

確認対象ごとの認識性能を考慮する効率的対話制御法

安田 宜仁[†] 堂坂 浩二[†] 相川 清明[†]

音声対話システムは、音声認識誤りに対処するために、確認対話と呼ばれるユーザ要求を把握するためのやりとりを実施する。ユーザの要求内容は属性と値の対で表現される。確認対話の手順は、確認する属性の組合せと確認の順序に応じて様々である。この様々な手順の中からできるだけ少ないやりとりの回数でユーザの要求内容の確認を終える対話制御法を提案する。本手法では、確認手順の各時点の音声認識率を予測し、確認を終えるために必要なやりとりの回数を推定する。音声認識率はシステムが認識対象とする語彙に基づいて予測する。認識語彙は、対話の進行上認識する必要のある一般的な語彙と、確認に対してのユーザ応答として可能な語を含む語彙から成り、認識誤りを少なくするために属性間の語彙の依存関係を考慮したうえで決定する。コンピュータ上に実装された模擬ユーザとのシミュレーション対話を行い、提案法が従来法に比べて少ないやりとりの回数で確認対話を終えることができることを実証した。

An Efficient Dialogue Control Method Based on Recognition Rate for Confirmation Targets

NORIHITO YASUDA,[†] KOHJI DOHSAKA[†] and KIYOAKI AIKAWA[†]

To address the problem of recognition errors, spoken dialogue systems prompt the user to confirm the user's request, which is called a confirmation dialogue. A user's request is represented as a set of attribute-value pairs. A confirmation dialogue is made according to the combinations of attributes to be confirmed at a given time and the order of the combinations and involves various procedures. We propose a new dialogue control method that enables the system to select, among the various procedures, one that completes the confirmation of the content of a user's request in as few exchanges as possible. In this method, the system predicts the recognition rate at each step of the confirmation procedure and estimates the number of exchanges to complete the step. The predicted recognition rate is determined based on the recognition vocabulary. The recognition vocabulary comprises a general vocabulary that the system has to recognize to carry out the dialogue and a vocabulary that covers the words in the answers to the confirmation questions. The vocabulary for answers from the user is determined dynamically whereby dependencies among attributes is taken into account in order to minimize recognition errors. Dialogue experiments performed using a computer-simulated user prove that the proposed method outperforms conventional ones.

1. はじめに

音声対話システムとは、コンピュータと人間が音声によるやりとりを通して、あらかじめ決められた種類の仕事を行うようなシステムのことである。この仕事を音声対話システムが遂行するタスクと呼ぶ。音声認識技術と言語処理技術の発展を背景に、これまでに様々なタスクでの音声対話システムが開発されて

きた。たとえば観光案内¹⁾や、飛行機での旅行案内²⁾、天気情報案内³⁾、TV番組予約⁴⁾といったタスクを行うシステムが開発されている。音声対話システムの1つのタスクは1つ以上の部分タスクに分割することができ、各部分タスクにおいて、システムは音声を通じてユーザの要求内容を把握し、要求内容に応じた処理を行う。

しかし、現在の技術では人の音声を完全に認識することは不可能であり、音声認識誤りを避けることはできない。このため、音声認識誤りを考慮せずにユーザの発話の認識結果だけから処理を行おうとしても、ユーザが要求している内容をシステムは把握できずに、タスクを遂行することができない。このため、システムは、伝達された要求内容の確認をユーザに対して行う。

[†] 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所

NTT Communication Science Laboratories, NTT Corporation

現在、東京工科大学メディア学部

Presently with School of Media Science, Tokyo University of Technology

さらに、ユーザの発話内容が過少であったり、あるいは音声認識結果の脱落誤りによって、システムに伝達された内容だけではユーザの要求内容を把握できない場合がある。この場合システムは不足している情報について、ユーザに発話を要求するのが一般的である。以上のような、システムからの確認や情報の要求にともなう行われる会話のやりとりのことを確認対話と呼ぶ。システムからの確認内容に誤りがあれば、ユーザは訂正を行うことになるが、訂正が正しく認識されるとは限らない。もしも音声認識精度が悪く、訂正内容をシステムがうまく把握できない場合には、ユーザは訂正を繰り返し行わなければならない可能性がある。タスクを確実に遂行するためには確認対話は必須であるが、このような繰返しを何度も行わなければならないようではユーザにとっては煩わしいので、確認対話のためのやりとりの回数は少ない方が望ましい。

ユーザの要求内容は属性と値の対の集合として表現される。この属性の種類および値の値域はタスク設計時に決められる。通常のタスクにおいては、複数の属性によってユーザの要求内容が表現されるため、属性の値を確認対話を通じて確定していく手順は、確定する属性の組み合わせ方と、確定の順序によって様々である。これらの属性の組合せや順序の違いはやりとりの回数に影響を与える。

まず、属性の組合せに関して、仮にユーザの発話が認識誤りなくシステムへ伝えられるとすれば、一度にまとめて確認や情報要求をすれば、少ないやりとりの回数で確認対話を終えることができるだろう。しかし、現実には音声認識誤りの可能性は無視できない。複数の属性をまとめて確認対象とした場合、それらの属性の値となりうる語彙をすべて認識対象とする必要がある。一般に、認識語彙を増やせば増やすほど認識精度は低くなるので、多くの属性をまとめて確認対象にしてしまうと、認識誤りが増え、その結果訂正を繰り返すような効率の悪い対話の原因になる可能性がある。一方で、音声認識誤りを減らすために1属性ずつ確認することによって、認識対象の語彙を絞り、認識精度を高くすることによって不要な対話を防ぐことも考えられる。しかし、個別に確認を行ってしまうと、たとえば音声認識の誤りがなかったとしても、少なくとも属性の数だけのやりとりをすることになり、やりとりの回数に関して非効率である。

次に、確定していく順序に関しては、所属と名前や、県名と市名などのようにどちらか一方の属性の値が確定することによって、他方の候補の値が定まったり、絞り込めたりする場合がある。このように属性間に語

彙の依存関係がある場合は、確定していく順序が異なると確認対話のやりとりの回数に影響を与える。

音声対話システムがユーザからの要求を効率的に把握するために、従来より様々な方法が提案されている。文献5)、6)では、対話を通じて認識率は一定である場合に、いくつかの確認手順を対話をモデル化したうえで数学的に分析、比較している。文献7)では、属性の確定までの平均対話回数を音素認識行列と音素系列で表した語彙定義から求め、平均対話回数が少なくなるような属性の確定順序を選択する方法も提案されている。しかし、これらの従来法は、いくつかの事前に決めた確認手順の中での優劣や、属性は1つずつ順に確定していくことを前提にしたうえで属性の確定順序について検討しており、複数の属性をまとめて確認することを含めて網羅的な確認手順の中から最適なものを選ぶものではなかった。前述のとおり、確認対象の順序の違いや組合せの違いは確認対話の効率に影響を与えられとされる。また、対話制御のためのシステム戦略をマルコフ決定過程によってモデル化し、最適化のために強化学習を用いる手法^{1),8)}も存在する。学習のためには相当数のシステムとユーザとの対話例が必要となる。本手法で対話例を集めずに対話の効率を高めることを議論する。これらの従来法のいくつかは確認の方法として間接的な確認を含めている。間接的な確認とは、確認以外のシステムの発話に確認内容を含み、ユーザからの否定がなければ発話に含まれた確認内容を正しいと見なす方法である。間接確認は、もし確認内容に誤りが無い場合は、ユーザが肯定の意図を示さなくても対話が進行するので、やりとりの回数を減らすことができる。しかし、もし確認内容に誤りがある場合には、ユーザは否定の代わりに肯定の返事をしてしまったり、沈黙してしまったりすることによって、ユーザが訂正をうまく行えなくなる傾向がある⁹⁾。このため、ユーザの要求内容を誤って把握してしまうことの代償が大きいタスクなどでは確認と訂正をより確実に行うために直接確認を行うことが望ましい。本手法では確認はすべて直接確認によって行うという前提の下で議論する。

本稿では、対話の各時点において必要な語彙に応じた認識率を推定し、その認識率の下で期待されるやりとり回数が少ないという意味でできるだけ効率的な確認手順を採る対話制御の方法を提案する。

2. 対象となる音声対話システム

提案法で対象とする音声対話システムの構成図を図1に示す。システムは音声理解部、対話制御部、発

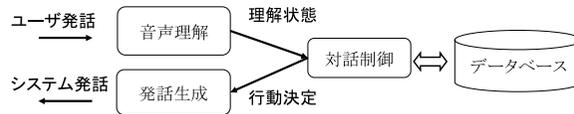


図 1 システム構成図

Fig. 1 System's architecture.

話生成部とデータベースより構成される。音声理解部はユーザ音声からユーザ発話内容を理解し、理解した結果をシステムの理解状態として保持する。対話制御部はシステムの理解状態を受け取り、システムの次の行動を決定し、決定した内容を発話生成部に伝える。発話生成部は対話制御部によって決定された内容に沿った発話を生成する。

音声理解部によって作成されるシステムの理解状態は、

(属性名 値 確定フラグ)

の 3 つ組を要素とするような集合として表現される。属性の種類と各属性のとりうる値の範囲はタスク設計時にシステム設計者が決定する。確定フラグとはその属性の値がすでに確認済みであることを表すフラグであり、システムからユーザに対して属性の値の確認を行った際に、その確認に対するユーザからの肯定の返事をシステムが受け取った場合に「確認済み」となる。

ユーザ要求を把握するためにシステムは確認発話あるいは情報要求発話を行う。確認発話とは、属性の値は入っているが未確定の属性の値について、その属性の値についての確認を行う発話である。たとえば「神奈川県ですね?」「明日の 10 時からですね?」などである。システムからの確認発話に対して、ユーザから「はい」「そうです」のような肯定の返事を受け取った場合にシステムは確定フラグを「確認済み」とする。一方「いいえ」「ちがいます」のような否定の返事を受け取った場合には、システムは確認対象となっていた属性の値を初期化する。情報要求発話とは、まだ値が定まっていない属性について、その属性についてのユーザ発話を求めるための発話である。たとえば「場所はどこですか?」「曜日と場所を指定してください」などである。システムからの情報要求発話に対しては、ユーザは「兵庫県です」「月曜日の北海道です」のように要求された属性についての応答をすることを求められる。これら確認発話と情報要求発話によって行われるシステムとユーザとの会話のやりとりのことを確認対話と呼ぶ。

システムは確認と同時に、確認対象属性以外の属性についての情報要求を行うような発話は行わないもの

とする。このようなシステム発話(たとえば「月曜日ですね? 場所はどこですか?」)は、確認内容に誤りが含まれていない場合は少ないやりとりの回数で対話を終える可能性があるが、確認内容に誤りが含まれている場合には、間接確認同様にユーザがうまく訂正を行えなくなることが予想されるからである。

提案手法ではユーザからの肯定否定の意図は誤りなく伝わるという仮定の下でやりとり回数の推定を行う。しかし、現実の音声認識においては、「はい」「いいえ」といった肯定否定の意図を表す発話を誤りなく認識することは困難である。このため、提案法を用いるシステムにおいてはできるだけ確実な方法でユーザからの肯定否定の意図を把握することが望ましい。たとえば、ユーザ発話の韻律情報を用いて、ユーザが訂正を行っているのかどうかを判定する手法が提案されている¹⁰⁾。あるいは、肯定否定に対応する専用の意図伝達方法を別途用意し、ユーザに教示を事前に行うことも考えられる。たとえば、「はい」ではなく「はいそのとおりです」、「いいえ」ではなく「いいえ間違っています」を必ず用いてもらうことにより音声認識の誤りを減らす方法や、肯定否定に対応するボタンを押したり、スイッチを切り替えたりするといった方法などである。

システムが確認対話中に用いる音声認識器は、確認の対象に応じて語彙を切り替えることができるものを用いる。この際の認識語彙は対話の進行上認識する必要のある一般的な語彙と、確認あるいは情報要求に対してのユーザ応答としてシステムが受け付け可能な語彙とによって構成する。認識語彙から現在の対話で不要な語彙を省くことにより、より高い精度での認識が行えることが見込まれるからである。システムが認識語彙から認識率の推定を行うために、認識器の性能の目安を知る必要がある。この目安として何らかの認識語彙の下での認識率は既知であるとする。

システムは与えられたすべての属性の値を確定できた場合に、ユーザの要求内容を把握したものとしてデータベースの検索や情報の案内などといった、ユーザの要求に応じた処理を行う。

3. 提案法

3.1 概 略

システムが確認対話中に行うことができる行動は、確認行動と情報要求行動のいずれかであるとする。確認行動とは、対象とした属性集合について、値が確定するまで確認発話を繰り返し行う行動である。情報要求行動とは、対象とした属性集合についての情報要求発話を行い、ユーザから属性集合のそれぞれの値を聞き出す行動である。システムは現在の理解状態と確認対話終了の状態、つまり必要な属性の値がすべて確定した状態を比較することによって、各属性についてシステムが行わなければならない行動を決定することができる。確認対話終了までにシステムがとるべき行動の列をシステムの確認プランと呼ぶ。本手法は、確認対話終了までのやりとり回数がある可能性のあるシステムの確認プランを網羅的に生成する。確認プラン中の各行動時に必要な認識語彙は、プラン中のその行動の前の行動までが終了したと仮定したうえで、対象の属性の値域と、属性間の語彙の依存関係を用いて推定する。この認識語彙に基づいて、その行動を行う際のシステムの音声認識の認識率を予測し、行動終了までに必要なやりとり回数の推定を行う。各確認プランでの確認対話終了までに必要なやりとり回数を、確認プラン中の各行動で必要なやりとり回数の和として推定し、確認対話終了までに必要なやりとり回数が最も小さい確認プランを採用し、システムの次の行動を決定する。

3.2 可能な確認プランの網羅的作成

現在の理解状態においてシステムが選択可能な行動の組合せのうち、期待やりとり回数の推定結果が異なる可能性があるものをすべて網羅する手順を説明する。確認プランとは確認対話終了までの行動の列であり、各行動は行動の種類（確認行動あるいは情報要求行動）と対象の属性集合から成る。各属性について、確認対話終了までに必要な行動は、値が埋まっていない属性については情報要求行動と確認行動であり、値が埋まっている属性については確認行動であると決定することができる。確認プランはどれも各属性について必要な行動を過不足なく含んでいる。各確認プランの違いは、対象属性の組合せ、および行動の順序の違いである。しかし、組合せと行動の順序の違いがあるからといって、期待やりとり回数という意味では違いがない場合もある。たとえば、お互いの対象属性間に語彙の依存関係がないような2つの行動は、それらの行動をどちらを先に行ってもそれぞれの行動を行う

時点で必要な認識語彙に違いはないので、これらの順序の違いは期待やりとり回数には影響を与えない。期待やりとり回数に影響を与える可能性があるという意味において考慮すべきは、一度の行動において対象とする属性の組合せと、依存関係がある2つの属性を対象として含んでいるような2つの行動の順序である。行動の対象とする属性の組合せについて、対象属性が異なる組合せになれば、認識語彙が異なるので、行動の対象属性の組合せの違うものはすべて異なるプランとして取り扱う。また、依存関係がある2つの属性を確定する場合、一方の属性の値が確定することによって他方の属性の値の範囲が絞り込まれる。たとえば名簿や電話番号簿を扱うシステムにおいて「所属」と「名前」を属性として持つような場合に、「所属」を含む行動を先に行えば、「名前」の値域は特定の所属の人の名前に絞られる。一方「名前」を含む行動を先に行えば、「所属」の値域は特定の名前の人を持つ所属に絞られ、先ほどの「所属」を含む行動を先に行った場合とは認識語彙が異なる。

以上より可能な確認プランを網羅的に作成するアルゴリズムは以下のとおりである。現在の理解状態において、値が埋まっていない属性の集合を R (情報要求が必要な属性の集合)、確定フラグが「確定済み」でない未確定の属性の集合を C (確認行動が必要な属性の集合) とする。なお、情報要求が必要な属性は確認行動も必要になるので、 R は必ず C の部分集合となる。 R のすべての分割から成る集合 P_R 、 C のすべての分割から成る集合 P_C をそれぞれ作成する。 P_R の各要素 (R の分割) の各要素に対して、情報要求を表す「R」のラベルを付与する。同様に、 P_C の各要素に対して、確認を表す「C」のラベルを付与する。こうして、 P_R 、 P_C の各要素 (属性の集合) と行動を対応付けることができる。「C」ラベルが付いた属性集合はその属性集合を確認対象とした確認行動であり、「R」ラベルが付いた属性集合はその属性集合を情報要求対象とした情報要求行動である。ラベル付け後の P_R と P_C の直積集合 S を作成する。 S の各要素は、必要な情報要求行動と確認行動を過不足なく持つような、行動の集合となる。これら行動の集合 (S の要素) それぞれについて、確認行動よりも情報要求行動が前になるように並べ換える。このような並びは複数存在する場合があるが、やりとり回数の推定には影響はないので無作為に1通りの並びを選択する。こうして並べ換えられた行動の列を $K_i = a_{i1}, a_{i2}, \dots, a_{i|K_i|}$ ($i = 1, \dots, |S|$) と呼ぶ。各 K_i から確認プランを得るために以下の操作により確認プランを作成する。もし K_i 中のすべて

チャンネル	8	確定
録画日	11/17	確定
録画開始時刻	9:00	未確定
録画終了時刻	10:00	未確定
画質	未定	未確定

図 2 ビデオ予約タスクにおける理解状態の例

Fig. 2 An example of system's understanding in a video reservation task.

の行動に関して、依存関係がある属性を異なる行動の対象としていなければ(つまり、依存関係がある属性はすべて同一の行動の対象となっていれば)、 K_i をそのまま 1 つの確認プランとする。そうではなく、もし K_i 中の行動に、依存関係がある 2 つの属性を異なる行動の対象として含んでいる場合はそのような行動の対 (a_{il}, a_{im}) を 1 つ選び、入れ替えによって作成される 2 つの列 $(\langle a_{i1}, a_{i2}, \dots, a_{il}, \dots, a_{im}, \dots, a_{i|K_i|} \rangle)$ と $(\langle a_{i1}, a_{i2}, \dots, a_{im}, \dots, a_{il}, \dots, a_{i|K_i|} \rangle)$ を作成する。作成された列に、依存関係がある 2 つの属性を異なる行動の対象として含んでいるけれどもまだ列の入れ替えを作成していないような行動の対があるならば、同様の操作により入れ替えによる列の作成を再帰的に繰り返す。このようにしてできたすべての列をそれぞれ確認プランとする。

例として、ビデオ予約タスクを考える。属性として「チャンネル」、「録画日」、「録画開始時刻」、「録画終了時刻」、「画質」があり、このうち「録画開始時刻」と「録画終了時刻」の間に依存関係があるとすると、このビデオ予約タスクにおいて図 2 のような理解状態を持つ場合、つまり、

情報要求行動が必要な属性集合:

$$R = \{ \text{画質} \}$$

確認行動が必要な属性集合:

$$C = \{ \text{録画開始時刻}, \text{録画終了時刻}, \text{画質} \}$$

である場合の確認プランの生成例を以下に示す。 R のすべての分割から成る集合 P_R 、および C のすべての分割から成る集合 P_C に、確認を表すラベル(「 C 」あるいは「 R 」)を付与する。

$$P_R = \{ \{R(\text{画質})\} \}$$

$$P_C = \{ \{C(\text{画質} \text{ 録画終了時刻}) C(\text{録画開始時刻})\}, \\ \{C(\text{画質} \text{ 録画開始時刻}) C(\text{録画終了時刻})\}, \\ \{C(\text{画質}) C(\text{録画開始時刻} \text{ 録画終了時刻})\}, \\ \{C(\text{画質}) C(\text{録画開始時刻}) C(\text{録画終了時刻})\}, \\ \{C(\text{画質} \text{ 録画開始時刻} \text{ 録画終了時刻})\} \}$$

次に、 P_R と P_C の直積集合 S を以下のように作成する。

$$S = \{ (R(\text{画質}) C(\text{画質} \text{ 録画終了時刻}) C(\text{録画開始時刻})), \\ (R(\text{画質}) C(\text{画質} \text{ 録画開始時刻}) C(\text{録画終了時刻})),$$

$$(R(\text{画質}) C(\text{画質}) C(\text{録画開始時刻} \text{ 録画終了時刻})), \\ (R(\text{画質}) C(\text{画質}) C(\text{録画開始時刻}) \\ C(\text{録画終了時刻})), \\ (R(\text{画質}) C(\text{画質} \text{ 録画開始時刻} \text{ 録画終了時刻})) \}$$

「録画開始時刻」と「録画終了時刻」の間に依存関係があるので、これらを異なる行動で対象としている行動については、行動の入れ替えを行った列を作成し、確認プランを作成する。

$$(R(\text{画質}) C(\text{画質} \text{ 録画終了時刻}) C(\text{録画開始時刻})) \\ (R(\text{画質}) C(\text{録画開始時刻}) C(\text{画質} \text{ 録画終了時刻})) \\ (R(\text{画質}) C(\text{画質} \text{ 録画開始時刻}) C(\text{録画終了時刻})) \\ (R(\text{画質}) C(\text{録画終了時刻}) C(\text{画質} \text{ 録画開始時刻})) \\ (R(\text{画質}) C(\text{画質}) C(\text{録画開始時刻} \text{ 録画終了時刻})) \\ (R(\text{画質}) C(\text{画質}) C(\text{録画開始時刻}) C(\text{録画終了時刻})) \\ (R(\text{画質}) C(\text{画質}) C(\text{録画終了時刻}) C(\text{録画開始時刻})) \\ (R(\text{画質}) C(\text{画質} \text{ 録画開始時刻} \text{ 録画終了時刻}))$$

たとえば、先頭に記したプランは、まず画質を情報要求行動によって要求し、次に画質と録画終了時刻を確認行動によって確定し、録画開始時刻を確認行動によって確定するプランである。

3.3 語彙の決定と認識率の推定

確認対話の際にシステムが認識対象とすべき語彙は、「はいそうです」「もう 1 回お願いします」のような対話上必ず認識する必要のある一般的な語彙と、確認に対してのユーザ応答として可能な語を含む語彙、つまり、確認対象となっている属性に入る可能性がある語彙とによって構成する。たとえば、ビデオ予約タスクにおいて録画日時を確認する際には、録画開始時刻、録画終了時刻、録画日時などは認識語彙に含まれるが、チャンネルや画質に関する語彙は録画日時という属性とは無関係な語彙なので、認識語彙に含まれない。属性間には、たとえば県名と市名のように、属性間に語彙の依存関係がある場合がある。このような語彙の依存関係がある場合、確認対話中に、依存関係にある一方の属性が確定していれば、他方の属性に関して確認や情報要求を行う際には、候補を絞ることができ、認識精度を高めることができる。このような語彙の依存関係に関する知識は事前に与えてあり、システムが利用できるかと仮定する。なお、語彙の依存関係に関する知識を与えてあることは無理な仮定ではない。なぜなら、このような語彙の依存関係に関する知識は、音声理解部においてもシステムが矛盾を含んだ理解状態を持ってしまふことを避けるために必要とされる知識であるため、本稿が対象にしている対話制御部が必要としなくても音声理解規則作成時に作成されていると思つてよい。属性間の語彙の依存関係は、各属性がと

りうるそれぞれの値について、もしその値が確定した場合に他の属性においてとりうる値として表す。たとえば、

{(県名 徳島)((市名 阿南市 小松島市 鳴門市 徳島市))} は、属性「県名」の値として「徳島」が確定すれば属性「市名」としてとりうる値は「阿南市」「小松島市」「鳴門市」「徳島市」のいずれかであることを示す。以上から確認対話中の認識語彙は、確認対象になっていない属性の値を除いたうえで、確定している属性との依存関係の制約を満たす語と、対話の進行全般に必要な語によって構成する。

次に、確認の際に期待できる認識率を予測する方法を考える。音声認識の精度に影響を与える要因は様々であるが、システムが対話中に変更できる要因である語彙との関係に着目する。一般に、認識対象の語彙数が多ければ認識率は下がり、逆に語彙数が少なければ高い認識率を期待できることが知られている。このような関係として、誤認識率がパープレキシティの平方根に比例するという経験則が知られている^{11)~13)}。本手法は、語彙の出現頻度を等確率とみなしたうえで、この関係を利用して認識率の予測を行う。他の属性との依存関係がない属性だけを対象とした行動については、各属性の値域より認識語彙を決定することができる。依存関係がある属性を対象とした行動のうち、依存関係にある属性がすべて同一の行動に含まれている場合も、その行動の際にはすべての属性の値域を含めることにより決定することができる。しかし、依存関係がある属性が異なる行動に含まれている場合には、依存関係にある属性を含む行動のうち、後から行われる行動での認識語彙は、属性値域のみからは決定できない。なぜなら、語彙の依存関係は、値が確定した場合に他の属性のとりうる値として与えられているので、先に行われる行動での属性の値が定まっていなかった状態では、依存関係に関する知識を直接利用できないからである。そこで本手法では、このような場合に、先に行われる行動に含まれる属性の値が未定の場合は、その属性がとりうる値それぞれについて、その値が確定した場合の、後から行われる行動での語彙を算出し、そこで期待される認識率の平均を後の行動を行う際の認識率とする。これにより、値が未知の属性に対しても依存関係の影響を考慮した確認プランの評価を行うことを狙う。

3.4 属性の確定に必要なやりとり回数の推定

確定すべき属性が与えられた場合に、確認行動、情報要求行動それぞれについて、各行動が完了するまでに必要なやりとりの回数を推定する方法を考える。シ

ステムからの確認に対してユーザは必ず肯定あるいは否定の返事を行い、ユーザからの肯定否定に関する意図はシステムに誤りなく伝わると仮定する。否定の場合にはユーザは確認されたすべての項目に対して訂正内容を発話すると仮定すれば、やりとり回数が t 回で終わる確認行動とは、 $t-1$ 回の誤りを含んだ確認の後に、正しい確認が行われると思うことができ、認識率を r とすると期待やりとり回数は以下のように求めることができる。

$$r + 2(1-r)r + 3(1-r)^2r + \dots = \sum_{t=1}^{\infty} tr(1-r)^{t-1} = \frac{1}{r} \quad (1)$$

システムからの情報要求行動では、システムからの要求を行い、ユーザからの応答によって理解状態を更新する。この行動のやりとり回数は認識率にかかわらず 1 と定めることができる。

3.5 確認プラン決定

3.2 節で作成した各確認プランを、期待されるやりとり回数を用いて比較する。各確認プラン中の各行動が終了するために必要なやりとりの回数を 3.4 節で述べた方法により推定し、これら各行動に必要なやりとり回数の和をその確認プランにおいて期待されるやりとり回数とする。作成したすべての確認プランの中で、推定したやりとり回数が最も小さいようなプランをその時点におけるシステムの確認プランとする。もし、推定したやりとり回数が同一であるプランが複数ある場合は、その中から無作為に 1 つを選択しシステムの確認プランとする。システムは、システムが発話する際はそのつどこの確認プランの決定を行い、決定されたプラン中の最初の行動に応じて発話を行う。

4. 評価

システムと模擬ユーザとのシミュレーション対話実験によって提案手法のユーザ要求確定までのやりとり回数の評価を行った。模擬ユーザはユーザの振舞いを模擬するプログラムである。模擬ユーザを用いることにより、多数の対話を様々な状況でシミュレーションを行うことができる^{14),15)}。模擬ユーザとシステムとの対話は音声を直接やりとりするのではなく、意味表現の交換によって行う。意味表現は発話意図を表す記号と、属性と値を要素とする集合によって表現される。シミュレーション対話は模擬ユーザとシステムが 1 発話ずつ交替しながら進められる。確認発話は、確認という発話意図と、属性名と値の対から成る集合をユーザプログラムへ伝えることによって行われる。情報要

求発話は、情報要求という発話意図と、属性名の集合を伝えることによって行われる。模擬ユーザは対話開始時に、その対話を通じてシステムへ伝達するユーザ要求を決定する。このユーザ要求は属性の依存関係を満たす範囲で無作為に決定する。模擬ユーザは、システムからの情報要求に対しては要求されたすべての属性について属性名と値の対を伝え、システムからの確認発話に対しては確認内容がすべて正しければ肯定の意図を伝え、確認内容に誤りが含まれていれば否定の意思を伝える。さらに、確認内容に誤りが含まれている場合には、模擬ユーザは否定の意思に加えて一定の確率で訂正内容を伝達するとした。否定の場合の訂正内容を伝える割合に関して、本手法でのやりとり回数での推定では、3.4節で述べたように、ユーザは必ず訂正内容も伝えると仮定している。しかし、実際に必ずユーザが訂正を行うという仮定は無理がある。文献16)では特定のタスクにおける Wizard of OZ 方式の実験において、システムからの確認が誤っている場合にはユーザ応答の 77%は内容語を含むと報告しているが、一般のタスクにおける先験的な知識はない。このため、本シミュレーション対話実験では、訂正内容を発話する確率の異なる 3 種類の模擬ユーザプログラムを用いての実験を行った。

模擬ユーザ 1: 否定の際に、否定の意図のみを伝え、まったく訂正を行わない。

模擬ユーザ 2: 否定の際に、否定の意図に加え、確率 0.5 で訂正を行う。

模擬ユーザ 3: 否定の際に、否定の意図に加え、必ず訂正を行う。

音声認識誤りを模擬するために、情報要求に対する応答と、確認に対する否定の応答に含まれる訂正内容に関して、語彙数の平方根に比例した確率で、属性ごとに置換誤りを発生させた。語彙数に応じて変化する認識器の性能の違いを、同一語彙数下での認識率の違いとし、この認識率を変化させることによって、異なる性能の認識器での比較を行った。本実験ではこの特定の語彙数を 500 とし、500 語における認識率を 0.6 から 1.0 まで変化させた。語彙数と認識率の関係は、システムが認識率の推定に使う関係と同一となっている。しかし、実際の認識の正否は語彙数以外にも様々な要因によって変化するため、システム自体が行う認識率推定は必ずしも正確ではないはずである。そこで、システムの推定結果に対しては誤りを含むようにした。推定誤りの範囲を十分大きくとるため、システムの推定した誤認識率に対して最大上下 50% の範囲の乱数でユーザ発話のつど決定するとした。この範囲は、た

例えばタスク 1 において、語彙数 500 語における認識率が 0.6 のときにすべての項目をまとめて認識対象にしようとした場合には認識語彙は 1,100 語となり、乱数の最大の振れ幅で推定誤りがあった場合では、認識率が約 0% となってしまう。このため推定誤りの設定範囲として十分大きいと考えられる。認識率によらず、模擬ユーザからの肯定・否定の意図は誤りなくシステムへ伝わるとした。

提案手法と比較するために、確認戦略の異なる 4 つの方法と比較した。これらの確認戦略を以下に述べる。

比較手法 1 (個別要求・個別確認): 事前に決められた順序に従って属性を個別に確定する。属性の値が未定ならば情報要求行動をその属性についてのみ行い、属性に何らかの値が入っていれば確認行動をその属性についてのみ行う。属性の値が確定するまで以上を繰り返すことを、すべての属性について行う。情報要求および確認を行う順序は固定で、事前に決められる。

比較手法 2 (一括要求・個別確認): 情報要求行動は一括して行い、確認行動は事前に決められた順序に従って行う。理解状態中に値が未定の属性があればそれらの属性すべてについての情報要求行動を行い、未定の属性がなければ未確定の属性について個別に確認行動を行う。以上を未確定の属性がなくなるまで繰り返す。

確認を行う順序は固定で、事前に決められる。

比較手法 3 (個別要求・一括確認): 情報要求行動は事前に決められた順序に行い、確認行動は一括して行う。理解状態中に値が未定の属性があれば、事前に決められた順序に従って個別に情報要求行動を行い、未定の属性がなければすべての属性についての確認行動を行う。以上を確認行動が成功するまで繰り返す。情報要求を行う順序は固定で、事前に決められる。

比較手法 4 (一括要求・一括確認): 情報要求行動も、確認行動も一括して行う。理解状態中に値が未定の属性があればそれらの属性すべてについての情報要求行動を行い、未定の属性がなければすべての属性についての確認行動を行う。以上を確認行動が成功するまで繰り返す。

比較手法 1 および比較手法 3 における情報要求および確認を行う順序は、実験結果が最小のやりとり回数を与えるような順序とした。比較手法 2 における情報要求の順序は実験結果に影響を与えないので無作為な

ただし、実験では推定結果や推定誤り設定の乱数の結果にかかわらず最低でも 0.01 の認識率はあるとした。

順序とした．これらの比較手法は，対話例や学習データを必要としない方法の典型例を集めたものといえる．文献 6) においてあげられた 4 種類の戦略のうち，間接確認を除いた 3 種類の戦略を含んでいる．文献 7) は，確定していく属性は一度に 1 つであり，本実験での比較手法は実験結果が最小となるような確定順序であるので，比較手法 1 あるいは比較手法 2 に含まれると考えられる．

提案法がタスクに依存せず動作することを検証するため，必要な属性数と属性間の依存関係の異なる 2 つのタスクで評価を行った．

タスク 1： 3 つの属性を持ち，それぞれの語彙数は 500, 300, 100 である．どの属性間にも語彙の依存関係はない．他に対話の進行に最低限必要な語彙として 200 あり，この語彙はつねに認識対象である．

タスク 2： 5 つの属性を持ち，それぞれの語彙数は 1,000, 300, 100, 100, 30 である．このうち属性 1 (1,000 語) と属性 2 (300 語) については属性に語彙の依存関係を設定した．依存関係の設定方法は 3,000 の異なる対象を属性 1 と属性 2 によって特定できるようにするために，3,000 の対象を属性 1，属性 2 とともに各値に 1 つずつ分配したうえで，残りの分配先を正規乱数で決定した．属性 1 での値と属性 2 での値が同一であれば対象は一意に決定されるよう重複を排除した．この関係は，たとえば 3,000 人の社員を 1,000 の姓と 300 の部署名によって特定するような状況を想定している．よくある姓は姓が判明しただけでは様々な部署が考えられるが，珍しい姓は部署名なしに特定することができるといった具合である．これらのほかに対話の進行に最低限必要な語彙が 200 あり，この語彙はつねに認識対象である．

各タスクについて，語彙数 500 語における認識率を 0.6 から 1.0 まで 0.01 刻みで変化させ，各 10,000 対話ずつのシミュレーションを行った．タスク 1 のやりとり回数の平均のグラフを，図 3，図 4，図 5 に示す．語彙数 500 語における認識率 0.75, 0.85, 0.95 および，グラフからは結果を読みにくい 0.61, 0.62, 0.77, 0.88 における本手法の平均の数値と，その認識率において最小の平均だった比較手法の平均の値を表 1 に示す．タスク 2 のやりとり回数の平均のグラフを，図 6，図 7，図 8 に示す．語彙数 500 語における認識率 0.65, 0.75, 0.85, 0.95 およびグラフからは結果を読みにくい 0.88, 0.89, 0.90, における本手法の平均の数値と，その認識率において最小の平均だった比較手法の平均の値を表 2 に示す．

図より，提案法は比較手法よりも，おおむねどの範

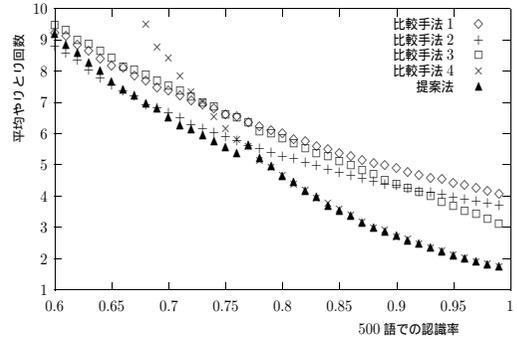


図 3 語彙数 500 語での認識率に対する確認対話終了までの平均やりとり回数 (タスク 1, 模擬ユーザー 1)

Fig.3 Average number of exchanges to complete a confirmation dialogue under 500-words vocabulary (task 1, simulated-user 1).

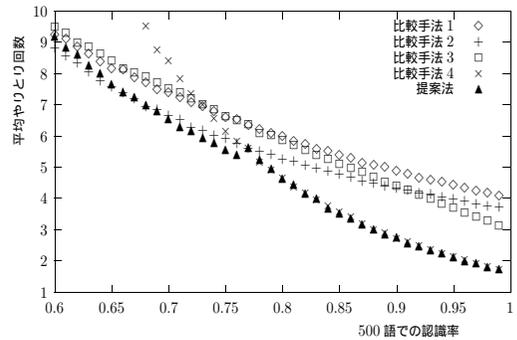


図 4 語彙数 500 語での認識率に対する確認対話終了までの平均やりとり回数 (タスク 1, 模擬ユーザー 2)

Fig.4 Average number of exchanges to complete a confirmation dialogue under 500-words vocabulary (task 1, simulated-user 2).

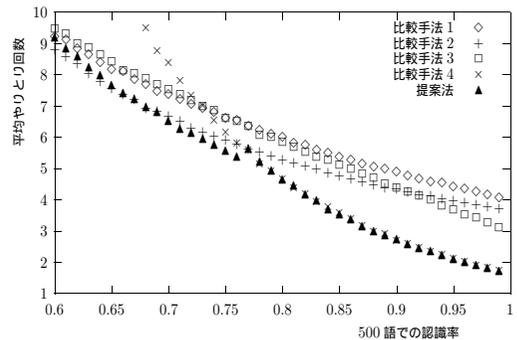


図 5 語彙数 500 語での認識率に対する確認対話終了までの平均やりとり回数 (タスク 1, 模擬ユーザー 3)

Fig.5 Average number of exchanges to complete a confirmation dialogue under 500-words vocabulary (task 1, simulated-user 3).

囲においても少ない平均やりとり回数でユーザ要求を把握できていることが見てとれる．表に示したように

表 1 語彙数 500 語での認識率に対する確認対話終了までの平均やりとり回数 (タスク 1)
 Table 1 Average number of exchanges to complete a confirmation dialogue under 500-words vocabulary (task1).

	確認手法	語彙数 500 語での認識率						
		0.61	0.62	0.75	0.77	0.78	0.85	0.95
模擬ユーザ 1	提案法	8.8	8.6	5.6	5.6	5.2	3.5	2.1
	比較手法最小	8.6	8.4	5.9	5.6	5.1	3.6	2.1
模擬ユーザ 2	提案法	7.5	7.3	5.0	5.0	4.6	3.2	2.0
	比較手法最小	7.4	7.3	5.3	4.8	4.6	3.2	2.0
模擬ユーザ 3	提案法	6.3	6.1	4.4	4.4	3.9	2.8	2.0
	比較手法最小	6.4	6.3	4.5	4.1	3.9	2.8	1.9

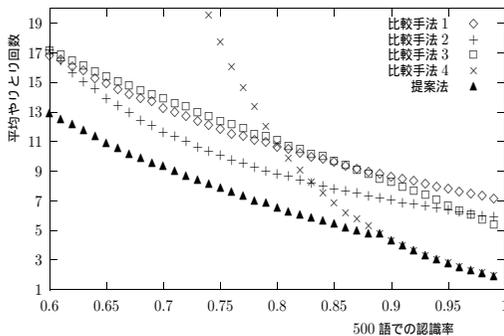


図 6 語彙数 500 語での認識率に対する確認対話終了までの平均やりとり回数 (タスク 2, 模擬ユーザ 1)

Fig. 6 Average number of exchanges to complete a confirmation dialogue under 500-words vocabulary (task 2, simulated-user 1).

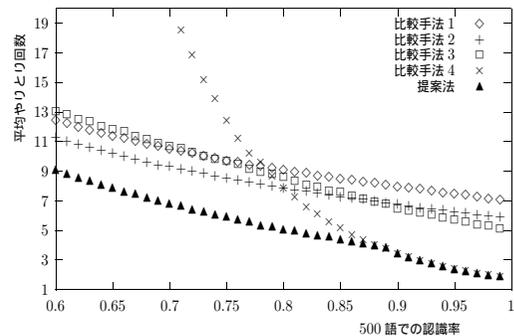


図 8 語彙数 500 語での認識率に対する確認対話終了までの平均やりとり回数 (タスク 2, 模擬ユーザ 3)

Fig. 8 Average number of exchanges to complete a confirmation dialogue under 500-words vocabulary (task 2, simulated-user 3).

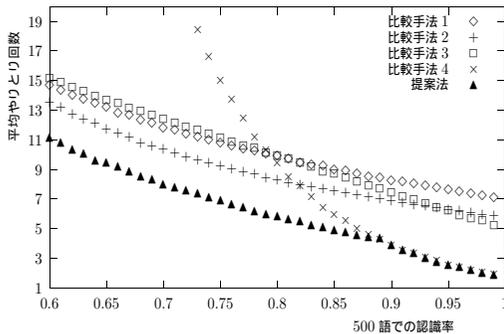


図 7 語彙数 500 語での認識率に対する確認対話終了までの平均やりとり回数 (タスク 2, 模擬ユーザ 2)

Fig. 7 Average number of exchanges to complete a confirmation dialogue under 500-words vocabulary (task 2, simulated-user 2).

一部の点において、比較手法よりも提案法のやりとり回数が必ずしも少ない場合があるが、ほぼ同等の性能である。これは、実験条件では認識率の推定は誤りがあるとしたため、システムが実際あるべき戦略に比べ、認識誤りの危険を過少評価し、たくさんの属性を一度の行動対象にするような貪欲な戦略をとってい

たり、逆に、認識誤りの危険を過大評価した結果、語彙数が少なくなるように少数の属性のみを一度の行動対象にするような控え目な戦略をとっている場合があるためであると考えられる。たとえば認識率が非常に高い範囲では、まとめて情報要求を行いまとめて確認を行う戦略 (比較手法 4) が有効であるはずだが、提案法において現実の認識結果に比べ認識率の推定が低く見積もられば、システムはより控え目な戦略をとり、未確定の属性すべてではなく一部のみを確認対象にしている場合があると考えられる。

比較手法はいずれも認識率に応じた挙動の変更を行わない。仮に本手法の仮定同様、認識器の基本性能としてシステムの認識器の特定語彙数下での認識率が与えられるとし、その値に応じて比較手法のうち最も有利な戦略を決定することも考えられる。たとえばタスク 2, 模擬ユーザ 1 の実験結果では、語彙数 500 語での認識率が 0.6 の場合には比較手法 1 を用い、0.61 から 0.84 の範囲では比較手法 2 を用い、0.85 以上では比較手法 4 を用いるような組合せを行えば最低のやりとり回数の戦略を得ることができる。しかし、このような切替点は認識器の基本性能が与えられたからと

表 2 語彙数 500 語での認識率に対する確認対話終了までの平均やりとり回数 (タスク 2)
Table 2 Average number of exchanges to complete a confirmation dialogue under
500-words vocabulary (task2).

	確認手法	語彙数 500 語での認識率						
		0.65	0.75	0.85	0.88	0.89	0.90	0.95
模擬ユーザ 1	提案法	10.9	7.9	5.5	4.8	4.9	4.3	2.8
	比較手法最小	13.9	10.1	6.9	5.3	4.8	4.3	2.8
模擬ユーザ 2	提案法	9.5	6.9	4.9	4.4	4.4	3.9	2.6
	比較手法最小	11.8	9.3	6.0	4.6	4.3	3.9	2.6
模擬ユーザ 3	提案法	7.9	5.9	4.4	4.0	3.9	3.5	2.4
	比較手法最小	10.2	8.5	5.2	4.0	3.8	3.5	2.4

いて自動的に決まるわけではなく、タスクごとに与える必要がある。これに対し本手法では戦略切替え点を事前に与える必要はない。

本手法のグラフには不連続な点が見られる。提案法では推定した認識率を用いて推定したやりとり回数に応じた確認プランを選択し、確認プラン中の最初の行動を実行する。行動の違いは、対象属性の組合せと行動の種類の違いのみで規定されるので、離散的な違いしかない。このため、推定された認識率の変化の範囲が、最も期待やりとり回数の小さい確認プランを変えない範囲では同一の行動が選択され、推定された認識率の変化の範囲が、最も期待やりとり回数が少ない確認プランを変えるような範囲にある場合に、選択する行動が代わることになる。以上が本手法において不連続な点が生じる理由であると考えられる。一方、4つの比較手法はいずれも認識率によらず同一の戦略をとるため、認識率の変化が行動に変化を与えることはなく、不連続な点を生むことはない。

3種類の模擬ユーザの結果の違いに注目すると、確認内容に誤りが含まれている場合に、より頻繁に訂正内容を伝達するほど、どの確認手法においてもより短かいやりとりの回数で対話を終えることができている。

5. おわりに

本稿では、確認の対象ごとに異なる音声認識率を考慮したうえで、効率的にユーザの要求内容を確定するための対話制御法を提案した。提案法は、確認対象を定めた場合に、値が確定するまでに必要なやりとりの回数を推定することにより、やりとり回数が最小の確認プランを選択する。属性間に依存関係がある場合の評価を、平均的な語彙の制約を用いて行うことにより、値が未定であるけれども他の属性と依存関係を持つような属性の確定に必要なやりとり回数の推定を実現した。システムと模擬ユーザとのシミュレーションによる対話実験の結果、認識率の性能（同一語彙数下での認識率）を変化させた場合でも、比較手法の中の最

のものよりも、より短いやりとり回数あるいは同等のやりとり回数でユーザ要求を確定できることを示した。

実際の音声対話システムにおいては、事前に予測した音声認識精度よりも、認識後に得られる音声認識結果の信頼度が有用な情報であると考えられる。このような音声認識結果の信頼度を用いる方法として、文献17)では音声認識結果の信頼度に応じて確認が必要かどうかの判定や誘導発話内容の決定を行う方法が提案されている。このような音声認識結果の信頼度を用いる提案法の拡張として、音声認識結果の信頼度を確率値と見なし、式(1)の初項として計算することにより、やりとり回数の推定の精度を高めることが考えられる。

謝辞 日頃よりご指導いただき、NTTコミュニケーション科学基礎研究所石井健一郎所長、メディア情報研究部村瀬洋部長、熱心に討論くださるNTTコミュニケーション科学基礎研究所マルチモーダル対話研究グループの諸氏に感謝します。

参 考 文 献

- 1) Singh, S., Litman, D., Kearns, M. and Walker, M.: Optimizing Dialogue Management with Reinforcement Learning: Experiments with the NJFun System, *Journal of Artificial Intelligence Research*, Vol.16, pp.105-133 (2002).
- 2) Stallard, D.: Talk'n Travel: A Conversational System for Air Travel Planning, *Proc. 6th Applied NLP*, pp.68-75 (2000).
- 3) Glass, J. and Hazen, T.J.: Telephone-Based Conversational Speech Recognition in the JUPITER Domain, *Proc. ICSLP* (1998).
- 4) 中野幹生, 堂坂浩二, 宮崎 昇, 平沢純一, 田本真詞, 川森雅仁, 杉山 聡, 川端 豪: TV番組の録画予約を受け付ける実時間音声対話システム, 情報処理学会研究報告 SLP-22, pp.41-42 (1998).
- 5) Niimi, Y. and Kobayashi, Y.: Dialog control strategy based on the reliability of speech recognition, *Proc. ICSLP* (1996).
- 6) Niimi, Y. and Nishimoto, T.: Mathematical

- Analysis of Dialogue control strategies, *EU-ROSPEECH*, Vol.3, pp.1403-1406 (1999).
- 7) 井本貴之, 相川清明: 平均対話回数を用いた対話設計方法, 日本音響学会講演論文集, Vol.2-Q-13, pp.165-166 (1997-3).
 - 8) Litman, D.J., Kearns, M.S. and Walker, M.A.: Automatic Optimization of Dialogue Management, *COLING* (2000).
 - 9) Boyce, S.J. and Gorin, A.L.: User Interface Issues for Natural Spoken Dialog Systems, *Proc. 1996 International Symposium on Spoken Dialog*, pp.65-68 (1996).
 - 10) Levow, G.-A.: Characterizing and Recognizing Spoken Corrections in Human-Computer Dialogue, *COLING-ACL*, pp.736-742 (1998).
 - 11) 中川聖一, 伊田政樹: 連続音声認識のタスクの複雑さを表す新しい尺度, 信学論, Vol.J81-DII, No.7, pp.1491-1500 (1998).
 - 12) Rosenfeld, R.: A Maximum Entropy Approach to Adaptive Statistical Language Modeling, *Computer, Speech and Language*, Vol.10, pp.187-228 (1996).
 - 13) Markhoul, J. and Schwartz, R.: State of the Art in Continuous Speech Recognition, *Voice Communication between Human and Machines*, National Academy of Sciences, pp.165-198 (1994).
 - 14) Eckert, W., Levin, E. and Pieraccini, R.: Automatic evaluation of spoken dialogue systems, *TWLT13: Formal semantics and pragmatics of dialogue* (1998).
 - 15) Watanabe, T., Araki, M. and Doshita, S.: Evaluating Dialogue Strategies under Communication Errors Using Compute r-to-Computer Simulation, *Trans. Institute of Electronics and Communication Engineers of Japan*, Vol.E81-D, No.9, pp.1025-1033 (1998).
 - 16) Hirasawa, J., Miyazaki, N., Nakano, M. and Aikawa, K.: New feature parameters for detecting misunderstandings in a spoken dialogue system, *Proc. ICSLP*, pp.154-157 (2000).
 - 17) 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, pp.3078-3086

(2002).

(平成 15 年 6 月 2 日受付)

(平成 16 年 7 月 1 日採録)



安田 宜仁

1997 年京都大学総合人間学部基礎科学科卒業。1999 年同大学大学院人間・環境学研究科修士課程修了。同年日本電信電話(株)入社。現在, NTT コミュニケーション科学基礎研究所勤務。音声対話システムの研究に従事。日本音響学会会員。



堂坂 浩二 (正会員)

1984 年大阪大学基礎工学部情報工学科卒業。1986 年同大学大学院博士前期課程修了。同年日本電信電話(株)入社。現在, NTT コミュニケーション科学基礎研究所勤務。音声対話システム, 言語生成, 文脈理解の研究に従事。博士(情報科学)。情報処理学会平成 9 年度論文賞受賞。ACL, 電子情報通信学会, 人工知能学会, ソフトウェア科学会各会員。



相川 清明 (正会員)

1975 年東京大学工学部電子工学科卒業。1980 年同大学大学院博士課程修了。同年日本電信電話公社武蔵野電気通信研究所入所。以来, 連続音声認識, 聴覚モデル, ニューラルネット, 音声対話システムの研究に従事。1989 年米国カーネギーメロン大学客員研究員。1992 年~1995 年国際電気通信基礎技術研究所主任研究員。2003 年東京工科大学メディア学部教授。現在に至る。工学博士。IEEE, ASA, 日本音響学会, 電子情報通信学会各会員。