2(対話戦略毎)

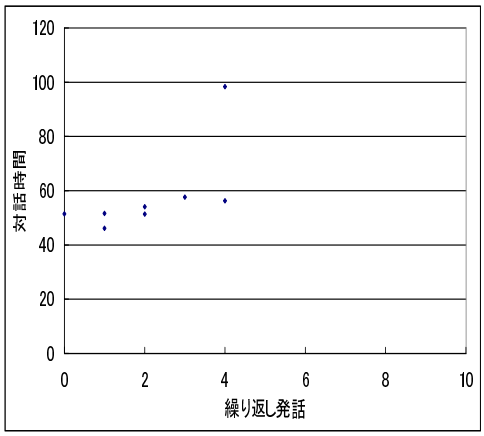


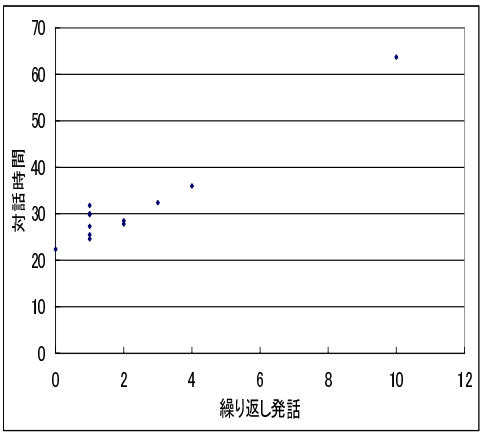


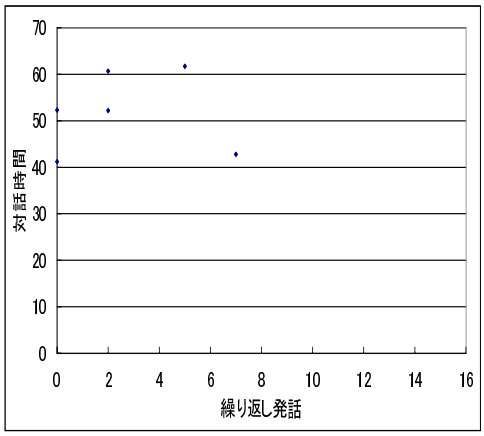





160
140
120
100
80
60
40
20
0
対話時間
0
5
10
15
20
繰り返し発話


120
100
80
60
40
20
0
対話時間
0
2
4
6
8
10
繰り返し発話


70
60
50
40
30
20
10
0
対話時間
0
2
4
6
8
10
12
繰り返し発話


70
60
50
40
30
20
10
0
対話時間
0
2
4
6
8
10
12
14
16
繰り返し発話

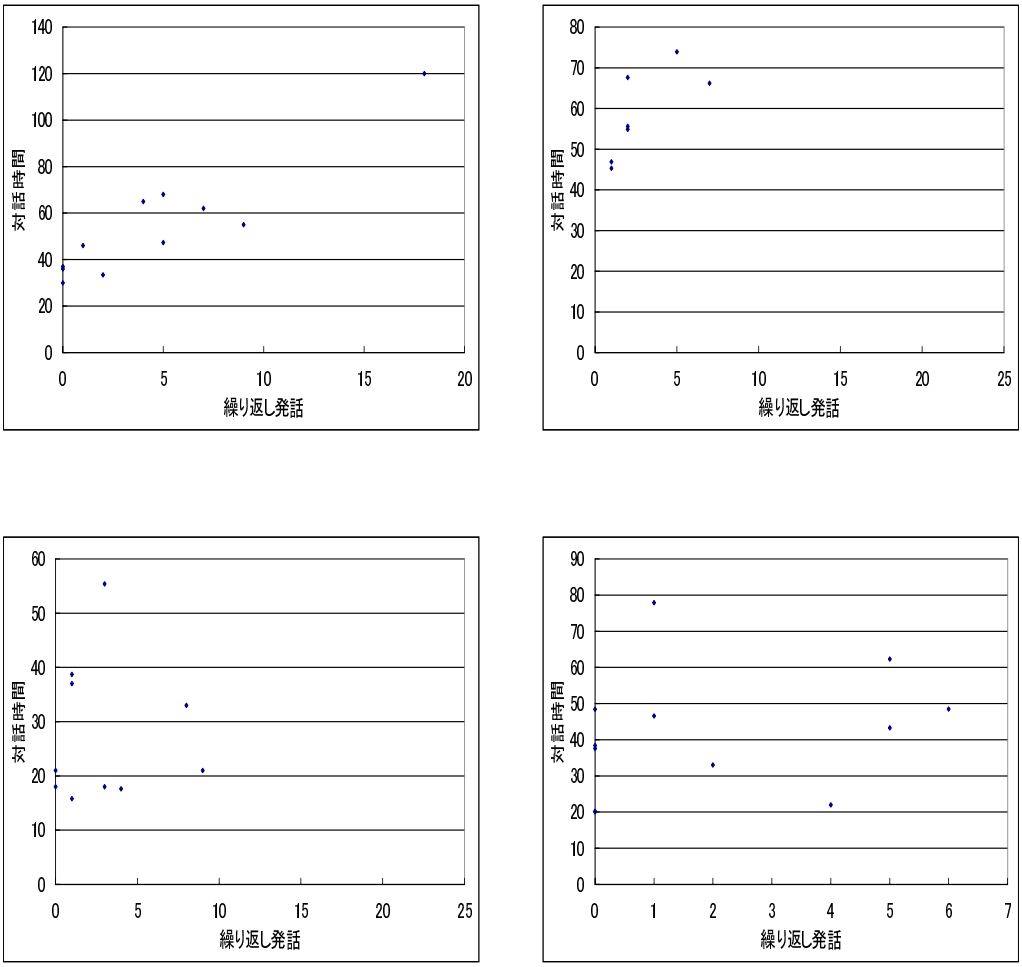


図 .10: 繰り返し発話と対話時間の散布図 (1B〜12B)

繰り返し発話と対話時間の相関係数 (対話戦略毎)

対話戦略	相関係数	対話戦略	相関係数
1A	0.577	1B	0.939
2A	0.654	2B	0.688
3A	0.965	3B	0.993
4A	0.973	4B	0.754
5A	0.899	5B	0.990
6A	0.919	6B	0.639
7A	0.758	7B	0.968
8A	0.627	8B	0.04
9A	0.820	9B	0.921
10A	0.969	10B	0.734
11A	0.214	11B	0.02
12A	-0.116	12B	0.028

表 .5: 繰り返し発話と対話時間の相関係数

AppendixG

使いやすさの重回帰式 (対話戦略毎)

重回帰分析に用いた目的変数 Y は，使いやすさ (5 段階評価アンケート値) 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．以下の重回帰式は，すべて標準偏回帰係数で表わされている．

-	標準偏回帰係数			
対話戦略	対話の円滑性 (x_1^*)	システム表現 (x_2^*)	対話ターン (x_3^*)	対話時間 (x_4^*)
1A	0.468	0.528	-0.408	-0.049
2A	0.362	0.592	0.035	0.005
3A	0.245	0.741	-0.002	-0.15
4A	0.441	0.414	-0.942	1.017
5A	0.043	0.657	0.702	-0.445
6A	0.451	0.641	-0.058	0.482
7A	0.284	0.959	0.056	0.517
8A	0.607	0.326	0.0824	0.066
9A	0.283	0.678	-0.04	0.088
10A	0.294	0.797	0.639	-0.792
11A	0.172	0.441	-0.509	0.124
12A	1.613	0.167	-1.87	1.57
1B	0.37	0.626	0.069	-0.049
2B	0.125	0.316	0.416	-0.192
3B	0.35	0.607	-0.042	0.288
4B	0.482	0.48	-0.154	0.279
5B	0.533	0.52	-0.22	0.27
6B	0.12	1.01	-0.1	-0.27
7B	0.45	0.446	0.52	-0.42
8B	0.84	0.438	-2.22	-1.99
9B	0.68	0.25	0.48	-0.32
10B	0.3	0.47	0.47	-0.23
11B	0.74	-0.26	0.77	-0.28
12B	0.66	0.61	0.58	0.14

AppendixH

システム対話戦略

対話戦略	1	2	3	4	5	6	7	8	9	10	11	12
直直直	○											
直直直+まとめ		○										
直直相			○									
直直相+まとめ				○								
直相相					○							
直相相+まとめ						○						
相相相							○					
相相相+まとめ								○				
間間									○			
間間 +まとめ										○		
無無無											○	
無無無+まとめ												○

- 直は直接確認，相は相槌，間は間接確認，無は無応答をそれぞれ表す.
- 2 種類の異なるシステムの表現により，それぞれ対話戦略 A と B とし，合計 24 パターンある.
- 表現 A は,「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)」の組み合わせ.
- 表現 B は,「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)」の組み合わせ.