

効率的な音声対話制御方式に関する一考案

大森 久美子

NTT 情報流通プラットフォーム研究所
kumiko@isl.ntt.co.jp

東田 正信

(株) 国際電気通信基礎技術研究所
higasida@ctr.atr.co.jp

現行の音声対話インタフェースに対するユーザ不満を解消するため、現状の音声認識の技術的限界やマシンパワーを考慮した上で、利用者意図の確定を目的とした実用レベルの音声対話制御方式を考案した。

考案方式は、人間同士の対話における相手の発話内容を聞き取る仕組みの分析を通して、発話語の使用頻度等の偏りを利用して、事前に認識対象範囲を予測し絞込む方式（“思い込み認識”）、事前予測が外れた場合に、利用者に気付かれないように自然に対話を進める方式（“聞いているふり認識”）という2つの対話戦略を併用することで、対話の自然性を保持した上で、大多数の候補から効率良く利用者の意図を確定するという特徴を有する。

考案方式の有効性検証のために、日本全国の住所確定をタスクとしたプロトタイプシステムを試作し、性能、ユーザ満足度等の評価を行った。その結果、考案方式は現行方式と比べ、性能面だけでなく利用者満足度も大幅に向上することが確認できた。

New Dialogue Control Method that Efficiently Ascertains Customer Intent through Speech Recognition

Kumiko Ohmori

NTT Information Sharing Platform Labs
kumiko@isl.ntt.co.jp

Masanobu Higashida

ATR International
higasida@ctr.atr.co.jp

This paper describes a new dialogue control method that utilizes speech recognition processes we propose based on human dialogue analysis. This method provides users with stress-free voice input through real-time responses and a naturally-controlled dialogue.

We found that dialogue control strategies derived from human operators who skillfully obtain the user intent, even when faced with unfamiliar words or an inability to hear clearly, can be adopted to solve the problems of speech recognition technology.

When a listener only prepares a small number of words to recognize a word spoken by a partner, “Presupposition-type Listening (PSL)” is initiated because the listener believes the spoken word will most likely fall into the prepared word set selected by preference or familiarity. When a listener feels uncertainty in the results, he usually asks the partner to repeat the word or queries the partner on related information to support his results. When the listener experiences failure in this process, he extends the word set to handle a large number of words and gains time to advance through the listening for larger words or winnow the extended word set by querying the partner for some additional information, while maintaining a smooth discourse. These processes are called “Pretense-type Listening (PTL)” because the listener employs these while not letting the partner know that he could not fully hear what was said.

We developed a prototype that simulates operators who efficiently identify a series of address segments. The experimental results show that the dialogue time and number of interactions required to obtain clearly the user's intent is reduced to roughly 2/3 compared to existing systems. The customer satisfaction level was also raised 184% on average, based on five-grade user satisfaction questionnaires. The important aspect of the results is not the improvements in function and performance, but the shift in user reaction from a dislike based on the system's limitations to excitement over the variety of new functions and possibilities.

1. はじめに

近年、音声認識技術の発達により、この技術をマンマシンインタフェースとして利用する音声対話制御方式の研究が盛んに行われている[3][4][5][9].

音声による入力、利用者にとっては、キー入力と比べ入力速度が速く、簡単であるという利点、サービス提供者にとっては、オペレータの人的削減、24時間サービス提供可能等、様々な効果を持つ。

実用サービスのフロントエンドとして音声を利用するためには、利用者の意図を確実に確定することが大前提となる。しかし、音声認識技術を利用して応用システムを構築するには現状、様々な制約条件がある。例えば、システム主導型、かつ一問一答形式の孤立単語認識においても、対象語彙数が増えれば増えるほど誤認識は避けられず、処理時間もかかる。このため、音声入力インタフェースは、まわりくどい、聞き間違えが多い、使い勝手が悪い等、利用者満足が得られておらず、ごく限られた分野でしか応用されていない[1][6].

我々は、現行音声認識技術の認識性能及び認識精度上の制約を考慮した上で、利用者の意図確定を目的とした新しい音声対話制御方式を考案した。考案対話制御方式は、利用者との対話を通して認識結果の曖昧性を解消し、短時間で効率良く利用者の意図を確定可能、かつ利用者満足度を獲得可能な方式であることを、試作を通して確認した。

2. 現行の利用者意図獲得方式

2.1 音声認識技術

例えば、住所や姓名の確定をタスクとするカスタマケアやチケット販売等の実用サービスにおいて、音声を入力インタフェースとして適用する場合を考える。実用レベルでは、何十万語もある住所を、都道府県、市区町村等、どのレベルから入力されても認識可能できる、何十万通りもある姓名を一度聞いただけで正しく認識できることが好ましい。しかし、現状の音声認識技術を利用してシステム構築する際、様々な困難が予想され、一般的には、以下の2点に留意する必要がある[2][7][8].

- 対話の自然性を保つことが可能な時間内で、認識可能な語彙数には限界がある。
- 認識精度は、発話場所の雑音等の使用環境や、発話者の年齢や話し方等の入力音声の状態に大きく依存し、常に認識装置の最善の処理性能、精度が得られる保証はない。

このため、名詞や動詞のような単語単位の認識に着目した場合でも、対話の自然性を保つ適切な時間内に処理を完了し、かつ認識精度をある程度保証しようとする、認識対象語彙数を減らさざるを得ない。そのため、大語彙を認識対象としなければならない実用サービスにおいて音声入力インタフェースを適用する場合、効率良く利用者の意図する情報を確定するためには、現状、以下のような方策を考えざるを得ない。

2.2 現行対話制御方式

現行インタフェースにおける利用者の意図確定を

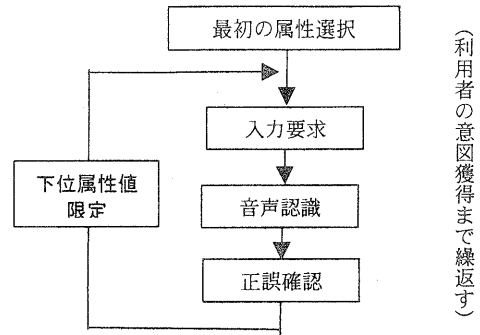


図1：現行インタフェースの対話制御方式

目的とした対話制御方式を図1に示す。

現行インタフェースは、一般的にシステムにとって都合の良い対話制御を行う。

- (1) 利用者の意図確定に必要な属性を階層化することで大語彙を分類し、各階層における語彙数を減少させると同時に、各ノード下の語彙数の減少を図る。上位層から順番に入力を要求し、[確定・下位層の絞込み]を繰返すことにより、利用者の意図を確定する。
- (2) すべてのレベルにおける認識結果に対して、正解が確認できるまで、利用者に対して、認識結果の確度の高い順に、提示し正誤確認を繰返す。各属性を確実に確定することにより、次の対象である下位層の認識対象語彙数を減少させる。

このために、現行インタフェースは、

- (1) 利用者の意図確定のために、属性の階層数分 [入力要求・正誤確認の繰返し] ターンを必要とする。
- (2) 確定するまで次の入力へ進むことができない、という問題点があり、この2点が利用者に対して、まわりくどさや聞き間違えの多さ等、不快感を抱かせる原因となり、「音声認識は実用レベルに達していない」という評価につながっていると考えられる。

3. 提案対話制御方式

3.1 人間の聞き取りモデル

人間同士の対話においても、相手の発話状況によって聞き取りにくいこと、あるいは非常に類似した語彙と聞き違えるという現象があるように、利用者とシステム間の音声によるやりとりにおいても、同様の現象が想定されても不思議ではない。通常、このような不具合に対して人間は、聞き取った情報を確定するために、聞き返したり、関連する様々な傍証情報を獲得する方向へ対話を進めたり、という対話によるインタラクションによって認識精度の曖昧性をカバーするという戦略をとっている。

そこで我々は、テレフォンサービスにおける利用者とオペレータの対話を分析した。人間同士の対話において相手の発話を聞き取る際、以下の2つの認識プロセスが組合わされていることを発見した。我々

は、これらのプロセスをそれぞれの性質より、“思い込み認識”と“聞いているふり認識”と命名した。

(1) 思い込み認識

オペレータは、利用者が発話する内容が広範囲に渡っている場合でも、日常生活習慣や話者の特徴等から発話内容を事前予測し、対象を自らで絞込むことにより利用者の発話を聞き取る準備（心構え）をしている。

(2) 聞いているふり認識

オペレータは、“思い込み認識”による事前予測が外れたと分かった場合、利用者にその事実を知らせずに、さりげなく聞き返したり、聞かれても利用者が不自然に感じない他の関連する事項を尋ねる。得られた複数の情報の関連性から総合的に利用者の意図を絞込み、確定を試みる。

例えば、住所を確定するプロセスを考える。オペレータは、利用者に対して「市名」の入力を要求したが、完全に聞き取れなかった場合、「その市は何県にありますか？」或いは「さらに詳しい地名がわかりますか？」等、尋ねることで、オペレータは、得られた県名、或いはその下の字名を、市名の認識結果と組合わせ市名確定を試みる。また、我々の日常生活における電話応対等で、相手の氏名を聞き損ねた場合は、「御社のお名前を頂戴できますか？」或いは「先日お越し頂きましたのは何日でしょうか？」等の質問をすることで、氏名に対する認識結果に対して、相手の会社名や訪問日を合わせた総合的判断を行うことにより氏名確定を試みる場合が多い。

そこで、オペレータが利用者の発話を聞き取った直後の状態を、以下の3通りに分類できると考えた。

(レベル1) 正しく聞き取れた（予測した語彙が発話された）と判断。

(レベル2) やや曖昧だが何となくは聞き取れた（予測した語彙であると思われるが確信がもてない）と判断。

(レベル3)：聞き損ねた（聞きなれない予測外の語彙が発話された）と判断。

例えば、氏名の聞き取りにおいて、“サトウ”や“マツモト”のように、聞いたことがあり容易に想像できる氏名に対しては、事前予測が的中することが多い。従って上述のレベル1の判断が可能である。一方、“アシダ”と“ハシダ”のような非常に類似した2語が認識対象の中に入っている場合は、レベル2と判断することが多いようである。“ヨスケ”や“ヒガシダモリ”のような聞いたことがない珍しい氏名に対しては、事前予測したカテゴリーの中に思い当たった発話語が存在しないことが多く、事前予測が外れたというレベル3の判断になる。

3.2 現行方式との比較

利用者とオペレータ間の対話における情報確定過程を考察するために、利用者の意図確定の過程における情報エントロピーの減少過程に焦点をあてる。

例として、日本全国の住所において大字を確定する場合を考える。住所は、都道府県（語彙数 47）、市区町村（語彙数 4,100）、大字（語彙数 18万）の順に

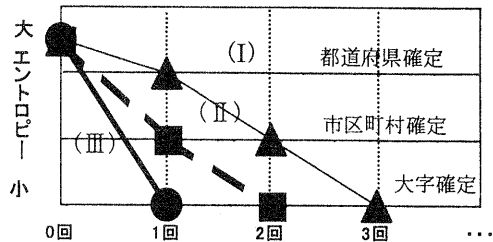


図2：情報エントロピー減少過程

階層化される。現行方式の情報確定プロセスにおける情報エントロピー減少過程を、図1の(I)に示した。対話開始時、情報エントロピーは17.5bitである。現行方式では、都道府県確定により情報エントロピーは11.9bit、市区町村確定により5.5bitと順に情報確定まで減少していく。住所が大字レベルまで確定した時点で情報エントロピーは0となる。現行方式の情報エントロピー減少過程は緩やかであり、利用者不満の原因となっている。我々は、図2における(II)、(III)のような、より急速に情報エントロピーを減少させることが可能な対話制御の実現を目指した。すなわち、住所確定をタスクとする場合は、最初から大字、あるいは市区町村レベルからの入力を可能にすることで、より早く、効率良く情報エントロピーの減少を狙うプロセスの考案に努めた。

4. 提案対話制御方式

我々は前述の対話手法をコンピュータ上で動作可能なシステムに展開するために、3.1節で述べた“思い込み認識”、“聞いているふり認識”という2つの認識プロセスを併用することにより、聞き手が話者の発話を聞き取る過程をモデル化した。提案モデルを図3に示す。

4.1 思い込み認識のモデル化

人間同士の対話での、聞き手が話者の発話を聞き取る過程において、聞き手はあらかじめ発話内容を予測している、という特徴を、実時間内に認識処理を行うことが不可能な大語彙を認識対象とする場合に置き換え“思い込み認識”をモデル化した。対話の中で使用される単語の使用頻度の偏りを利用して、利用者発話の集中する属性値を事前予測し、優先的に認識対象語彙（以下、「思い込み対象語」と呼ぶ）とすることで、仮想的に認識対象語彙数の減少を図る。“思い込み認識”の適用により、多くの属性値を持つ下位の属性に対しても、実時間の応答が可能になる。また、下位の属性に属する語彙（属性値）であっても、発話頻度の高い語彙は、思い込み対象語に指定しておくことにより、下位属性からの入力要求も可能になる。

音声認識エンジンを使用する対話においては、認識対象語彙を限定しなければならないことから、“思い込み認識”のプロセスをうまくあてはめることが重要になる。そこでまず始めに、音声認識の結果、算出される認識尤度の相対的な関係を利用して、音声認識結果を以下のようにモデル化した。

認識レベル 1：第 1 位の尤度 \geq 規定閾値 a

利用者は思い込み対象語を発話したと判断。

認識レベル 2：第 n 位までの尤度和 \geq 規定閾値 b

利用者発話を思い込み対象語の上位 n 個に絞込むことができた と判断。

認識レベル 3：その他の場合

(a, b は認識エンジンに依存して設定する値)

4.2 聞いているふり認識のモデル化

“思い込み認識”による事前予測が外れた場合、聞き手は、関連する情報を話者から聞き出すことにより、対話の軌道修正を図る、という特徴を利用して“聞いているふり認識”をモデル化した。事前予測が外れた場合、利用者にその事実を知らせずに、自然に関連する別属性を尋ねる方向へ対話を誘導する。得られた複数の属性に間のつながりを利用して、利用者発話を総合的に絞込む。“聞いているふり認識”の適用により、事前予測した思い込み対象語に含まれない語彙を発話した利用者に対しても、属性毎の正誤確認を行わずに、意図確定が可能になる。“思い込み認識”と“聞いているふり認識”を併用することで、対話全体の正誤確認回数の減少、かつ利用者の意図確定までの対話時間の短縮が見込めることから、利用者満足度向上が期待できる。

4.3 対話制御プロセス

“思い込み認識”、“聞いているふり認識”を効率良く組み合わせ、対話制御プロセスを作成することにより、利用者の意図確定を行う手法について説明する。

プロセス 1：思い込み当り

4.1 節の認識レベル 1 に相当する。“思い込み認識”を実施する。その結果、思い込みが的中したと判断できる場合、認識結果第 1 位候補を、提示し正誤確認を行う。利用者から否定された場合は、“聞いている

ふり認識”へ移行する。

プロセス 2：思い込みたぶん当り

4.1 節の認識レベル 2 に相当する。“思い込み認識”の結果、上位に正解らしき候補が n 候補得られたが、正解を特定するのに迷いがある場合、各候補の尤度差の関係に応じてプロセスを使い分ける。ここでは n = 2 の場合を例に挙げる。2 候補の尤度差が大きい場合、発話語候補を直接提示するのではなく、有力候補が持つ他の関連する属性値をシステム側から提示し、利用者に確認する。例えば、住所の確定を行う場合、利用者の発話“台東区”に対して、認識の結果“台東区”と“江東区”が有力 2 候補として出力され、かつ両候補の尤度差が大きい場合、「それは上野があるところですよ？」のように、上位 2 候補を識別するのに有効と考えられる、その下の字名をシステム側から提示するプロセスをとる。これは、おそらく“台東区”が正解であると考えられるが確信が持てないので再確認を行う。一方、有力 2 候補の尤度差が小さい場合は、2 候補の識別に有効な関連属性値の入力を要求し絞込む。上記例において、認識の結果、東京都の“台東区”と大阪府の“大正区”が非常に近差で出力された場合、「それは何県にありますか？」のように、“台東区”と“大正区”を識別するのに有効と思われる都道府県名を関連情報として利用者に尋ねる。これは、どちらも同程度の正解の可能性を秘めていることから、都道府県名を情報として得ることで、絞込みを行うプロセスである。

プロセス 3：思い込み外れ

“聞いているふり認識”による関連属性に関する質問を利用者に行う。この対話に要する処理時間を利用して、思い込み対象語に含まれない候補に対する認識処理を並行して実施する。この対話による応対時間は、少なくとも数秒程度は見込める事から、

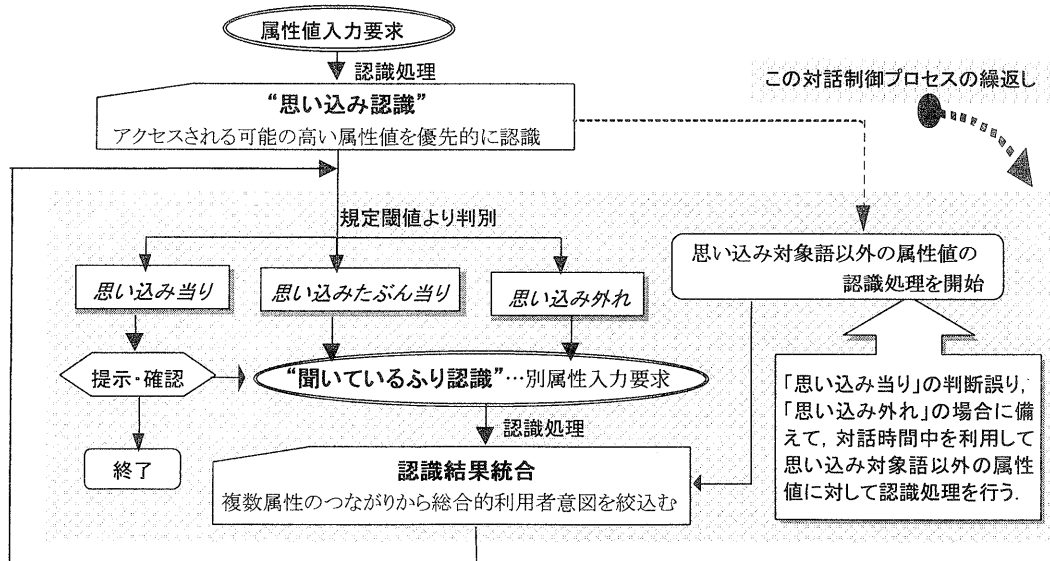


図 3：提案対話制御プロセス

思い込み対象語の数倍から数十倍程度の語彙数に対する処理が可能である。この結果と、取得した別属性候補の認識結果とを相互の接続性（都道府県と市区町村が上下関係にあるか、聞き取った姓と名の組合せを持つ人が想定しているデータベースに実在するかどうか、等）を利用して総合的に候補を絞込む。

“聞いているふり認識”の結果に対しても、規定閾値と認識尤度を比較することにより、上記 3 モデルを適用し、必要に応じてさらに別属性の入力要求、総合的判断を繰返すことで、利用者の意図の確定を行う。

5. プロトタイプシステム

提案対話制御方式の有効性を検証するために、日本全国の住所を市区町村のレベル（対象語彙数 4,100）まで確定することをタスクとしたプロトタイプシステムを PC 上に試作した。例えば、東京都港区、北海道札幌市中央区、神奈川県三浦郡葉山町等のように都道府県を最上位階層に設定し、市、区、町、村（2 階層、政令指定都市のように 3 階層目が区、村の場合は 3 階層）のレベルまでの確定を行う。利用者は、都道府県名からではなく、第 2、あるいは第 3 階層に属する市区町村名から最初に入力することができる。プロトタイプシステムにおける属性は以下のように階層化される（同音異義語は 1 語彙とカウントする）。

- ・階層 1: 都道府県 (対象語彙数 47)
- ・階層 2: 市区町村 (対象語彙数 4,100)
- ・階層 3: 区町村に続く大字 (対象語彙数約 18 万)

提案方式との比較を行うために、現行方式についてもプロトタイプシステムを作成した。現行方式は、図 1 に示したフローに従い、都道府県、市区町村の順に確定し、各々確定するまで正誤確認を繰返す方式である。

6. 評価試験及び結果

表 1: 評価試験結果 (40 名×20 件)

	現行方式 800件	提案方式	
		思い込み対象 400件	思い込み 対象外 400件
平均確定所要時間 (秒)	54.0	26.9	40.8
平均被験者発話数 (回)	5.04	2.47	3.29
平均正誤確認回数 (回)	2.59	1.20	1.28
平均 5 段階評価	2.38	4.38	

試作プロトタイプシステムを用いて評価試験を行った。試験では、音声、及び対話処理の研究に携わった経験のない男女 40 名（首都圏出身、10 代～40 代）を被験者とした。各被験者に対して、現行方式、提案方式の各方式を用いて、思い込み対象語の中から選択した 10 市区町村、思い込み対象語に含まれない中から選択した 10 市区町村の計 20 市区町村を確定する対話、計 40 件（20 件×2 方式）を実行した。

6.1 評価試験集計結果

現行、提案両方式とも、タスク達成率は 100% であ

り、指定した市区町村はすべて確定することができた。各対話に対し、市区町村確定までの所要時間、被験者の発話回数、正誤確認回数を測定したところ表 1 のような結果になった。

被験者に対して、現行方式と提案方式に対する満足度を 5 段階で評価してもらった（1: 不満 2: やや不満 3: 普通 4: やや満足 5: 満足）。

6.2 評価試験結果分析

評価試験結果より以下のようなことが分かる。

(1) 平均確定所要時間と被験者発話回数

提案方式において、思い込み対象語に含まれる市区町村の平均確定所要時間は、現行方式に比べ約 1/2 に短縮、思い込み対象に含まれない市区町村に関しては、約 3/4 に短縮した。被験者発話回数においてもほぼ同様の結果が得られている。提案方式は、利用者発話の偏りが大きければ大きいほど有効であり、実際にアクセス状況として、利用者発話に 8 割の偏りがあると仮定すると、確定までの所要時間、及びターン数とも約 2/3 に短縮できることが評価試験結果より推定できる。

(2) 被験者満足度

現行、提案両方式に対する被験者満足度を調査した結果を図 4 に示す。現行方式に対する満足度は平均 2.38 と“やや不満”レベルであったのに対して、提案方式に対する被験者満足度は“やや満足”を上回る 4.38 へと向上した。

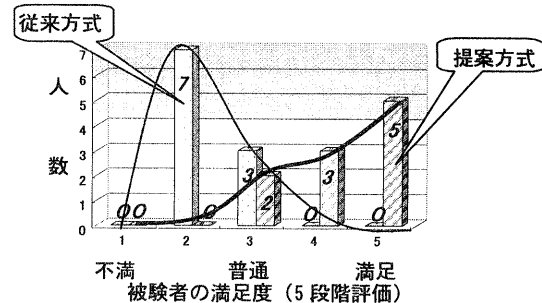


図 4: 評価試験における被験者満足

(3) 被験者満足度向上の要因分析

現行、及び提案方式の平均確定所要時間について比較した結果を図 5 に示す。図 5 より、全被験者とも提案方式は現行方式と比較して、対話処理時間が平均で 37.3% 短縮（平均被験者発話回数も 43.4% 減少）しており、その結果、被験者満足度が 1.8 倍向上したことがわかる。これは被験者へのアンケートにおける以下のコメントの数値的な反映と考えられる。

(アンケート) 提案方式の良いところは？

(括弧内は人数を表す)

- ・ 確定したいものから入力できる (32)
- ・ 確定までの時間が早い (23)
- ・ いちいち確認されない (30)

(4) 思い込み認識、聞いているふり認識の効果

提案方式において、4.3 節のプロセス 1 の場合は、

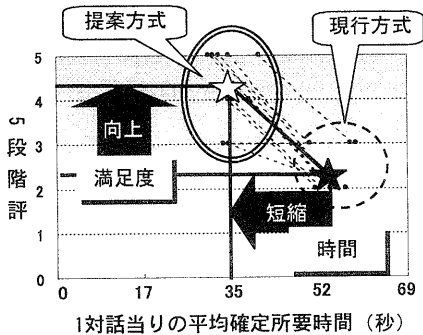


図 5：被験者毎の確定所要時間と満足度

1 回の [入力要求・正誤確認] ターンでの確定終了が実現する。思い込み対象語に含まれている市区町村を確定対象とした 100 件の評価試験対話において、1 回のターンで確定終了した対話は 88 件存在した。これに対して、現行方式は、都道府県、市区町村各々に対して順に確定を行うため、確定までに最低でも [入力要求・正誤確認] ターンを 2 回必要とする。

また、“聞いているふり認識”は、正誤確認せずに先へ進むことから現行方式に比べ、正誤確認回数の減少が期待できる。評価試験において現行方式と提案方式の平均正誤確認回数を比較した。現行方式は都道府県、市区町村各々に対して確定できるまで正誤確認を繰り返すことから、正誤確認回数は平均 2.59 回 (計 200 件の対話の平均) であった。これに対して提案方式は、1.29 回とほぼ半減した。提案方式は、4.3 節のプロセス 1 において、第 1 位が正解である場合、及びプロセス 2, 3 において“聞いているふり認識”による関連情報取得、統合した結果、正解が 1 位に上がってきた場合、正誤確認回数は 1 回、プロセス 1 による提示が利用者から否定された場合、及び“聞いているふり認識”による統合結果による提示が利用者から否定された場合に限り、2 回以上の正誤確認の必要が生じる。

これらの結果より、“思い込み認識”、“聞いているふり認識”の適用により、確定までの所要時間、及び被験者発話回数が減少したことが分かる。

7. 結論

音声認識機能を持つ利用者の意図確定を目的としたインタフェースにおいて、我々が提案した“思い込み認識”、“聞いているふり認識”を併用した対話制御方式の適用は、意図確定までの所要時間、利用者の発話回数ともに、現行方式に比べ、大幅に減少することが明らかになった。この結果、オピニオン評価 (図 4) により、これまでのインタフェースは「使えない」という評価がなされていたのに対して、我々の対話方式を適用したインタフェースは「使える」という評価を得ることができ、実用システム構築のための有力な手段であることを示すことができた。また、実際の CTI サービスへの適用を考えた場合、電話回線を介したサービス提供が一般的と考えられ

る。現在、住所確定プロトタイプシステムに改良を加え、電話帯域音声の入力を受付可能なシステムを構築し、同様の予備評価試験を行っている。電話帯域音声はマイク音声と比較して、精度劣化を避けることができないが、予備試験結果より、電話帯域音声に対しても、提案方式による対話制御により効率良く入力情報を確定可能であるという見通しが立っている。従って、電話帯域音声利用下での実用サービス提供を念頭におき、より頑健な対話制御方式確立を目指していく予定である。

謝辞

本研究の機会を与えて下さいました情報セキュリティプロジェクト、松本隆明プロジェクトマネージャー、セキュリティプラットフォームグループ、関根純前グループリーダー、本提案方式に関して様々なアドバイスを下さいました堂坂浩二氏、中野有紀子さん、水澤紀子さんに感謝致します。

References

- [1] 嵯峨山茂樹。「なぜ音声認識は使われないか・どうすれば使われるか？」情報処理学会 94-SLP-1, pp. 23-30, 1994.
- [2] Christopher S. “Voice Communication With Computers,” Van Nostrand Reinhold, A Division of Wadsworth, Inc. 1993.
- [3] Goodien D., Brill E., Glass J., Pao C., Phillips M., Polifroni J., Sneff S. and Zue V. “GALAXY: A human-language interface to on-line travel information,” Proc. ISCLP '94, pp. 707-710, 1994.
- [4] Zue V., Sneff S., Polifroni J., Phillips M., Pao C., Goodine D., Goddeau D. and Glass J. “PEGASUS: A spoken Dialogue interface for on-line air travel planning,” Speech Commun. 15, pp. 331-340, 1994.
- [5] Kuroiwa S., Tadedo K., Naito M., Inoue N. and Yamamoto S. “Error analysis of field trial results of a spoken dialogue system for telecommunications application,” IEICE Trans. E78-D(6), pp. 636-641, 1995.
- [6] Niimi Y. and Kobayashi Y. “Modeling dialogue control strategies to relieve speech recognition errors,” Proc. Eurospeech '95, pp. 534-537, 1995.
- [7] Schneiderman B. 「ユーザインタフェースの設計 (第 2 版)」日経BP出版, 1995
- [8] Zue V. “Conversational interfaces: advances and challenges,” Proc. Eurospeech, KN-9-KN-18, 1997
- [9] Niimi Y., Nishimoto T. and Kobayashi Y. “Analysis of interactive strategy to recover from misrecognition of utterances including multiple information items,” Proc. Eurospeech '97, pp. 2251-2254, 1997.