

## 誤認識に対処した自然で効率的な音声対話戦略の評価

杉本 夏樹<sup>†</sup> 矢野 浩利<sup>†</sup> 北岡 教英<sup>††</sup> 中川 聖一<sup>†</sup>

<sup>†</sup>豊橋技術科学大学 情報工学系

〒441-8580 愛知県 豊橋市 天伯町 雲雀ヶ丘 1-1

E-mail: †{sugimoto,hyano,nakagawa}@slp.ics.tut.ac.jp

<sup>††</sup>名古屋大学 大学院情報科学研究科

〒464-8603 愛知県 名古屋市 千種区 不老町 1

E-mail: ††kitaoka@nagoya-u.jp

あらまし 音声対話におけるシステムの誤認識を、対話を通じて自然かつ効率的に暗黙に回復する対話戦略を提案する。自然かつ効率的に誤認識を回復するために、確認発話を多様しない戦略を目指す対話戦略として、各対話ターンにおいて複数の理解候補を保持し、対話を通して最適な仮説を探索する。その際、理解候補を収束させる効率性の尺度と、理解候補と矛盾しない無矛盾性の尺度をシステム応答の選択に用いる。まず、計算機シミュレーションと対話聴取によって評価を行い、自然で効率的な戦略であることを示した。次に、これらの手法を用いた音声対話システムを構築し、実際に被験者が使用して客観的・主観的に評価を行った。

キーワード 音声対話システム, 誤認識からの回復, 効率的な応答, 自然な応答

## Evaluation of a Spoken Dialog Strategy for Natural and Efficient Recovery from Misrecognition

Natsuki Sugimoto<sup>†</sup> Hirotooshi Yano<sup>†</sup> Norihide Kitaoka<sup>††</sup> Seiichi Nakagawa<sup>†</sup>

<sup>†</sup> Spoken Language Processing Laboratory, Toyohashi University of Technology

1-1 Hibarigaoka, Tenpaku-cho, Toyohashi City, Aichi Prefecture, 441-8580 Japan

E-mail: †{sugimoto,hyano,nakagawa}@slp.ics.tut.ac.jp

<sup>††</sup> Speech, Text and Human behavioral Signal Processing Laboratory, Nagoya University

1 Furo-cho, Chikusa-ku, Nagoya City, Aichi Prefecture, 464-8603 Japan

E-mail: ††kitaoka@nagoya-u.jp

**Abstract** We proposed a novel dialog strategy which can recover implicitly from misrecognition through a spoken dialog. To recover the misrecognition with naturally and efficiently, we aim at the strategy without confirmation. Our strategy kept multiple understanding hypotheses at each turn and searched a globally optimal hypothesis across user's utterances in a whole dialog. As for a dialog strategy, we introduced the new criteria based on 'efficiency for convergence' and 'consistency with understanding hypotheses' to select an appropriate system response. We developed a spoken dialog system using these techniques, and showed that our strategy is efficient and natural in simulation dialog and real dialog experiments.

**Keywords** spoken dialog system, recovery from misrecognition, efficient response, natural response

### 1 はじめに

近年、音声対話システムの研究が非常に盛んになっている。こと実用的な音声対話システムを設計することにおいて、音声認識器やテキスト音声合成システムのような音声言語処理システムの単なる連結は本質とは云えず、そこには実用的な対話管理戦略の設計が必要とされる。その戦略の指標として、システムの効率性やタスク実行の自然性、インシアチブ管理などがある。これらの要素の評価はユーザの主観で大きく変動しタスクやシナリオに大きく依存するなど、対話戦略の設計を難しいものとしている。このため、汎用的で実用的な戦略は常に求められている。

コンピュータと人間が音声を通じてコミュニケーションをはかる場合、誤認識は避けて通れない。また、現在はインターフェースが未熟であるために、誤認識からの回復が困難である。現在の商用音声カーナビの対話戦略を例に取ると、誤認識したまま対話を進めるのを防ぐために、1項目の入力ごとに確認発話が行われる。しかし、ユーザの発

話を受け取る度に確認発話を行うとタスク達成までのターン数が増え、ユーザにとっても煩わしいものである。そもそもカーナビ操作を音声で行う利点は、アイズフリーやハンズフリーによって運転に集中させるためであることを考えるとこれは矛盾であり、ユーザの運転への集中力を妨げる戦略は有害と云える。

本稿では、先行研究 [1][2] である確認発話を多用しない対話手法を、自然性や効率性から評価を行い、実用的な対話管理戦略を模索する。一般に確認発話を行わないことで、ユーザの発話を誤認識したまま対話を進めてしまう恐れがあるが、発話の理解候補を複数保持させることで、曖昧性を残して対話を進めていき、理解候補から正しい意図を導くようなシステム応答を行う。そしてユーザから効果的な情報を得ることで、システムが対話を進めていくうちに誤認識からの自然な回復が成される戦略を目指すと共に、計算機シミュレーションや被験者実験を通して主観的な尺度(満足度など)と客観的な尺度(ターン数など)から戦略を評価する。

関連研究はこれまでも多く行われており、駒谷らは音声認識結果の信頼度を用いて必要な場合にのみ効率よく確認発話を行う方法 [3] を提案しており、伊藤や東中らは複数理解候補を順序付けて保持することで、単純に音声認識結果の第一候補を最優先するよりも複数の対話状態を保持することが優位であることを示している [4,5]。翠らは、音声認識誤りによる曖昧性を、 $N$ -best 結果の重なり具合からリスクを減らすことで効率的な確認発話を実現している [6]。堂坂らは、確認コストと伝達コストからなるデュアルコスト法を用いて、効率的な対話制御を実現した [7]。また、評価の試みとして Walker は主観的な基準の評価分析のために客観的な尺度の線形結合で近似できるフレームワーク [8] を提唱し、渡辺らは、計算機シミュレーションの方法論 [9] を示すことで対話戦略の評価を行っている。Yang らは、実際の人間間対話を模した協力努力を最小にするイニシアチブ戦略をシミュレーションで評価 [10] し、Oliver は、実際の音声対話シミュレーションの為のユーザモデルに背景知識共有プロセスを付与してシミュレートするためのモデルを提案した [11]。

本稿では、複数理解を保持した状態で効果的に理解を収束させつつ、対話履歴との矛盾を少なく誤認識からの回復ができる応答生成法を提案し、評価を行う。

## 2 複数理解候補を保持する 音声対話理解方法 [1][2]

一般に、音声対話中ではシステムによる誤認識が生じる。誤認識から回復できる対話戦略として、確認発話がある。確認発話を行う対話は図 1 のようなものがある。例中で確認発話を行わなければ、対話ターン数は大幅な削減が期待できる。しかし、「豊橋市」を「豊川市」と誤認識したまま対話が進行し、対話中にこれまでの理解と現在のユーザ発話において矛盾が起こり対話が破綻したり、最終的に誤った理解結果に至る。

System	市区町村名を入力してください
User	豊橋市に行きます
System	豊川市でよろしいですか？
User	いいえ、豊橋
System	豊田市でよろしいですか？
User	豊橋市だよ！
System	豊橋市でよろしいですか？
User	はい
System	町名を入力してください

図 1: 毎回確認発話を行う対話例

これは、音声認識の結果の第一候補のみを信じて対話を進めることに起因する。各認識において、複数得られる認識候補を有効に用いれば、この危険は低減される。

そこで図 2 のように、認識の複数候補に基づいたユーザ発話の理解候補を複数持ったままシステムは対話を行う。各行が 1 つの理解候補に対応し、それぞれの候補には認識結果より信頼度の値を持たせる。信頼度の値は、通常の連続単語認識結果における対象単語区間の対数尤度と、別に動作させる音節列認識結果の同区間の対数尤度の差 (対数事後確率) をとったものを用いる。また、図中で理解候補中の ( ) で囲まれた単語は、ユーザは発話していないがシステムの知識より候補が予想できている状態を示す。この例の場合、「静岡県清水区」という認識結果に基づいた理解では、県名は自動的に決まっていることを示す。

## User 発話 豊橋市のコンビニに行きたいです

システムの理解候補				
ランク	県名	市名	町名	目標物
1	(愛知県)	豊橋市	—	コンビニ
2	富山県	—	—	コンビニ
3	(愛知県)	豊川市	—	コンビニ

図 2: 1 つのユーザ発話からの複数理解候補生成の例

ユーザ発話の度に、ユーザ発話の複数認識候補と現在の各理解候補を組み合わせることで最新のシステムの理解を構成する。そのため、理解候補は対話を繰り返すことで増加していく。

複数理解候補を持ったまま対話を進めていくと、理解候補中には新たに得られるユーザ発話を考慮すると、理解に一貫性がなく矛盾している候補が出現することがある。ここでいう矛盾とは、このような理解候補については、これまでの対話理解が誤っているか、最新のユーザ発話の認識が誤っていると考えられる。そのため、このケースに該当する理解候補については信頼度を下げることとする。すると、対話を進めていくことで過去のユーザ発話の認識結果を統合して矛盾のない理解候補は信頼度が高くなり、過去の認識結果において正解が最尤でなくても、後に誤りからの回復が可能になると考えられる。

## 3 システム応答の決定手法 [1][2]

### 3.1 応答が理解候補を絞り込む度合いの尺度

現在の理解候補集合  $N$  において、システムが質問  $q$  を行った場合にユーザから回答  $a$  が得られる確率を以下のように求める。

$$P(a|N, q) = \sum_n P(a|n, q) \cdot P(n) \quad (1)$$

ここで理解候補  $n$  が正解である確率  $P(n)$  は信頼度を考慮して計算する必要があるが、理解候補の信頼度は 2 節で述べたように連続単語認識と音節列認識の対数尤度の差を用いているため、信頼度の尺度  $Conf(n)$  を元に理解候補が正しい確率を求めるには信頼度の値と正解確率の関係について統計を取る必要がある。今回は擬似的に  $P(n) = Conf(n) / \sum_{m \in N} Conf(m)$  を用いた。そして、質問  $q$  の場合に予想できる回答  $a$  の出現確率と矛盾する数を掛け合わせる。図 2 の例について考えると、ここで「県名を教えてください」という質問を想定した場合、「愛知県です」と回答が得られればランク 2 に富山県があるので矛盾 1 個となり、「富山県です」と回答が得られればランク 1, 3 に愛知県があるので矛盾する数が 2 個となる。そして、 $n$  において  $a$  が矛盾することを  $I(a, n) = 1$ 、矛盾しないことを  $I(a, n) = 0$  として、候補が絞り込めるスコア (効率性尺度; Efficiency score) を以下のように定義する。

$$S_e(q) = \sum_n \{1 - P(n)\} \sum_a I(a, n) \cdot P(a|N, q) \quad (2)$$

### 3.2 応答が理解候補に共通する度合いの尺度

3.1 節の尺度で選んだ応答は、誤認識した結果から得られた場合には、ユーザにとっては対話上不自然で、また、

システムとしては自らの誤認識をユーザに伝えてしまうことになる。ここで、現時点の複数候補のいずれかで正しい理解を表現されていると仮定すると、多くの理解候補と矛盾しない応答が、ユーザにとって一貫性があり自然で、かつ誤解を悟られない応答であると考えられる。図2の例について考えると、ここでシステムが「市区町村名を教えてください」という応答を選択すると、理解候補の1位と3位ではすでにユーザが発話した内容について2度尋ねることに相当してしまい矛盾している。また、「愛知県でよろしいですか?」とシステムが応答すると(実際には誤った理解であるが)2位の理解候補にとっては富山県と理解しているため矛盾している。このケースのようにシステムが行う応答が、各理解候補に対してなるべく矛盾しているものは避けたい。そこで、 $n$  において  $q$  が矛盾することを  $I(q,n) = 1$ 、矛盾しないことを  $I(q,n) = 0$ 、理解候補  $n$  の信頼度を  $Conf(n)$  として、理解候補に共通である尺度 (=無矛盾性尺度; Consistency score) を以下のように定義して用いる。

$$S_c(q) = \sum_{n \in N} \{1 - I(q,n)\} \cdot Conf(n) \quad (3)$$

### 3.3 最終的なシステム応答の選択

上で定義した尺度において、2つの値の重みつき和が最大となる質問  $Q$  をシステムの応答として選択する。

$$Question = \operatorname{argmax} w_n \cdot S_e(Q) + w_c \cdot S_c(Q) \quad (4)$$

本稿では、以下の入力促しと確認発話の質問に対してこのスコアを計算し、最もスコアの高かったものをシステムの応答とした。

- 「県名を教えてください」
- 「市区町村名を教えてください」
- 「町名を教えてください」
- 「目標物を教えてください」
- 「県名は〇〇でよろしいですか?」
- 「市区町村は△△でよろしいですか?」
- 「町名は□□でよろしいですか?」
- 「目標物は××でよろしいですか?」

上の4つは新たな情報を要求するものであり、複数の理解候補に共通に使える傾向がある。一方、下の4つは確認発話であり「愛知県でよろしいですか?」という応答である。候補を1つに絞ることを目的とするもので、複数候補において共通する割合は小さい。また、この8つとは別にタスク完了のため最終的に確認する発話を行う。最終的な確認は、上の8つの質問のスコアが一定閾値以下だった場合、これ以上有効な応答はないと考え行う。

### 3.4 システムの評価尺度

以前の報告 [1] では、システムをターン数平均やタスク達成率といった効率性の側面からのみ評価していたが、本稿では自然性の側面からもシステムを評価するために、その尺度を提案する。

#### 3.4.1 肯定発話率

1つめの尺度として、システムの確認発話に対する全応答中に「はい」「そうです」などの肯定発話がどんな割合で含まれていたかを表す肯定発話率を評価する。これは、一般のユーザは否定の発話は心理的負担があり、嫌うであろうとの考えに基づいている。

#### 3.4.2 無矛盾率

2つめの尺度として、ある対話戦略を搭載したシステムが一对話を終了するまでにユーザの視点から対話に矛盾があったかどうかを判断する無矛盾率を評価する。ここでいう矛盾とは、ユーザが既に入力済みの項目について再度入力を促されること、及び入力済み項目に対する確認発話が発話内容と異なるなどシステムが誤認識をさらした場合の2点である。対話中に一度も上記の矛盾が存在しない場合にのみ矛盾がないとしてカウントされ、評価中に行ったタスクの回数中にどんな割合で存在したかを調べる。ただしこれらは試験者側の定義であり、被験者実験においては彼らは独自の基準に則って主観評価すると予測される。よって同時に被験者ごとの矛盾の定義を自由記述してもらうことで普遍的な矛盾の定義づけも行い、今後に活かしたい。

## 4 言い直し検出を用いた対話処理 [1][2]

一般にシステムの誤認識に気付いた際、ユーザは言い直し発話を行う [12]。言い直し発話とは、図1のユーザの1発話目と2発話目のように、過去に発話した内容と同じ内容を繰り返すことによってシステムの誤認識を伝えるものである。ユーザの言い直し発話を検出できれば、システムは自身の誤認識に気付くことができ回復が容易になる。

言い直し判定には DP マッチングを用いた。マッチした区間を用いて単語単位で言い直し判定を行うことで理解候補中のどのカテゴリが訂正されたかの判定が可能である。言い直しを検出したときの効果は、以下の通りである。

- システムの確認発話に対しユーザの言い直しを受け取った場合、提示した理解が誤っていることになる。よってその候補の信頼度を下げる
- 言い直し前と後の単語は同じ発話だと考える。よって、2つの単語の認識結果において共通して出現した認識結果は信用できるため信頼度を上げる

信頼度の上下度合いは現在は先見的に与えているが、将来的には言い直し判定で得られるスコアを言い直し判定の正解率に統計に基づいて変換することが望ましい。

## 5 音声対話システムの構築

### 5.1 対話システムの構成

本音声対話システムの構成図を図3に示す。ユーザの発話は、本研究室で開発された音声認識器 SPOJUS [13] の CFG 版と言い直し検出部に送られ、それぞれの結果が理解更新部に送られる。そこで、過去のユーザ発話を踏まえた現在の複数理解候補と、得られた複数の認識結果、言い直し検出結果を基に最新の理解候補へと更新する。最新の理解候補に基づき、提案したシステム応答の選択手法により応答が選ばれ、音声合成器を通じてユーザに発話される。この動作をシステムが要求する情報(県名、市名、町名、目標物名)すべてが保持されるまで繰り返す。

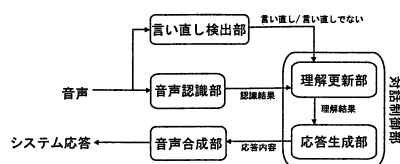


図3: 対話システムの構成図

## 5.2 タスク

本システムで扱うタスクは、カーナビゲーションを想定した目的地設定とする。タスク達成に必要な項目は県名、市名、町名、目標物であり、最初の発話についてはどの項目を入力するかはユーザの自由とする。その後の対話はシステム主導で行い、ユーザの意中の結果が得られるまで対話を行う。システムが要求するカテゴリはこれまでの例のように、県名、市名、町名、目標物である。また、カテゴリは県名-市名-町名と上下関係があり、目標物についても一部のものについては相関がある。本システムではアイズフリーの実現のために、音声だけでシステムの理解を図り、ディスプレイには現在の理解提示を行わない。本システムで扱う認識辞書の内訳は、県名:5、市区町村名:22、町名:139、目標物:29 でパープレキシティは 91.7 である。

## 5.3 提案手法による対話例

構築したシステムで得られた対話の一部を図 4 に示す。例では、ユーザの最初の発話 (User1) に対する理解候補を絞り込むために、システムは県名を尋ねる質問を行った (System2)。それに対するユーザの 2 発話目 (User2) の認識結果より、1 発話目の「調布」という候補が誤っている可能性が高いと判断した。ここで認識結果にシステムの尋ねた市名でない候補(「見付」)が出現しているが、これはユーザの返答として考えられないものとし、理解候補から除外する。まだ、「大須」と「豊橋市」で迷っているためシステムは市名を尋ねる応答を行った (System3)。このときのユーザの 3 発話目は 1 発話目の言い直しと判定され、それぞれの認識結果の N-best 結果を統合すると豊橋市のほうが正しいであろうと理解した。この例では結果としてユーザと対話を繰り返すことで、誤認識からの回復ができています。

## 6 評価実験

### 6.1 実験条件

提案した対話戦略の評価を目的地設定タスクで行った。評価基準としては、ユーザがシステムを使用した際のタスク完了までのターン数 (システムの応答に対してユーザが発話すごとに 1 ターンとする)、タスクの達成率 (タスクが未達成の場合、もう一度タスクをやり直す)、タスクの無矛盾率、確認応答中の肯定表現(「はい」など)の割合、ユーザの満足度などを求める。

本稿では以下の対話戦略を比較する。

- ベースライン 1 (毎回確認)
- ベースライン 2 (尤度の低い発話の場合だけ確認)
- 提案手法

#### 6.1.1 計算機シミュレーション

擬似的にユーザ発話を作り、想定単語認識率ごとに 1000 回のシミュレーション<sup>1</sup>を行った。発話の信頼度については、正解の認識候補の信頼度の方が誤った候補よりも高くなるが多くなるよう調整した分布を使用した。提案手法における 5 のシステム応答の生成には、3 節の方法を用いる。2 つの尺度の重みは互いに 1.0 とした。毎回確認発

<sup>1</sup>先行研究 [1][2] では、カテゴリ語彙数の差によって生じる認識率  $R$  の差をシミュレートするため、認識率に差を付けていたが、100%近くの認識率で不具合が出るため、県名の認識率を  $1 - (1 - R) \times 0.8$ 、町名の認識率を  $1 - (1 - R) \times 1.2$  と修正。

System1:目的地を設定してください

User1:豊橋市のコンビニに行きます

認識結果

- 大須 コンビニ
- 豊橋市 コンビニ
- 調布市 コンビニ

理解候補

1:	(愛知県)	(名古屋市中区)	大須	コンビニ
2:	(愛知県)	豊橋市	—	コンビニ
3:	(東京都)	調布	—	コンビニ

System2:県名を教えてください

User2:愛知県です

認識結果

- 愛知県
- 見付

理解候補

1:	愛知県	(名古屋市中区)	大須	コンビニ
2:	愛知県	豊橋市	—	コンビニ
6:	(東京都)	調布	—	コンビニ

System3:市名を教えてください

User3:豊橋市だよ

認識結果

- 豊橋市
- 愛知県

理解候補

1:	愛知県	豊橋市	—	コンビニ
2:	愛知県	(名古屋市中区)	大須	コンビニ

図 4: 構築したシステムの対話例

話を行う対話システムについては、ユーザの発話に対して確認発話を行う。確認発話に対して肯定が得られた場合は、次の応答は空いているカテゴリについて尋ねる。尤度が低い場合だけ確認を行う対話システムでは、認識成功では半分程度、誤認識時では 1/10 程度で確認発話がかかるように信頼度の閾値を設定した。いずれの手法においても、確認発話で誤認識を確認した際には、ユーザは否定ではなく訂正を行うこととする。タスク完了条件は全てのカテゴリに対して、ユーザの肯定発話を得られた時とする。また、本実験ではタスク失敗と見なすターン数を 16 とし、これを上限とした。シミュレータのタスク完了までの対話の流れとしては次のようになる。

- ユーザの目的地をランダムに決定 (タスク完了条件となる)
- 最初の応答「県名を入力して下さい」を発話
- 目的地的に基づくユーザ発話 (テキスト) を生成
- 誤認識も含めたユーザ発話の認識結果とその信頼度を擬似的に作成
- 認識結果とこれまでの理解状態から最新の理解候補を作成
- タスク完了条件を満たしていなければ最新の理解候補からシステム応答を生成し 3 へ

音声認識部での想定単語認識率は、60%から100%まで5%刻みで実験し、評価項目としてターン数平均とタスク達成率、無矛盾率、各矛盾の発声回数、肯定発話率、肯定発話と否定発話の回数を求めた。

### 6.1.2 シミュレーション対話の被験者による聴取

上記のシミュレーションから得られた対話ログから対話音声試料（ユーザ発話は録音音声・システム発話は合成音声、車内雑音を重畳）を作成し、被験者に聴取してもらい手法の比較を行う。提案手法と比較手法2種から一対を交互に聴取させ、前後の順番を入れ替えた6パターンで主観評価を行ってもらい。順番を入れ替える、均一化により先の聴取の影響を減らし、被験者にシステムの内容を察せられるのを防ぐためである。この方法はシステムの習熟時間がかからず短時間で評価できる、実対話の聴取に比べて認識率など基準を揃えるのが容易でありながら対話音声試料を大量に作りやすいという利点があり、また被験者に実対話と錯覚させやすいのが特徴である。しかし対話を第三者が聴取するので、対話当事者の評価でない欠点がある。

音声認識部の想定単語認識率は60%、75%、90%の3段階を使う。被験者は1つの認識率で24対話（1パターンの1手法ごとに2回の聴取で6パターン）聴取してもらい、3段階の認識率をそれぞれ7人ずつの被験者によって主観評価を行った。比較項目としては効率性、自然性、システムの賢さ、システムの満足度を5段階で相対評価すると共に、試験者側と被験者側の矛盾や自然性の定義の摺り合わせを行うために、自然性の判断基準や改良の要望を記述してもらった。

#### 6.1.3 被験者による実使用

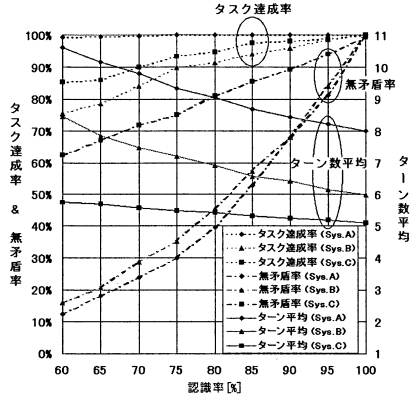
試験者側が提示する目的地（1手法ごとに6箇所）を対話システムを通して被験者に順々に入力してもらい、5.1.2の聴取と同様に効率性、自然性、システムの賢さ、システムの満足度を5段階で絶対評価すると共に、各設問ごとの判断基準や改良の要望を記述してもらった。対話システムの習熟には最大30分までの練習時間を設けた。誤った確認発話には否定以外にも訂正が可能なことや提案手法には複数カテゴリの入力可能なこと、また、合成音声自体の質や応答遅延時間に関しては評価基準から外すなど、事前に幾つかの注意事項を与えた。対話戦略の比較には、システムごとのユーザ満足度等の平均を用いる。

## 6.2 実験結果

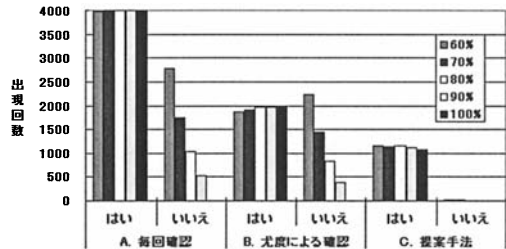
### 6.2.1 計算機シミュレーション結果

計算機シミュレーションの実験結果を図5に示す。同時に、矛盾回数と確認発話への応答回数の分析も同図にまとめた。効率性・自然性ともに提案手法が優位であり、「いいえ」が激減しているなどシステムの誤認識をユーザに悟られずに対処していることが判る。戦略Aはターン数平均・無矛盾率とも悪いものの、高いタスク達成値を有するロバストなシステムであることが示された。一方、戦略Bではターン数平均・無矛盾率が良いが肯定発話の出現割合が悪く、効率よく不自然なシステムとなっている。また信頼度次第で誤認識を確認なく受容する特性上、タスク達成率が低い。

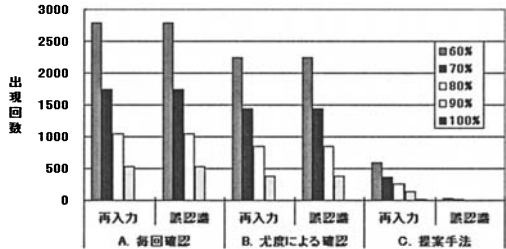
提案戦略Cでは複数カテゴリの発話を許していたが、結果が安定しない上に悪かったので比較手法と同様に単一カテゴリの発話のみで評価してある。結果から、提案手法においてタスク完了までにかかる平均ターン数の減少が見られ効果が確認できた。



(a) ターン数平均、無矛盾率、タスク達成率



(b) 矛盾回数



(c) 確認応答回数

図5: シミュレーションによる対話戦略の比較

### 6.2.2 シミュレーションの聴取による結果

聴取実験における各認識率ごとの比較手法と提案手法の比較結果を図6に示す。提案戦略Cは満足度においてベースラインA、Bを上回っている一方、高認識率の条件下では自然性が尤度による確認手法と同等以下であった。矛盾や不自然さの記述の結果、確認発話の減少はシステムの使い易さに繋がる事が再確認でき、また試験者側の定義した矛盾が被験者側と一致していることが判明したが、それ以外の判断基準も多く、特に提案手法で起きる非シケンシャルな対話の流れが自然性を損なうと感じられたようである。一方、低認識率条件下では効率よく入力できることが、自然でもあり賢いシステムであると感じた被験者が多かった。評価手法としての感想では、モーラ数の差や誤り音素数などに関係なくランダムに作成した誤認識候補の間違い度合が気になる、システム発話を録音音声にしたり話速を速めることで満足度が上がるとの意見もあった。

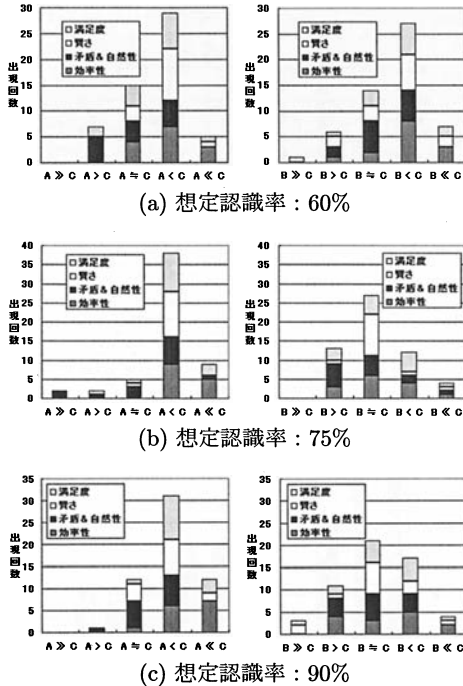


図 6: 聴取実験のアンケート結果

### 6.2.3 実使用による結果

被験者実験の結果は、使用実験におけるシステムごとの平均を表 1 に示す。提案戦略 C は満足度においてベースラインを上回っている一方、高い認識率下のためか自然性が戦略 B を下回っている等、シミュレーションの聴取結果と同様な結果となった。実使用では辞書のサイズのためか殆ど誤認識は起きなかったが、苦手な単語では何度も入力を繰り返す、肯定否定表現との誤認識が存在するなどの特徴があり、また言い直し発話よりも否定発話と再入力を使用する頻度が高いなどシミュレーション結果との乖離の原因となった。今後のシミュレーションに反映したい。

また、戦略 B 及び提案手法におけるタスク達成率はあまり高いものではなかったが、原因としては重みや信頼度閾値の調整不足や、誤答がたまたま高いスコアを得た時に正答を認識しても理解候補更新のルールが未熟なために効果的な応答をシステムが出来ず、対話がループした例が存在したことが挙げられる。その場合でも正答が 2 位以下に存在することが多く、正答を揃い上げる応答で対処できる可能性がある。

表 1. 実使用による対話戦略の比較

対話戦略	主観評価項目				客観評価項目				
	効率性	自然性	賢さ	満足度	ターン数平均	認識率	タスク達成率	肯定発話率	無矛盾率
System.A	2.90	3.43	2.98	2.98	9.95	89.3 %	90.8 %	84.2 %	50.0 %
System.B	3.55	4.00	3.43	3.60	5.55	96.7 %	67.8 %	71.8 %	83.3 %
System.C	3.90	3.55	3.86	3.81	3.74	92.3 %	65.5 %	98.3 %	80.9 %

## 7 まとめ

本稿では、システムの誤認識を対話を通じて自然かつ効率的に回復する対話戦略を、目的地入力タスクによって主観・客観の双方で評価した。誤認識の対処のために確認発話を多様な戦略を指した対話戦略として、各対話ターンにおいて複数の理解候補を保持し、対話を通して最適な仮説を探索した。その際には、理解候補を収束させる効率性の尺度と、理解候補と矛盾しない無矛盾性の尺度を求め、その和により最終的なシステム応答を決定した。これらの手法を用いた音声対話システムを構築し、実際に被験者が使用して評価を行った。また、計算機シミュレーションと対話聴取によっても評価を行い、自然で効率的な戦略であることを示した。今後は理解候補更新のルールの高精度化や新たな応答候補を使用する、また PARADISE で満足度を客観評価で分析するなどにより、更に効率的で自然な対話が期待できる。

## 参考文献

- [1] 北岡 教英, 矢野 浩利, 中川 聖一: 「誤認識に対処した自然で効率的な音声対話戦」, 情報処理学会研究報告, 2006-slp, pp61-67, (2006.5)
- [2] 矢野浩利, 北岡教英, 中川聖一: 「誤認識の修復機能を備えた音声対話システムの構築」, 言語処理学会, B1-4, pp.38-41, (2006.03)
- [3] 駒谷 和範, 河原 達也: 「音声認識結果の信頼度を用いた効率的な確認・誘導を行う対話管理」, 情報処理学会論文誌, Vol.43 No.10, pp3078-3086, (2002.10)
- [4] 藤原敬記, 伊藤敏彦, 荒木健治, 甲斐孝彦, 小西達裕, 伊東幸宏: 「認識信頼度と対話履歴を用いた音声言語理解手法」, 電子情報通信学会論文誌, Vol.J89-D, No.7, pp.1493-1503, (2006.7)
- [5] 東中 竜一郎, 中野 幹生, 相川 清明: 「複数文脈を用いる音声対話システムにおける統計モデルに基づく談話理解」, 情報処理学会研究報告 (SLP-45-17), pp.101-106, (2003.2)
- [6] 翠 輝久, 駒谷 和範, 清田 陽司, 河原 達也: 「音声対話によるソフトウェアサポートタスクのための効率的な確認手法」, 電子情報通信学会論文誌, Vol.J88-DII, No.3, pp.499-508, (2005.3)
- [7] 堂坂浩二, 安田宣仁, 相川清明: 「システム知識制限下での効率的音声対話制御法」, 自然言語処理, Vol.9, no.1, pp.43-63, (2002.01)
- [8] M. Walker, D. Litman, C. Kamm, and A. Abella, "Paradise: A framework for evaluating spoken dialogue agents," in Proc. of the 35th Annual Meeting of the Association for Computational Linguistics, pp.271-80, (1997)
- [9] T. Watanabe, M. Araki, and S. Dshita, "Evaluating dialogue strategies under communication errors using computer-to-computer simulation", in Proc. of the Trans of IEICE Vol.E81-D, No.9, pp.1025-1033, (1998)
- [10] Fan Yang and Peter A. Heeman, "Exploring Initiative Strategies Using Computer Simulation", In Proc. of Interspeech, pp.106-109, (2007)
- [11] Olivier Pietquin, "Learning to Ground in Spoken Dialogue System" In Proc. of ICASSP, pp.IV165-168, (2007)
- [12] 北岡 教英, 角谷直子, 中川 聖一: 「音声対話システムの誤認識に対するユーザの繰り返し訂正発話の検出と認識」, 電子情報通信学会論文誌, Vol.J87-D-II No.7, (2004.7)
- [13] 日本語連続音声認識システム SPOJUS-SYNO, <http://www.slp.ics.tut.ac.jp/SPOJUS/>