大語彙を対象とした音声対話制御手法に関する研究

平成 16 年度

大森 久美子

もくじ

第 1 章
序論

1.1　はじめに

近年，音声認識技術や言語処理技術，コンピュータ性能の向上によって，音声対話システムへの期待が急速に高まっている[10，14，17，18，42，51，58，65，66，68，69，70]．音声対話システムとは，利用者との音声による対話を通して，利用者の要求内容を決定し，情報検索をはじめとする各種タスクを実行するシステムのことである．音声対話システム実現のための対話制御手法も数多く発表されている[37，40，43，44，50，52，53，54，55，56，57，63]．

音声は，人間同士の通常のコミュニケーション形態の 1 つであり，操作に習熟を必要としないため，誰もが簡単に利用可能な入力手段として期待されてきた．また，入力速度はキーボードと比べて 3〜4 倍，手書き文字と比べて 8〜10 倍も速いと言われる[15，16]．さらに，他の器官を同時に使って並行作業が可能であるという利点を有する．また，各種サービスにおいて，音声対話システムは，運営コストの削減，24 時間，365 日のサービス提供を可能にすることから，入力インタフェースとしての需要も高い[62]．

しかしながら，音声対話システムの実用化はほとんど進んでいない[42]．実用を指向した高度な音声対話システムを実現するためには，音声認識の性能向上という大きな課題がある．音声認識の性能は，利用者の入力が単語であるか文章であるかなど，対象とする発話様式によって大きな影響を受けると言われる．システムが受理できる語彙や文法の範囲を広げるほど，誤認識は起こりやすくなる[30，33]．最も単純な特定話者を対象とした単語認識でも，対象語数が増えるほど誤認識は増大し，利用者の途中放棄によってタスクが達成できずに終わる対話も多い．

実用サービスの入力インタフェースに音声認識を適用するためには，不特定多数の利用者の入力に対して，迅速かつ正確に応答する必要がある．そのためには，認識性能の向上も必要であるが，それ以上に誤認識を修正するための対話制御，及びさらなる誤認識を防ぐための対話戦略が重要である．

1.2　本研究の目的

本研究の目的は，不特定多数の利用者が入力する住所や姓名などの大語彙を対象とした音声対話システムの実現である．現状の音声対話システムでは，利用者の発話を 100%の精度で認識できないため，認識結果の正誤は，発話者本人へ確認しなければ判断できない．そして，誤認識の場合は，正解を導き出すために，利用者に対して再入力を要求し，認識，提示確認というプロセスを繰返す．認識対象が大規模になるほど誤認識は起こりやすく，修正のためのプロセスの繰返しは利用者にとって大きな負担となる．これが，音声対話システムが実用に至らない理由である．

本研究では，現状の音声対話システムと人間同士の対話の相違を分析し，システムと利用者との間に人間と同様の対話を実現するための対話制御手法を提案する．音声対話は，目的によって様々な形態が考えられるが，本研究では，システムと利用者が単なるおしゃべりをするのではなく，利用者の入力の確定を目的とした目的指向型対話に焦点を当てる．

提案対話制御手法は，思い込み応答と属性の有効度を利用した対話を特徴とする．提案手法を音声対話システムに適用することによって，利用者満足度が高いシステムの実現を目指す．

1.3　本論文の構成

本論文の構成を以下に述べる．第 2 章では，音声認識技術の現状と音声対話システムの研究事例について紹介し，利用者にとって使い勝手のよい音声対話システムについて考える．第 3 章では，利用者の入力を単語に制限したシステム主導の音声対話システムに焦点を当て，大語彙を対象としたシステムの課題，及び本研究において解決すべき課題について論じる．第 4 章では，人間同士の対話の分析を通して，思い込み応答を提案する．思い込み応答とは，人間同士の対話と同様の聞き取り傾向を実現するための応答戦略である．第 5 章では，大語彙に対する絞り込み効果を属性の有効度として定義し，属性の有効度を利用した誤認識修正のための対話制御手法を提案する．第 6 章では，思い込み応答と有効度が大きい属性を尋ねる対話を利用した大語彙確定のための対話制御手法を提案する．市販の音声認識エンジンを使用して，個人姓 87,944 種を対象とした音声対話システムを実装し，評価実験を通して，提案対話制御手法の有効性を検証する．第 7 章では，個人姓以外の対象への

適用例として，177,747 種の住所を対象とした音声対話システムについて述べる．
第 8 章では，結論，及び今後の課題について論じる．

第 2 章
音声認識技術と音声対話システム

今日，ネットワークやマルチメディア技術が発展して，コンピュータが身近なものになりつつある．こうした中で，誰でも簡単にコンピュータやネットワーク上の情報へアクセスできるようにするための手段として，ヒューマンインタフェースへの期待は非常に大きい．音声によるインタフェースもその 1 つである．音声認識はその核になる技術である．本章では，音声認識技術の現状と課題について述べ，音声認識を利用した対話システムの研究動向を紹介する．そして，実用的な音声対話システムについて考える．

2.1　音声認識技術の現状

2.1.1　音声認識技術の形態

音声認識とは，音声波に含まれる意味内容に関する情報を電気回路やコンピュータによって抽出し，判定することである[15，16]．音声認識の手段としては，予め辞書に登録した音素や単語などを単位とする標準パターンと入力音声とを比較し，最も類似している標準パターンを選択して，そのカテゴリーの音素，単語などが発声されたと判定する場合が多い．

音声認識の形態は，誰の音声でも認識できる不特定話者型と，学習を行った特定の話者の音声のみを認識する特定話者型に分けることができる．人間は同じ単語でも個人によって少しずつ異なった発声をするため，不特定話者型認識は特定話者型認識よりも難しい．さらに，音声認識の形態は，区切って発声した単語を認識する孤立単語音声認識と，単語の連続を認識する連続音声認識の 2 通りに分けることができる．連続音声認識は，比較的少数の語彙を対象として，主として音響的特徴を用いて認識する連続単語音声認識と，比較的多数の語彙を対象として，言語的知識を用いてその意味内容を理解しようとする文音声認識に分けることができる．孤立

単語音声認識は，単語の境界を検出するために各単語間にポーズを必要とする．これに対して，連続音声認識では，ポーズを入れることなく発声された単語の連続や文章が認識できる．すべての形態において，データベースに登録されている単語や音素の組合せで構成できない発話は，未知語と判断され誤認識となる．

HMM (Hidden Markov Model： 隠れマルコフモデル) の導入によって, 人間の制約のない自由な発話を認識する方向へ研究が進められているが，文音声認識は対象を特定話者に限定しても難しい．日常で使われているような自然発話の認識は，助詞落ちや倒置，言い淀みや言い直しなどが多く含まれるため非常に困難であり，認識対象語彙や文法に多くの制約を加えない限り，実現は不可能である[9，45]．

また，連続単語音声認識に関しても，隣り合う単語に依存して単語の発音が変化してしまうことや，滑らかに発声された音声から単語の境界が検出しづらいこと，1発声中に含まれる認識誤りは単語数に応じて増加することなどが原因となり，認識対象語彙，及び連続する単語数を小規模に制限しなければ，利用者に対して高精度な結果の提供は困難であることが知られている．

次節では，市販の音声認識エンジンに対する性能評価を通して，音声認識技術の現状について説明する．

2.1.2 市販の音声認識エンジン

2.1.1 節で述べたように，音声認識技術は，コンピュータなどのハードウェアの処理能力の向上，及び統計的認識手法の導入によって，文音声認識が可能なレベルに達した．日本語の音声認識エンジンや認識装置が相次いで市場に投入され，それらを用いた電話による情報案内サービスを導入する企業も増えている[21, 22, 23, 24, 26，28，29，62]．しかし，音声認識技術は急速に進歩したが，利用者の自発的な発話を 100%の精度で正しく認識できるまでには至っていない．対象話者，及び受理可能な語彙や文法の範囲を広げるほど，誤認識は起こりやすくなる．

表 2.1 に，現在，製品化されている主な音声認識エンジンについて，不特定多数の話者を対象とした場合の一度に認識可能な最大語彙数を示す[62]. 一度に認識可能な最大語彙数とは，孤立単語音声認識のみではなく，それらの組合せから構成される連続単語音声認識，及び規定した文法に基づいた文音声認識において，認識対象として一度に受理可能な単語数を意味する．

表 2.1：市販の音声認識エンジン

ベンダー名	発話単位	最大認識可能語彙
Speechworks	単語，連続単語	12 万語
L&H	単語，連続単語	1,000 語
Nuance	単語，連続単語	制限なし (コンピュータの性能に依存)
KDDI	単語，連続単語，文	20,000 語
Philips	単語，連続単語	20 万語
NTT-IT	単語，連続単語	ソフトウェア上，制限なし
NEC	単語，連続単語	標準 5,000 語，最大 20 万語
NTT データ	単語，連続単語	最大 10,000 語

音声認識エンジンの現状を把握するために，表 2.1 の各認識エンジンに対して，最大語彙数を認識対象として与えた場合の性能評価を実施した．評価では，日本人の苗字を認識対象として，孤立単語音声認識を実行した．日本人 3,000 万人の苗字は，漢字表記の相違は考慮せずに平仮名で表記した場合の種類を数えると，全部で 87,944 種類存在する[20]．認識可能語彙数には上限なしと発表しているNuance, NTT-IT, 20 万語と記載しているPhilips, NEC, 12 万語と記載しているSpeechworksの音声認識エンジンに対しては，全個人姓 87,944 種を認識対象として与えた．実験室において，同一話者がマイクを通して発声した姓 100 種を録音し，録音した姓を各音声認識エンジンに入力した．表 2.2 に，正解率[1]，及び結果出力までに要した処理時間を示す．実用サービスにおいて音声認識を利用するためには，不特定多数の利用者の入力に対して，迅速かつ正確に応答する必要がある．しかし，音声認識エンジンの現状は，個人姓のような大語彙に対して，認識精度，結果出力までの処理時間ともに，実用レベルであるとは言い難い．

次節では，代表的な音声対話システムについて紹介し，音声認識の性能を踏まえた上で，音声対話システム実現のための理想的な対話制御手法について述べる．

[1] 利用者から入力された音声と類似度が最大と判定された認識対象単語とが一致した割合．

表 2.2：市販の音声認識エンジンに対する認識性能評価

ベンダー名	認識対象姓数	処理時間と正解率
Speechworks	87,944 語	正解率 35%, 処理時間 121 秒
L&H	1,000 語	正解率 65%, 処理時間 2 秒
Nuance	87,944 語	正解率 46%, 処理時間 15 秒
KDDI	20,000 語	正解率 69%, 処理時間 12 秒
Philips	87,944 語	正解率 25%, 処理時間 182 秒
NTT-IT	87,944 語	正解率 47%, 処理時間 489 秒
NEC	87,944 語	正解率 40%, 処理時間 46 秒
NTT データ	10,000 語	正解率 67%, 処理時間 11 秒

2.2 音声対話システム

2.2.1 音声対話システムと対話制御手法

音声対話システムとは，利用者との音声による対話を通して，利用者の要求内容を決定し，タスクを実行するシステムのことである．タスクとは，システムごとに定められた作業のことであり，例えば，各種予約や株価などの情報案内サービス，個人スケジュール管理などが挙げられる．近年の音声認識技術や自然言語処理技術の発展に伴い，様々なタスクにおいて音声対話システム実現に向けた取組みが行われている[13，31]．

利用者のどのような発話に対しても誤認識することなく応答し，タスク達成に向けて効率のよい対話を進行できるシステムが，音声対話システムの理想であると言えるのかもしれない．これまで，理想的な音声対話システムの実現に向けて，誤認識を補い，利用者との円滑な対話を実現するための対話制御手法が必要とされてきた．対話制御とは，音声対話システムにおける利用者との対話の進め方の規定である．一般的に，対話の主導権に焦点を当てると，システムの質問に対して利用者が答えていく“システム主導型”，逆に利用者が自発的に色々と話題を出し，システムから情報を引き出していく“ユーザ主導型”，対話の進行とともに主導権が交代する“主導権交代型”の 3 種の対話制御手法が存在する．人間同士の対話に最も多いのは，主導権交代型対話である．

音声対話システムの実現を目指した取組みでは，模擬対話システムを用いて収集した人間と機械との多数の対話例から，システムの質問に対する利用者の答え方や発話のタイミングなどを分析し，対話制御を決定するのが一般的である．模擬対話システムとは，システムになりすました人間が利用者と対話するシステムのことである．模擬対話システムを用いた検討では，利用者の発話を可能な限り網羅するために，利用者の発話に基づいて，システムが用意すべき認識対象語彙リストについて検討し，人間の対話に近い応答を目指す取組みが多い[54].

次節では，これまでに提案されてきた音声対話システムについて紹介する．

2.2.2 代表的な音声対話システム

本節では，代表的な音声対話システムの一例として，最初の実用音声対話システムと言われる Bell Northern Research による Automated Alternate Billing Service, イタリアの CSELT のシステムと MIT の Zue らのグループの取組み，Philips による実用システム Train Time Table Information System について紹介する．

(1) Bell Northern Research

Automated Alternate Billing Service は，1989 年に Bell Northern Research によって実用化されたコレクトコールの申込みをタスクとした音声対話システムである[8]．このシステムは，顧客 A から顧客 B へのコレクトコールの申込みがあった場合，システムは顧客 B に電話をかけ，コレクトコールを受けるか否かを尋ねる．顧客 B の回答が Yes の場合は顧客 A からの電話を顧客 B に繋ぎ，No の場合は顧客 A にその旨を伝える．このシステムにおける認識対象語彙は，Yes，No とそれらの同義語を合わせた計十数単語である．評価では，タスクが達成できた対話は 95%であり, システムが誤認識するたびに繰返される Yes, No の確認が利用者の負担となり，5%の対話が途中放棄されたと報告されている．

(2) CSELT

CSELT では，イタリア国内の列車の出発時刻，到着時刻，料金などの案内をタスクとする電話を通した 2 種の対話システムを開発した[7]．1 つはシステム主導，孤立単語音声認識による対話システム RAILTEL, もう 1 つは主導権交代型の自然発話

を対象とした対話システム Dialogos である．RAILTEL は，簡単な質問と確認を繰返すシステムであり，認識対象語彙は，駅名とシステムからの確認に対する回答である Yes，No などを合わせた計 533 単語である．これに対して，Dialogos の認識対象語彙は，駅名以外に，都市名や列車名などを合わせた計 3,471 単語であり，言語モデルとして 358 個のクラスを単位としたバイグラムを利用している．利用者の自然発話から出発駅や到着駅などのキーワードを抜き出し，該当する列車の情報を案内する．

無料電話によるフィールドテストの結果，RAILTEL は，67%の利用者に対して誤りなく出発駅と到着駅を確定し，列車の情報を案内することができた．そして，確定した対話は，平均 13 ターン，160 秒を要した．ターンとは，システムの質問とそれに対する利用者の回答の組を意味する．一方，Dialogos のタスク達成率は 70%であり，14%はシステムの不備，16%は利用者の途中放棄によって失敗に終った．さらに，CSELT では，これら 2 種のシステムの比較実験を実施している．被験者は 18 歳から 60 歳までの男女 24 名ずつの計 48 名である．興味深い結果として，システムの使いやすさ，疲れやすさ，対話の人為性，効率に関しては，システム主導，孤立単語音声認識の RAILTEL の方が，自然発話音声認識システムである Dialogos よりも高く評価された．対話が自然であったかどうか，及び合成音声の明瞭度については，両システムともにほぼ同じ結果となっている．

(3) MIT

MIT の Zue らのグループは 1989 年以来，ケンブリッジ市内の案内をタスクとした VOYAGER，航空座席の予約をタスクとした ATIS，世界 500 都市以上の天気状況案内をタスクとした JUPITER などの音声対話システムを開発している[70]．いずれも電話を通した自然発話を対象としている．JUPITER の認識対象語彙は，都市名など 1,500 単語である．JUPITER の評価実験では，1,500 対話，8,000 文を収集した結果，タスク達成率は 60%であり，システムが規定した文法範囲外の文章が入力されたためタスクが達成できなかった対話が 30%，システムの不備が 10%であったと報告されている．

(4) Philips

Philips は 1,200 都市を結ぶ列車に関する情報を電話で提供する実用システム Train Time Table Information System を開発した[4]．このシステムは，システム主導，孤立単語音声認識を用いたシステムであり，1994 年にドイツで，1996 年に

スイスで，1997 年にオランダで同様のシステムが発表されている．認識対象語彙は 1,850 単語であり，10,000 呼に対する評価の結果，タスク達成率は 80%程度であり，音声認識の失敗と認識対象リストにはない都市名が入力されたことが，タスク未達成の主な原因として報告されている．

2.2.3 実用を意識した音声対話システム

現状，不特定多数の利用者を対象とした実用レベルの音声対話システムを実現するためには，認識対象語彙と文法に多くの制約を設けなければならない．しかし，2.2.2 節で述べた音声対話システムの一例からも分かるように，システムが受理できる語彙や文法を制限しても誤認識は起こる．一方，音声認識の性能を考慮して認識対象語彙や文法を制限するほど，何をどのように発話すればよいのかというシステムに対する知識を持つ利用者でなければ，システムを使いこなすことができなくなる．すなわち，語彙や文法に多くの制約を設けると，不特定多数の利用者を対象としたシステムの実現からは遠ざかってしまう．

音声対話システム実現のためには，システムの制約を利用者に意識させずに，利用者の発話をシステムが受理可能な語彙や文法の範囲内に抑えるように誘導し，少ないやり取りでタスクが達成できるような対話制御が好ましい．これまで，認識性能を考慮した対話制御手法に関しては，数多くの研究事例がある．それらは，模擬対話システムを用いて収集した対話例から最適な対話戦略を学習するなど，多量のデータから統計的な対話モデルを提案する手法[60]や，有限オートマトンを用いて対話の流れを記述する手法[51，68]など，実験室レベルの提案が多く，不特定多数の利用者を対象とした実用システムへの適用は難しい．

これに対して，Biermann ら[6]は，システムの制約を予め利用者に伝え，利用者の発話をシステムが受理可能な範囲内に制限することは，音声対話システムの目標である円滑なコミュニケーションの妨げになると指摘する．Biermann らは，システムが主導権を取りながら利用者に質問し，利用者の回答を単語に制限したシステムが最も効率がよく，利用者からみても，システムとの対話が順調に進んでいることが実感できるシステムであると考察している．Biermann らは，システムの効率とは，タスク達成までのシステムと利用者とのやり取りの回数と，システムと利用者との間で自然な対話が実行できたかどうかで測定できると述べている．

また，シュナイダーマン[5]は，自然発話による対話システムは，利用者がシステ

ムやタスクに対して必要な知識を持ち，タスクの範囲が限定されている場合には有効であるが，制約を何も設けないシステムは，誤認識を修正するための余分な対話が必要になり，今後，広く普及することはないと考察している．さらに，2.2.2 節で述べた CSELT の 2 種の対話システムに対する評価でも，システム主導，孤立単語音声認識の方が，主導権交代型の自然発話音声認識よりも使いやすく，効率がよいと評価されている．

一般に，対話システムの評価は，タスク達成率や利用者の主観評価を基準にする場合が多い．現在，音声対話システムに対する定量的な評価手法は確立されていない．しかし，過去の研究において，タスク達成までの対話の回数や，誤提示，及び誤認識を解決するための同じ質問の繰返しなどが，利用者満足度に大きく影響することが明らかになっている[19，46]．

本研究では，音声認識の性能を踏まえ，システム主導かつ利用者の入力を単語に制限することが，利用者満足度が高い音声対話システムの実現に繋がると考える．そして，タスク達成までの対話回数を可能な限り少なくし，誤提示や同じ質問を繰返さない対話制御の提案を目指す．

次章では，利用者の入力を単語に制限しても，大語彙を対象とした音声対話システムの実現が困難である現状について述べ，本研究において解決すべき課題を明確にする．

第 3 章
大語彙を対象とした音声対話システム

前章で述べたように，実用を指向した音声対話システムは，システム主導，孤立単語音声認識という構成が好ましい．本章では，利用者の入力を単語に制限したシステムにおいて，住所や姓名などの大語彙を対象とした場合の課題について述べる．

3.1　音声対話システムの課題

3.1.1　誤認識を考慮した対話戦略

実用を指向した音声対話システムでは，利用者の入力を単語に制限しても，100%の精度は保障できないため，誤認識への対処が必須となる．認識結果のみから正誤を判断してしまうと，利用者の要求通りにタスクを実行できない可能性があり，特に予約や申込みなど，誤認識すると利用者に迷惑がかかるタスクでは，念入りな確認が必要になる[12]．

音声認識において，認識結果の正誤は，発話者本人に確認する以外に判断する手法はない．しかし，確認対話の最中にも誤認識を生じる可能性があるため，確認のための対話は必要最小限度に抑えることが好ましい．確認のための対話戦略には，大きく分けて 2 種類存在する[52]．

(1)　直接確認対話

「佐藤です」という利用者の入力に対して，「佐藤さんですか？」のように，入力ごとに認識結果を提示して確認する手法である．無駄な確認を省くために，認識結果の信頼度が高い場合は提示確認，低い場合は再入力を要求する確認手法も直接確認対話と呼ぶ．

(2) 間接確認対話

確認のための対話を省くために，次のシステムの質問に認識結果の確認を含ませる間接的な確認方式である. 例えば,「佐藤です」という利用者の入力に対して,「佐藤，何さんですか？下のお名前もおっしゃって下さい」のように，次の質問と前の質問に対する入力の確認を同時に行う.

直接確認対話は，正誤確認の繰返しによる利用者の負担が大きい．一方，間接確認対話については，誤認識修正のための音声コマンドと類似した語彙が認識対象に含まれている場合や，雑音環境などの悪条件下では，修正コマンド自体の認識が困難であり, 修正に時間がかかるという評価が多い. 新美ら[52]は数式を用いて, また, Watanabe ら[64]はコンピュータ同士のシミュレーションを用いて, 両確認対話の有効性を評価している．小坂ら[41]は，両者の利点を採用して，音声認識結果の信頼度が高い場合は間接確認対話，信頼度が低い場合は直接確認対話を実行する対話システムを提案しているが，認識結果に対する信頼度の推定方法が大きな課題であると考察している.

また，堂坂ら[13]は，無駄な確認を避けて効率のよい対話を実現するために，デュアルコスト法を提案している．デュアルコスト法とは，確認対話そのものの長さを表す確認コストと確認対話終了後のタスク達成までの対話の長さを表す情報伝達コストという 2 種のコストを導入し，確認コストと情報伝達コストの和を最小化する方向へ対話を制御することによって，対話全体の効率化を図る対話制御手法である. 堂坂らの手法は，自然発話を対象としたシステムにおいて，利用者の発話内容がシステムの制約を超えている場合に有効であることが示されている.

現状の音声対話システムでは，提示確認の結果，誤認識の場合は，利用者に対して入力要求，提示確認を繰返す．したがって，どのような確認手法を用いても，誤認識の場合は，修正のためのプロセスが繰返されるため，利用者の負担は解消されない．利用者の負担を解決するためには，音声認識の性能を踏まえ，実用レベルの精度が期待できる適切な語彙数を考慮してシステムを実現する必要がある.

次節では，現状の音声対話システムにおける適切な対象語彙数について考える.

3.1.2　音声対話システムにおける適切な対象語彙数

現在，実用化されている音声対話システムは，音声認識の性能を踏まえ，システムが主導権を取りながら利用者に質問し，利用者の入力を単語に制限している場合が多い[2][21，22，23，24，26，28，29]．実用システムにおける認識対象語彙数は非常に少ない．例えば，株価案内システムは 770 企業名，天気予報は 47 都道府県，星占いシステムは 12 星座のみであり，ほとんどのシステムが，0 から 9 までの数字と“はい”，“いいえ”のみの数単語に認識対象を制限している．2.2.2 節で述べた音声対話システムの中でも，実用化されたBell Northern Researchのコレクトコールの申込みシステムAutomated Alternate Billing Serviceの認識対象語彙は十数単語，PhilipsのTrain Time Table Information Systemは 1,850 単語であったが，いずれも，評価の結果，タスク未達成のままで終了した対話が存在する．

2.1.2 節で述べた市販の音声認識エンジンに対する性能評価から明らかなように，現状の音声認識技術を利用して，利用者に対して実用レベルの認識精度と処理速度を提供できる対象語彙数には限界がある．表 3.1 に，現在，日本において実用化されている音声対話システムの対象語彙数と複雑度を示す．

複雑度とは，認識対象語彙の音韻上の複雑さの程度を表す情報量である．認識対象語彙 L における音韻列 $w_1 \cdots w_n = W_1^n$ の生成確率を $P(W_1^n)$ とすれば，L のエントロピーは，以下の式 (3.1) で計算できる．

$$H_0(L) = -\sum_{w_1^n} P(W_1^n)\log P(W_1^n) \tag{3.1}$$

複雑度とは一音韻当たりのエントロピーに相当し，以下の式 (3.2) から計算できる．各音韻の後には平均して $2^{H(L)}$ 種の音韻が後続可能であることを意味し，複雑度が大きいほど，認識が困難な対象であると言える．

$$H(L) = -\sum_{w_1^n} \frac{1}{n} P(W_1^n)\log P(W_1^n) \tag{3.2}$$

[2] 単語の語尾に“です”，“ます”のような接尾語を付けるなど，予め想定できる利用者の発話は認識対象語彙として登録しているシステムが多い．

表 3.1：実用音声対話システムにおける認識対象語彙数と複雑度

サービス内容	対象語彙数	複雑度	実施企業（採用認識エンジン）
星占い	12 星座名	8.1	・NTT コミュニケーションズ (NTT-IT) [24] ・日本テレコム (Nuance) [29]
天気予報	47 都道府県名	9.1	・NTT コミュニケーションズ (NTT-IT) [24] ・日本テレコム (Nuance) [29]
電話番号案内	東京 38 市区名	9.4	・日本情報システム (Philips) [23]
株価案内	770 企業名	10.1	・野村證券 (Nuance) [26] ・UFJ つばさ証券 (Nuance) [28]
数字認識	0 から 9 まで 数字 10 種	7.9	・DELL カスタマーセンター (Nuance) [21] ・ソニーVAIO お客様センター (Nuance) [22]

表 3.1 に示したように，星占いシステムの対象である 12 星座名の複雑度は 8.1，株価案内システムの対象 770 企業名の複雑度は 10.1，パソコンの製造番号や会員番号などを構成する 0 から 9 までの数字の複雑度は 7.9，“はい”，“いいえ”の 2 単語の複雑度は 5.6 と非常に小さい．このことから，現状の音声認識技術を利用して，サービス提供が可能な対象語彙の複雑度は，10 程度が限界と言えるであろう．

3.2　住所，姓名の確定をタスクとした音声対話システム

3.2.1　住所，姓名の対象の複雑さ

住所や姓名の確定をタスクとした音声対話システムは，コールセンタの受付け業務などに広く適用できるため需要は高い．しかし，3.1.2 節で述べたように，音声対話システムは，天気予報や株価案内，星占いなどの対象語彙数の限られた分野でしか実現されていない．

住所は，日本全国に 47 都道府県，4,100 市区郡，173,600 町村字の計 177,747 種類の地名が存在する[3][35]．日本人 3,000 万人を対象とした場合，姓は 87,944 種類，名は 79,867 種類存在する[4][20]．表 3.2 に，住所，姓名の複雑度を示す．

表 3.1 に示した現状の実用システムの対象である星座名や数字と比較して，住所の複雑度は 13.3，姓は 16.7，名は 14.9 と非常に大きい．複雑度から考えても，住所や姓名は，規模が大きく確定が困難な対象であると言える．

表 3.2：住所，姓名の複雑度

対象	対象数	複雑度
住所	177,747 種	13.3
個人姓	87,944 種	16.7
個人名	79,867 種	14.9

3.2.2　住所，姓名の確定をタスクとした現状システム

これまで，住所や姓名の確定をタスクとした音声対話システムに関しては，多くの提案がある[1，3，11，67]．住所に関しては，ボイスポータルの天気予報サービスに代表されるように，都道府県，市区郡，町村字という階層的なデータ構造を利

[3] 漢字表記の相違は考慮せずに，平仮名表記した際の種類数．同じ読みを持つ地名は 1 種類と数える．
[4] 住所同様，平仮名で表記した際の種類数．

用して，上位から順に住所を絞り込みながら確定する提案が多い[24, 29]．これは，情報検索におけるディレクトリ検索方式を音声入力に適用したものである．日本全国の地名 177,747 種を一度に認識対象としても，高精度な応答は期待できないため，各質問の対象を小規模に設定して，リアルタイムかつ利用者の負担にならない応答を実現している[36]．

音声入力型ディレクトリ検索方式には，以下の課題がある．

(1) 入力対象を階層化して，上位から順に入力を確定する．そのため，利用者に対して，対象の階層数分の入力要求，提示確認の繰返しが必要になる．

(2) 上位階層が確定しなければ，下位の対象を絞り込むことができない．そのため，各階層において，認識結果が正解であるという確認が得られるまで，下位階層の入力要求へ進むことができない．

課題 (1) を解決するために，堂坂ら[13]は，個人姓名や部署名など，予め規定したスロットを順に埋めていくタスクにおいて，現在のスロットとその値からスロット間の依存関係を算出し，その後の提示確認回数を最小化する方法を提案している．これに対して，伊藤ら[34]は，タスク達成までの効率のみを優先した質問順序は機械特有であり，大語彙を対象とした音声対話システムが実用に至らない理由に繋がっていると考察している．

階層を利用した住所確定対話の実用性を検討するために，オペレータと利用者との対話を分析した．利用者から住所，姓名，年齢などの情報を聞き取り，商品カタログの送付を専門業務とするコールセンタ[5]への 1 日のアクセスを記録した．首都圏在住の利用者は，オペレータから住所を尋ねられると，港区，横浜市，千葉市のように市区郡名を最初に答える場合が多く，特に，東京 23 区在住の利用者の多くは，六本木，虎ノ門，千住のような字名を最初に答える傾向がみられた．都道府県名を最初に答える利用者は非常に少数であることから，都道府県名の入力を強制するのではなく，どの階層の地名が最初に入力されても対応できるシステムが好ましい．

一方，個人姓名の確定をタスクとした音声対話システムは，パソコンのサポート

[5] 所在地は東京都港区，1 日の平均アクセス数は 5,000 コール．

センタなどでの実用化例がみられる．しかし，階層的なデータ構造を持たない姓名には，ディレクトリ検索方式が適用できないため，数百名程度の事前登録会員のみを対象にするなど，対象を小規模に制限しなければ実現は困難である．夜間や休日などのオペレータ業務時間外に，不特定多数の利用者からのアクセスに自動応答で対応するサービスでは，利用者の入力を正しく認識し確定することが困難であるため，姓名や電話番号を留守番電話に向かって発話するように利用者に依頼し，後日，人間がその録音を聞いてコールバックする方法が一般的である[25]．新規加入申込みなど，予め対象が制限できない姓名を対象とした実用システムでは，実在者数の偏りを利用して，実在頻度順位上位から数千種の頻出姓名のみを対象としている場合が多い．しかし，これでは，希少姓名を持つ利用者からのアクセスに対応できないため，対象外の姓名を持つ利用者の満足度が獲得できない．現状，不特定多数の利用者の姓名を網羅した姓名確定システムの実用化例は存在しない．

3.3 本研究において解決すべき課題

実用サービスの入力インタフェースに音声認識を適用するためには，不特定多数の利用者の入力に対して，迅速かつ正確に応答する必要がある．しかし，利用者の入力を単語に制限して，システム主導に対話を進めても，高精度な応答が提供できる語彙数には限界があるため，住所や姓名などの大語彙を対象としたシステムの実現は極めて困難である．特に，姓名については，階層構造を用いた絞り込みができないため，利用者の入力に対して，すべての姓との類似度を計算し，最も類似している候補を利用者に提示確認するのが現状である．直接確認対話，間接確認対話のどちらの確認手法を採用しても，誤認識の場合は，利用者に対して入力要求，提示確認のプロセスを繰返す．対象が大語彙になるほど誤認識は起こりやすく，修正のための同じプロセスの繰返しは，利用者にとって大きな負担となる．

大語彙を対象としたシステムでは，類似した音韻を持つ候補が増えるため，人間同士の対話では，通常みられないような誤認識が起こりやすくなり，利用者は人間の対話との相違を感じる．そして，システムは，利用者に対して修正のためのプロセスを繰返す．本研究では，認識対象が大語彙の場合，人間同士の対話でも聞き間違いはあることから，システムが誤認識すること自体が利用者の負担の要因ではないと考える．人間の対話では起こらないような機械特有の誤認識と修正のための同じプロセスの繰返しが，利用者に大きな負担を与えている．

本研究では，大語彙を対象とした現状の音声対話システムには，以下の 2 つの課題があると仮定する．

(1) 人間同士の対話では起こらないような，利用者にとって予測不可能な機械特有の誤認識が多い．

(2) 誤認識の場合，利用者に対して入力要求，提示確認という修正のためのプロセスを繰返す．

本研究では，上記 (1)，(2) の課題を解決し，対象が大語彙であっても，利用者満足度が高い音声対話システム実現のための対話制御手法を提案する．大語彙とは，住所や姓名などの十万語程度の語彙数を意味する．人間と同様の聞き取り傾向を実現し，誤認識の場合，同じプロセスを繰返さずに，迅速かつ効率よく正解を導き出せるような対話制御の実現を目指す．

第 4 章
思い込み応答

我々は，初対面の相手に住所や名前を尋ねる際，聞き覚えがある地名や姓名に対しては，正しく聞き取ることができても，初めて耳にする地名や，日頃，聞き慣れない珍しい姓名については，聞き間違えることや正しく聞き取れたかどうかという確信を持てない場合が多い．

前章で述べたように，音声対話システムでは，対象が大語彙になるほど，人間同士の対話では，通常みられないような誤認識が起こりやすくなる．本研究では，この機械特有の誤認識が利用者の負担になっていると仮定する．本章では，人間同士の対話と同様の聞き取り傾向を利用者に提供するために，思い込み応答を提案する．個人姓 87,944 種を対象とした音声対話システムに思い込み応答を適用することによって，利用者の負担が解決できることを検証する．

4.1　思い込み応答

4.1.1　人間の対話における聞き取り傾向

人間は，聞き慣れない言葉を耳にした時，聞き覚えがある類似した言葉に聞き間違えることが多い．本節では，発話対象が住所や姓名のような大語彙の場合，人間は，すべての対象を網羅しているわけではなく，聞き慣れた言葉は間違いにくいが，聞き覚えのない珍しい言葉は，聞き慣れた中の類似している対象に間違いやすいのではないかと予想し，検証を行った．以下，人間が大語彙を聞き取る際の傾向を分析するために実施した，87,944 種の個人姓を対象とした聞き取り試験の概要を示す．

男女合わせて 10 名の被験者に，電話回線を通して個人姓 4,000 種の聞き取りを依頼した．被験者に対して，1 番から 4,000 番までの番号のみが記してある記録用紙を渡し，聞き取った結果の記入を指示した．被験者は，音声認識技術や対話システムに関する知識を持たない 20 代から 30 代の男女 5 名ずつの計 10 名である．被験

者に聞き取りを依頼した音声には，ナレータ業務を専門とする女性の発声の録音を使用した．被験者には，日本人の苗字が発話されることのみを事前に伝え，1 件につき 1 回の聞き取りを原則とした．1 つの姓について被験者から終了の合図が出るのを確認して，実験担当者は次の姓を回線に流す．これを 4,000 種の姓について繰返す．聞き取れなかった姓については，空欄にすることを認めた．試験では，実在頻度順位[6][20]が均一になるように 4,000 種の姓を選択した．表 4.1 に，試験に使用した姓の実在頻度順位，及び被験者の平均正解率を示す．

表 4.1 から，実在頻度順位が上位の姓ほど正解率が高いことが分かる．この結果から，日頃，よく耳にすると思われる実在者数の多い姓については正しく聞き取れるが，聞き慣れない希少姓ほど間違いやすいという人間の聞き取り傾向が確認できる．特に，実在頻度順位 5,000 位以内の姓については 93.8%，実在頻度順位 5,001 位から 10,000 位以内の姓については 93.7%と非常に高い正解率が得られている．

表 4.1：聞き取りに使用した 4,000 種の個人姓と被験者の聞き取り精度

実在頻度順位	データ数（件）	平均正解率（%）
1 位　～　5,000 位	400	93.8
5,001 位　～　10,000 位	400	93.7
10,001 位　～　20,000 位	400	73.9
20,001 位　～　30,000 位	400	62.6
30,001 位　～　40,000 位	400	51.1
40,001 位　～　50,000 位	400	48.1
50,001 位　～　60,000 位	400	60.6
60,001 位　～　70,000 位	400	59.2
70,001 位　～　80,000 位	400	60.8
80,001 位　～　87,944 位	400	60.9
計	4,000	66.5

[6] 日本全国，実在者数の多い順に個人姓を並べた順位．

次に，人間が聞き間違いやすい姓の特徴を掴むために，被験者の聞き間違い方に着目した．図 4.1 に，被験者が聞き間違えた姓の実在頻度順位（X 軸）に対して，間違えた先の姓の実在頻度順位（Y 軸）をプロットした．

図 4.1 のプロットはグラフの下部に集中している．分析したところ，不正解のうち実在姓への聞き間違いは 5,116 件存在し，そのうち約 8 割に該当する 3,990 件は実在頻度順位 10,000 位以内の姓への聞き間違いであり，実在頻度順位 20,001 位以降の姓への聞き間違いは全体の 1 割に満たない．さらに，5,116 件のうち 99.5%は，自分自身よりも頻度上位の姓に聞き間違えている．

この結果から，人間は，87,944 種の個人姓を聞き取る際，すべての姓を把握しているわけではなく，聞き慣れた頻度上位の姓の中から聞き取った結果を探し出そうとする傾向が強いと考えられる．試験を通して，頻度上位の姓ならば正しく聞き取れるが，希少姓については，頻度上位の類似姓に聞き間違いやすいという人間の聞き取り傾向が確認できた．

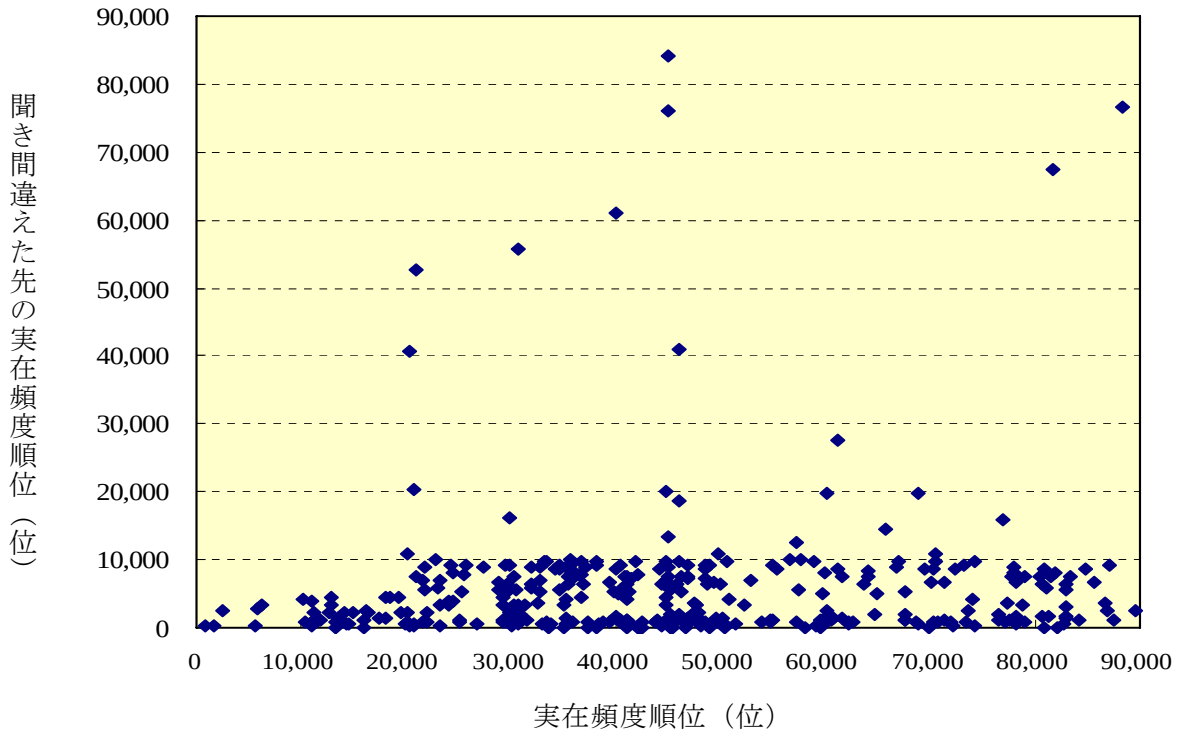

図 4.1：聞き間違えた姓と間違い先の姓の実在頻度順位との関係

4.1.2　音声対話システムへの思い込み応答の適用

4.1.1 節の聞き取り試験では，人間同士の対話でも聞き間違いは起こり，さらに，聞き間違いやすい対象と間違いにくい対象が存在することを検証した．本研究では，誤認識そのものではなく，人間同士の対話では起こらないような誤認識が利用者に負担を与えていると考える．人間が間違いやすい対象をシステムが同じように誤認識したとしても，利用者の負担にはならないと予想する．

そこで，人間が聞き間違いにくい対象についての精度を高め，機械特有の誤認識を解決するために，“思い込み応答”を提案する．思い込み応答とは，人間が聞き間違いにくい対象のみを認識対象に設定して，人間と同様の聞き取り傾向を実現する応答戦略である．こうすることで，設定した対象については，対象語彙数が制限されるため，認識精度，及び処理速度が大きく向上する．設定外の対象が発話された場合は誤認識となるが，これは人間も間違いやすい対象なので，その後の対話制御で迅速に正解を導き出すことができれば，初回の応答が誤認識であっても利用者の負担にはならない．すなわち，思い込み応答によって，機械特有の誤認識は回避できる．

思い込み応答では，利用者からアクセスされやすい対象を数多く思い込むほど，利用者に対して正解が提示できる可能性は大きくなる．しかし，認識精度と網羅率[20]はトレードオフの関係にあり，認識対象語彙数が増えるほど精度は低下する．

次節では，市販の音声認識エンジンを使用した実験を通して，思い込み対象の選択方法を提案する．

4.2　音声認識エンジンを用いた思い込み対象の分析

4.2.1　認識精度と網羅率

本節では，Nuance7.03[27]を使用した実験を通して，認識精度と網羅率，利用者満足度との関係から，思い込み応答の対象として選択すべき個人姓について分析する．

Nuanceは，予め用意した姓の仮名表記リストを渡すと，Nuance Grammar Builderを用いて認識対象辞書を作成する．そして，利用者の入力に対して，作成した認識対象辞書内のすべての姓との尤度を計算し，尤度の大きい順に候補を提示す

る．尤度とは，利用者の入力と認識対象辞書内の姓との類似度を表す値である[7]．実験において，提示候補数はデフォルト値の 10 とした[8]．すなわち，最大で 10 候補が出力される．

実験では，実在頻度順位 1 位から対象数を変化させた 14 種の個人姓認識対象辞書を用いて，認識精度と網羅率，利用者満足度との関係を調べる．被験者は 4.1.1 節で個人姓の聞き取りを依頼した 10 名である．被験者に対して，予め選択した個人姓 400 種の発声を依頼した．400 種は，実社会と同じ実在頻度分布を構成するために，各認識対象辞書に含まれる割合が，網羅率と一致するように選択した．表 4.2 に，Nuance に与えた認識対象辞書を構成する個人姓の実在頻度順位と網羅率，及び入力姓 400 種が各辞書に含まれる割合と複雑度を示す．複雑度は，3.1.2 節で述べた定義式 (3.2) を用いて計算した．例えば，辞書 A には，被験者の入力姓 400 種のうち 72.3%に当たる 289 種が含まれ，この 289 種に対して Nuance は尤度を計算する．残りの 111 種については，辞書 A に含まれていないため，認識結果に出現することはなく誤認識となる．

表 4.2 に示したように，87,944 種の個人姓のうち，実在頻度順位上位 5,000 種の姓で，実在者数の約 9 割を網羅できることから，個人姓は実在頻度に大きな偏りがあることが分かる．

表 4.3 に，被験者の入力に対する各辞書の平均認識精度を示す．思い込み対象の選択方法を分析するために，認識エンジンの出力結果が音声対話システムの第一応答として受入れられる精度か否かを被験者に尋ねた．表 4.3 の最右欄に，受入れ可能と答えた被験者数を示した．90%以上の認識精度を持つ認識対象辞書 C，D，E については，すべての被験者が音声対話システムの第一応答として受入れ可能な精度であると評価している．

網羅率に着目すると，全被験者が受入れ可能と評価した認識対象辞書 C，D，E の網羅率は 90%以上である．辞書 F 以降については，網羅率が 95%以上でも認識精度が 90%以下であるためか，利用者満足度が獲得できていない．また，辞書 A，B に関しては，認識精度が 90%以上でも網羅率が低いためか，利用者満足度も低い．したがって，認識精度のみではなく網羅率も考慮して思い込み対象を選択しなければ，利用者満足度は獲得できないことが分かる．

また，表 4.2 の複雑度に着目すると，対象語彙数が多くなるほど複雑度は大きく

[7] 0 から 100 までの値で算出される．
[8] N-best値＝10 と設定する．

なる．全被験者が受入れ可能と判断した中で，網羅率が最大，すなわち，最多の対象数で構成される辞書 E の複雑度は 9.8 である．これは，現状の音声認識技術を利用して，実用レベルの認識精度と処理速度を利用者に提供できる対象の複雑度は 10 程度であるという 3.1.2 節で述べた見解に一致する．表 4.2 の複雑度に基づいて考えると，Nuance を使用した実用サービスにおける認識対象語彙数は，最大で 10,000 語程度が限界であると考えられる．

表 4.2：Nuance に与えた認識対象辞書，及び入力に使用した 400 種の個人姓

辞書名	辞書を構成する姓の実在頻度順位	網羅率（%）	含まれる入力姓数（件）(400 件に占める割合（%))	複雑度
A	1 位 ～ 1,000 位	72.3	289 (72.3)	8.2
B	1 位 ～ 3,000 位	86.2	345 (86.2)	8.9
C	1 位 ～ 5,000 位	90.6	362 (90.6)	9.3
D	1 位 ～ 8,000 位	94.1	376 (94.1)	9.4
E	1 位 ～ 10,000 位	95.6	382 (95.6)	9.8
F	1 位 ～ 15,000 位	97.1	388 (97.1)	10.6
G	1 位 ～ 20,000 位	98.0	392 (98.0)	11.2
H	1 位 ～ 30,000 位	98.9	395 (98.9)	12.8
I	1 位 ～ 40,000 位	99.4	397 (99.4)	13.6
J	1 位 ～ 50,000 位	99.6	398 (99.6)	14.1
K	1 位 ～ 60,000 位	99.7	398 (99.7)	14.9
L	1 位 ～ 70,000 位	99.7	398 (99.7)	15.7
M	1 位 ～ 80,000 位	99.9	399 (99.9)	16.1
N	1 位 ～ 87,944 位	100.0	400 (100.0)	16.7

表 4.3：辞書ごとの認識結果，及び被験者による受入れ可否評価

辞書名	網羅率（%）	平均認識精度（%）	被験者評価結果 受入れ可人数（人）
A	72.3	96.0	6
B	86.2	95.9	7
C	90.6	94.1	10
D	94.1	91.2	10
E	95.6	90.8	10
F	97.1	80.3	8
G	98.0	76.5	8
H	98.9	66.3	7
I	99.4	67.2	4
J	99.6	57.4	4
K	99.7	42.1	2
L	99.7	44.7	0
M	99.9	48.3	0
N	100.0	44.4	0

4.2.2　人間と音声認識エンジンとの思い込み結果の比較

本節では，人間の聞き取り傾向と Nuance を使用した思い込み応答との比較を通して，思い込み応答の有効性を検証する．

実験では，思い込み対象として，4.2.1 節ですべての被験者が受入れ可能な精度であると判断した中で，網羅率が最大の認識対象辞書 E を Nuance に与える．入力には，4.1.1 節の聞き取り試験に用いたナレータの録音音声 4,000 種の中から選択した実在頻度順位 10,000 位以内の姓 400 種を用いた．表 4.4 に，入力 400 種に対する Nuance の認識結果を示す．4.1.1 節の聞き取り試験における 400 種に対する被験者の聞き取り精度と比較すると，頻度順位上位の姓 10,000 件を思い込み対象とした場合，思い込み対象に含まれる姓に対する音声認識エンジンの精度は，人間の聞き取り精度とほぼ同じであることが分かる．

次に，思い込み対象外の姓について考察する．Nuance に認識対象辞書 E を思い込み対象として与え，今度は，4.1.1 節のナレータの録音音声 4,000 種のうち，実在頻度順位 10,001 位以降の 3,600 種を入力する．Nuance は与えられた辞書 E にある

姓に対して尤度を計算するため，入力姓 3,600 種に対する結果はすべて誤認識となる．表 4.5 に，Nuance が最大尤度を算出した 3,600 種の誤認識結果と，4.1.1 節の聞き取り試験で被験者が実在姓に聞き間違えた 5,116 件の間違い先とが一致した割合を示す．辞書 E を思い込み対象とした場合，思い込み応答の精度，及び聞き間違い方ともに，人間の聞き取りとほぼ一致する．この結果から，思い込み応答は，利用者に対して人間と同様の傾向を持つ応答を提供できることが確認できた．

表 4.4：音声認識エンジンの出力結果と人間の聞き取り結果の比較

実在頻度順位（位）	試験件数（件）	平均認識精度（%）	
		音声認識エンジン	被験者
1 位 ～ 5,000 位	200	93.9	93.8
5,001 位 ～ 10,000 位	200	91.2	93.7
計	400	92.6	93.8

表 4.5：認識エンジンと人間の聞き間違い先一致度

実在頻度順位	聞き間違い先一致度（%）
10,001 位 ～ 20,000 位	87.3
20,001 位 ～ 30,000 位	87.9
30,001 位 ～ 40,000 位	86.5
40,001 位 ～ 50,000 位	89.1
50,001 位 ～ 60,000 位	91.1
60,001 位 ～ 70,000 位	87.3
70,001 位 ～ 80,000 位	91.4
80,001 位 ～ 87,944 位	90.9
計	88.9

4.3　住所への思い込み応答の適用

これまで，個人姓に焦点を当て議論を進めてきたが，思い込み応答の住所への適用は，現状の住所確定対話システムの課題を解決できる．現状のシステムでは，177,747 種の地名を対象とした場合，実用レベルの認識精度と処理速度を利用者に提供することは困難であるため，利用者に都道府県名の入力を強制し，上位階層から順に住所を絞り込みながら確定する．しかし，3.2.2 節で述べたように，コールセンタのオペレータに対して，都道府県名から発話する利用者は少ないことから，どの階層の地名が発話されても対応可能なシステムが好ましい．コールセンタへのアクセスの多くは，サービス提供地域や首都圏在住の利用者からであり，アクセスされる地名に偏りがあることが分かっている．この偏りを利用して，アクセス頻度の高い地名を思い込み対象として選択することで，どの階層の地名が最初に入力されても受理可能であり，かつ大部分の利用者に対しては，高精度な応答が提供できるシステムの実現に繋がる．住所への適用については，7 章で詳しく述べる．

第 5 章
属性を利用した対話による誤認識修正手法

前章で提案した思い込み応答は，人間が聞き間違いにくい対象のみを選択し，認識エンジンに設定するため，設定外の対象が発話された場合は誤認識となる．音声対話システムにおいて，タスク達成のためには誤認識の解決が必須となる．現状のシステムでは，誤認識の場合，利用者に対して入力要求，提示確認という修正のためのプロセスを繰返す．しかし，人間同士の対話では，聞き間違えた場合，話し手に何度も同じ発話を強要しない．これが人間の対話と音声対話システムとの相違である．思い込み応答によって，人間と同様の聞き取り傾向が実現できたとしても，その後の修正のための機械的なプロセスの繰返しは，思い込み応答の効果を消してしまうことになる

本章では，利用者に負担をかけずに，誤認識を修正するための対話制御手法を提案する．提案手法は，利用者に関連情報を尋ね，尋ねた結果に基づいて大語彙を絞り込み誤認識を修正する．本研究では，この関連情報を属性と呼ぶ．大語彙に対する絞り込み効果を属性の有効度として定義する．87,944 種の個人姓を用いて，属性の有効度の有用性を検証する．

5.1　人間の対話における誤認識修正手法

本節では，3.2.2 節で述べたコールセンタのオペレータと利用者との対話記録から，相手の発話が聞き取れなかった場合の人間の対応を分析する．利用者の発話を聞き損じた場合のオペレータの対応は，利用者にその事実を知らせずに再発話を要求するか，あるいは利用者に違和感を与えないような関連情報を尋ね，住所や姓名を絞り込むかの 2 通りに分けられる．

利用者の住所 500 発話，及び姓 500 発話に対するオペレータの対応のうち，利用者に再発話を要求した対話数は，住所は全体の約 4%に該当する 21 発話，姓については全体の約 70%に該当する 341 発話であった．住所については，約 8 割の利用者

が，都道府県名を省略して市区郡名から発話している．オペレータは，聞き取れなかった住所については，「都道府県からお願いできますか？」という階層を利用した質問をする場合が多い．姓については，再発話を要求した 341 発話のうち 173 発話は，再発話を聞いても確定できないままであった．このうち 7 割に当たる 120 発話については，オペレータは，次の質問で，「漢字で書くと大きい（おおきい）に森（もり）と書く大森（おおもり）様ですか？」，「太い（ふとい）に田んぼ（たんぼ）の田(た）と書く太田（おおた）様ですか？」などのように，利用者に漢字表記を尋ね，漢字に基づいて姓を絞り込み確定まで導いた．このように，オペレータは，聞き間違えた場合，利用者に対して，関連情報を尋ね対象を絞り込む方向へ対話を誘導し，少ないやり取りで正解を導き出そうとする．その際，オペレータは，対象の絞り込みが可能で，かつ利用者が回答に戸惑うことのない質問を投げかけようとする．また，確定が困難な姓名については，利用者からオペレータに対して，部首や別の漢字，あるいは固有名詞を例に挙げて，自分の姓名に使われている漢字を説明する[9]などの協力的な場面もみられた．

この分析から，人間は，誤認識した場合，話し手に対して入力要求，提示確認を繰返さないことが分かる．現状の音声対話システムでは，再入力された音声に対して再認識を繰返しても，絞り込みに繋がる新たな情報は獲得できないため，正解が提示できない場合も多く，同じプロセスの繰返しに対して利用者が負担を感じるのは明らかである．本研究では，利用者に再発話を要求するのではなく，オペレータと同様に，絞り込みに繋がる関連情報を尋ね，誤認識を修正する対話制御によって利用者満足度獲得を目指す．その際，関連情報として何を尋ねるべきかが課題となる．

次節では，尋ねるべき関連情報の選択方法について検討する．

[9]「三水（さんずい）に青い（あおい）と書く清水（しみず）です」や「人偏（にんべん）に口（くち）を書いて木（き）を書く保田（やすだ）です」，「木（き）が 3 つの森（もり）です」や「西郷隆盛（さいごうたかもり）の西郷（さいごう）です」や「富士山（ふじさん）の富士（ふじ）です」など．

5.2　絞り込みに有効な情報の判断手法

5.2.1　属性の有効度

本節では，大語彙に対する絞り込み効果を表す尺度として属性の有効度を定義する．有効度に基づいて，尋ねるべき属性を判断することによって，大語彙に対して効率のよい絞り込みが可能になる．

個人姓 87,944 種の属性として，文字数と頭文字を例に挙げる．姓の文字数とは，平仮名で表記した際の文字数であり，拗音は前の文字と併せて 1 文字と数える．“つ”，“じょ”などの 1 文字の姓から，“ひがしあそうばら”，“みなみきよせがわ”のような 8 文字の姓まで存在する．すなわち，文字数という属性は全部で 8 個の属性値を持つ．頭文字は平仮名 45 音と濁音 25 音，拗音は，文字数と同様に，前の平仮名と併せて 1 文字と数えると 30 音存在し，合計 100 個の属性値を持つ．文字数，頭文字ともに，各属性値に対する姓の分布が均一で，かつ認識精度が 100%ならば，文字数の確定によって姓の候補数は 8 分の 1 に，頭文字の確定によって候補数は 100 分の 1 に減少する．したがって，文字数よりも属性値を多く持つ頭文字の方が，姓に対する絞り込み効果が大きいことになる．しかし，各属性値に対する対象の分布は均一であるとは限らない．表 5.1 に，各文字数に対する姓の分布を示す．1 文字であることが確定した場合には候補数は 74 候補に減少するが，4 文字であることが確定しても 48,304 候補にしか減少しない．表 5.2 に，頭文字について，100 種の頭文字のうち該当する姓を多く持つ上位 20 種を示す．

表 5.1：文字数における個人姓の分布

姓の文字数	実在件数（件）	全体に占める割合（%）
1 文字	74	0.08
2 文字	2,299	2.61
3 文字	22,210	25.25
4 文字	48,304	54.93
5 文字	11,806	13.42
6 文字	2,635	3.00
7 文字	495	0.57
8 文字	121	0.14
計	87,944	100.00

表 5.2：頭文字における個人姓の分布（上位 20 種）

姓の頭文字	実在件数（件）	全体に占める割合（%）
か	6,065	6.90
い	5,004	5.69
お	4,480	5.09
し	3,933	4.47
た	3,903	4.44
あ	3,519	4.00
こ	3,294	3.75
う	3,026	3.44
は	2,925	3.33
な	2,791	3.17
と	2,742	3.12
み	2,711	3.08
さ	2,701	3.07
ま	2,561	2.91
く	2,506	2.85
ふ	2,214	2.52
ひ	2,213	2.52
き	2,181	2.48
や	2,163	2.46
つ	2,020	2.30
:	:	:
計	87,944	100.00

さらに，音声対話システムでは属性値の入力も音声であり，100 %の精度で正しい属性値を獲得できる保証はない．例えば，表 5.1，表 5.2 から分かるように，22,210 種存在する“3 文字”の姓であることよりも，2,561 種存在する頭文字が“ま”で始まるという情報の方が，個人姓に対する絞り込み効果は大きいが，“3 文字”よりも“ま”の認識が困難な場合は，文字数よりも頭文字の方が姓の絞り込みに有効な属性であるとは言い難い．すなわち，正しい属性値が確定できなければ，対象の曖昧性減少には繋がらないため，属性の有効度は，属性値の確定難易度も考慮して定義する必要がある．そこで，本研究では，属性の有効度$U(A_i)$を，属性の確定難易度$D(A)$と，属性値が確定した場合の曖昧性減少度合い$I(A_i)$を用いて，以下のように属性値ごとに定義する．

(1) 属性の確定難易度は，3.1.2 節で述べた複雑度の逆数を用いて，各属性に対して計算する．属性Aの属性値A_iの生成確率を$P(A_i)$とすると，属性Aの確定難易度$D(A)$は，次の式 (5.1) で定義する．

$$D(A) = \frac{1}{\sum_{A_i} \frac{1}{i} P(A_i) \log P(A_i)} = \frac{i}{\sum_{A_i} P(A_i) \log P(A_i)} \quad (5.1)$$

文字数の確定難易度を例に挙げて説明する．文字数の 8 個の属性値を認識するためには，1 文字に対しては“いちもじ”又は“ひともじ”，2 文字に対しては“にもじ”又は“ふたもじ”のように，音韻として複数の回答が予測できることから，全部で 11 種[10]の認識対象リストを用意する必要がある．文字数 = {いちもじ，ひともじ，にもじ，ふたもじ，･･･ ，はちもじ} の確定難易度は，式 (5.1) を用いて，次のように計算する．

[10] 3 文字 : “さんもじ”，4 文字 : “よんもじ”，5 文字 : “ごもじ”，6 文字 : “ろくもじ”，7 文字 : “ななもじ”，“しちもじ”，8 文字 : “はちもじ” の 11 種.

$$D(\text{文字数}) = -11 \times \frac{1}{P(\text{いちもじ})\log P(\text{いちもじ})} + \cdots\cdots + \frac{1}{P(\text{はちもじ})\log P(\text{はちもじ})}$$

$$= 0.18$$

(2) 属性値確定による曖昧性減少度合いは，属性値ごとに計算する．属性Aの属性値A_iが確定した場合の曖昧性減少度合い$I(A_i)$は，次の式 (5.2) で定義する．

$$I(A_i) = -\log \frac{N_{A_i}}{N} \tag{5.2}$$

N_{A_i} ･･･ 該当する属性値を持つ対象数

N ･･･ 全対象数

例えば，“さんもじ”という文字数の属性値が確定した場合の曖昧性減少度合いは，3 文字姓（$N_{\text{さんもじ}}$）は 22,210 種，全個人姓(N）は 87,944 種であることから，式 (5.2) を用いて，次のように計算する．

$$I(\text{さんもじ}) = -\log \frac{22{,}210}{87{,}944}$$

$$= 1.99$$

(3) (1), (2)を用いて，属性の有効度を次の式 (5.3) で定義する．有効度$U(A_i)$は各属性の属性値ごとに求める．

$$U(A_i) = D(A) \times I(A_i) \tag{5.3}$$

5.2.2 個人姓における属性の有効度

5.2.1 節で定義した有効度を個人姓 87,944 種に適用し，属性ごとの姓に対する絞り込み効果を比較する．個人姓を絞り込むための属性として，姓の文字数，頭文字，先頭に使われる漢字の読み仮名の 3 種を用いる．先頭漢字としては，姓の先頭に使用される常用漢字のみに焦点を当てる．漢和辞典[2]に掲載されている常用漢字 1,945 種のうち，姓 87,944 種の先頭に使用される漢字は 309 種のみである．漢和辞典に記載されている 309 種の常用漢字の読み仮名は，計 1,190 種類[11]存在する．この 1,190 種を先頭漢字の読み仮名の属性値として採用する．先頭漢字の読み仮名の有効度を求めるために，1,190 種の先頭漢字の読み仮名に対する確定難易度，及び先頭漢字の読み仮名が確定した場合の姓の曖昧性減少度合いを計算する．全体の約 5% に当たる 4,990 種の姓は，先頭が常用漢字でないため，読み仮名を付与することができない．したがって，先頭漢字の読み仮名が確定しても，これら 4,990 種の姓の絞り込みはできない．

5.1 節で述べたコールセンタにおけるオペレータと利用者との対話では，オペレータは，聞き取りにくい姓に対して，利用者に漢字情報を尋ね，確定を試みる場面が多くみられた．したがって，個人姓を絞り込むために漢字情報を利用することは，人間の対話と同様である．しかし，不特定多数の利用者の漢字の説明は，同じ姓や同じ漢字であっても多様であり，すべての発話に対応できるように認識対象を用意して漢字を正しく特定することは，実装上難しい．そこで，本研究では，個人姓の属性として，姓の先頭に使用される常用漢字の読み仮名を用いることを提案する．例えば，「増」に対しては“ぞう”，“そう”，“ます”，“ふえる”，“ま”，“ふ”という

[11] 漢字表記の相違は考慮せずに，平仮名で表記した際の種類数．(例) 増す，益す，枡，升，鱒は“ます”で 1 種類．

ように，訓読みは，読みに続く送り仮名を付けた形も属性値に含める．具体例を挙げると，“ます”という読みを持つ漢字は，「増す」，「升」，「鱒」，「益す」など複数存在し，“ます”と読む漢字を先頭に使う姓は，917 種類[12]存在する．したがって，“ます”という漢字の読みが確定した場合の曖昧性減少度合いは，式 (5.2) を用いて，次のように計算する．

$$I\ (\text{ます}) \quad = \quad -\log \quad \frac{917}{87{,}944}$$

$$= \quad 6.58$$

この曖昧性減少度合いに，先頭漢字の読み仮名 1,190 種の確定難易度 D(漢字読み) = 0.09 を乗算した値が，漢字読み属性値“ます”の有効度になる．表 5.3 に，文字数，頭文字，先頭漢字の読み仮名の属性値数と確定難易度を示す．3 種の属性の各属性値について，曖昧性減少度合いを計算し有効度を降順に示したものが，表 5.4，表 5.5，表 5.6 である．表 5.5 の頭文字，表 5.6 の先頭漢字の読み仮名については，有効度の大きい属性値と小さい属性値の一部を示す．

[12] ますだ (増田，枡田，益田，舛田，升田)，ますもと (増本，益本，升本)，ますざわ (増沢，鱒沢，升沢)，ますなが (増永，益永)，ましこ (増子，益子)，ぞうし (増子，草子) など，計 917 種類存在する．

表 5.3：各属性の確定難易度 $D(A)$

属性種類	文字数	頭文字	先頭漢字の読み仮名	個人姓全体
対象数	11	100	1,190	87,944
確定難易度 $D(A)$	0.18	0.15	0.09	0.07

表 5.4：文字数の確定難易度 $D(A)$，曖昧性減少度合い $I(A_i)$，有効度 $U(A_i)$

属性値	確定難易度 $D(A)$	曖昧性減少度合い $I(A_i)$	有効度 $U(A_i)$
1 文字	0.18	10.22	1.80
8 文字		9.51	1.68
7 文字		7.48	1.32
2 文字		5.26	0.92
6 文字		5.06	0.89
5 文字		2.90	0.51
3 文字		1.99	0.35
4 文字		0.87	0.15
平均		－	0.76

表 5.5：頭文字の確定難易度$D(A)$，曖昧性減少度合い$I(A_i)$，有効度$U(A_i)$

属性値	確定難易度 $D(A)$	曖昧性減少度合い$I(A_i)$	有効度$U(A_i)$
きょ	0.15	15.42	2.32
びゅ		15.42	2.32
づ		15.42	2.32
びゃ		15.42	2.32
きゃ		15.42	2.32
みょ		14.84	2.23
みゃ		14.84	2.23
ぎゅ		14.42	2.17
ほ		14.10	2.12
ひゅ		13.84	2.08
:		:	:
:		:	:
な		4.98	0.75
は		4.91	0.74
う		4.86	0.73
こ		4.74	0.71
あ		4.64	0.70
た		4.49	0.68
し		4.48	0.68
お		4.30	0.65
い		4.13	0.62
か		3.86	0.58
平均		−	1.36

表 5.6：先頭漢字の読み仮名の確定難易度$D(A)$, 曖昧性減少度合い$I(A_i)$, 有効度$U(A_i)$

属性値	確定難易度 $D(A)$	曖昧性減少度合い$I(A_i)$	有効度$U(A_i)$
まゆずみ		16.24	1.49
あくつ		15.66	1.44
たつみ		15.24	1.40
わび		14.92	1.37
あえる		14.66	1.35
あえ		14.66	1.35
しめる		14.66	1.35
ふもと		14.66	1.35
まきもの		14.66	1.35
もてなす		14.66	1.35
:		:	:
:	0.09	:	:
せん		5.73	0.53
しん		5.58	0.51
と		5.51	0.51
お		5.49	0.50
じょう		5.28	0.48
か		5.26	0.48
せい		5.19	0.48
こ		4.98	0.46
こう		4.85	0.44
しょう		4.19	0.38
平均		―	0.94

5.3　属性の有効度評価

本節では，評価プログラムを用いて，5.2.1 節で定義した属性の有効度の有用性を検証する．確定対象は 87,944 種の個人姓であり，個人姓の属性として，文字数，頭文字，先頭漢字の読み仮名を利用する．

5.3.1　有効度評価プログラム

市販の音声認識エンジン Nuance7.03 を使用して，以下，3 種の評価プログラムを実装する．評価プログラムにおいて，個人姓，及び各属性の認識対象辞書は Nuance Grammar Builder を用いて作成する．Nuance の N-best 値は 10 とする．

(1) 文字数評価プログラム

1. 文字数の入力を要求する．
2. 文字数 11 種を対象として，入力された文字数を認識する．
3. 姓の入力を要求する．
4. 認識結果 1 位の文字数で構成される姓のみを対象として，入力された姓を認識する．
5. 認識結果 1 位の姓の正誤を確認する．

(2) 頭文字評価プログラム

1. 頭文字の入力を要求する．
2. 頭文字 100 種を対象として，入力された頭文字を認識する．
3. 姓の入力を要求する．
4. 認識結果 1 位の頭文字で始まる姓のみを対象として，入力された姓を認識する．
5. 認識結果 1 位の姓の正誤を確認する．

(3) 先頭漢字の読み仮名評価プログラム

1. 先頭漢字の読み仮名の入力を要求する．
2. 先頭漢字の読み仮名 1,190 種を対象として，入力された先頭漢字の読み仮名を認識する．

3. 姓の入力を要求する．
4. 認識結果 1 位の読みを持つ漢字を先頭に用いる姓のみを対象として，入力された姓を認識する．
5. 認識結果 1 位の姓の正誤を確認する．

例えば，文字数を利用した評価プログラム（1）では，文字数の認識結果 1 位が“3 文字”ならば，3 文字の姓 22,210 種のみを対象として入力姓を認識し，認識結果 1 位の姓の正誤を確認する．すなわち，属性値の認識結果が誤りである場合，絞り込んだ対象の中に入力姓は含まれていないため，正解姓は出現しない．

5.3.2 有効度評価実験

本節では，属性の有効度の有用性検証のための実験について述べる．5.3.1 節で実装した評価プログラムに対して，被験者に個人姓 20 件と属性値の入力を依頼した．20 件の姓はランダムに選択した．被験者は，4.1.1 節とは異なる 20 代から 30 代の男女 10 名ずつの計 20 名である．20 名はいずれも，音声認識技術や対話システムに関する知識を持っていない．被験者には，入力を依頼する 20 件の姓のリストと各姓の文字数，頭文字，先頭漢字の読み仮名の属性値のリストを予め渡す．表 5.7 に，20 件の姓，及び属性値のリストを示す．同じ条件で評価するために，入力姓と各属性値は被験者の発声を録音したものを，評価プログラムの入力に用いた．

表 5.7：評価姓，及び属性値リスト

姓	姓（読み）	(1) 文字数	(2) 頭文字	(3) 先頭漢字の読み仮名
加藤	かとう	さんもじ	か	くわえる
高橋	たかはし	よんもじ	た	たかい
大森	おおもり	よんもじ	お	だい
中村	なかむら	よんもじ	な	ちゅう
増田	ますだ	さんもじ	ま	ます
影山	かげやま	よんもじ	か	かげ
久保	くぼ	ふたもじ	く	ひさしい
糸井	いとい	さんもじ	い	いと
竹林	たけばやし	ごもじ	た	たけ
迫	さこ	にもじ	さ	はく
喜多丸	きたまる	よんもじ	き	よろこぶ
小座	おざ	ふたもじ	お	ちいさい
茂	しげり	さんもじ	し	しげる
苗代沢	なしろざわ	ごもじ	な	なえ
唐	から	にもじ	か	とう
碇野	いかりの	よんもじ	い	いかり
宇和	うわ	にもじ	う	う
増棟	ますむね	よんもじ	ま	ふえる
上辻	かみつじ	よんもじ	か	うえ
和佐野	わさの	さんもじ	わ	かず

5.3.3　評価実験結果

表 5.8 から表 5.10 に，各姓について，被験者の入力に対する評価プログラムの出力結果をまとめる. 各表には, 正しい属性値が認識結果 1 位に出現した被験者数（属性値正解人数）とその割合（属性値認識率)，正解姓が認識結果 1 位に出現した被験者数（姓正解人数）と, 属性値正解者のうち正解姓が認識結果 1 位に出現した被験者の割合（姓正解率），及び式 (5.3) を用いて計算した有効度の値が示してある．文字数を用いた絞り込みの結果が表 5.8，頭文字を用いた結果が表 5.9，先頭漢字の読み仮名を用いた結果が表 5.10 である．図 5.1 は，3 種の属性について有効度と正解率の関係をグラフに示したものである.

各表から，すべての属性において，有効度の大きい属性ほど，高い姓正解率が得られていることが確認できる. 5.3.1 節の最後に述べたように, 評価プログラムでは, 何文字の姓か，頭文字は何かなど，属性値が正しく認識できなければ，正解姓を導き出すことはできない. 実験において, 頭文字を用いた場合の平均姓正解率は 60.7% であり，文字数の平均姓正解率 29.9%の約 2 倍であるのに，平均姓正解人数は，文字数は 4.0 人，頭文字は 4.2 人とほぼ同じ結果となっている．これは，有効度の値から判断すると，頭文字は，文字数と比較して，個人姓に対して約 2 倍の絞り込み効果を持つことになるが，頭文字の属性値認識率が非常に低いため，有効度の値通りの姓正解人数が獲得できていない．このことから，有効度の値通りの絞り込みを実現するためには，属性値誤認識による姓の絞り込み失敗を回避する必要がある.

次節では，属性値の認識率向上のための改善策を提案する.

表 5.8：文字数を用いた有効度評価結果

姓	有効度	属性値認識結果		姓認識結果	
		属性値正解人数(人)	属性値認識率(%)	姓正解人数(人)	姓正解率(%)
かとう	0.35	12	60.0	3	25.0
たかはし	0.15	14	70.0	4	28.6
おおもり	0.15	11	55.0	4	36.4
なかむら	0.15	13	65.0	4	30.8
ますだ	0.35	14	70.0	4	28.6
かげやま	0.15	11	55.0	4	36.4
くぼ	0.93	9	45.0	3	33.3
いとい	0.35	12	60.0	3	25.0
たけばやし	0.51	16	80.0	6	37.5
さこ	0.93	11	55.0	3	27.3
きたまる	0.15	15	75.0	3	20.0
おざ	0.93	13	65.0	7	53.9
しげり	0.35	15	75.0	5	33.3
なしろざわ	0.51	18	90.0	5	27.8
から	0.93	13	65.0	4	30.8
いかりの	0.15	16	80.0	4	25.0
うわ	0.93	11	55.0	3	27.3
ますむね	0.15	17	85.0	5	29.4
かみつじ	0.15	15	75.0	3	20.0
わさの	0.35	14	70.0	3	21.4
平均	－	13.5	67.5	4.0	29.9

表 5.9：頭文字を用いた有効度評価結果

姓	有効度	属性値認識結果		姓認識結果	
		属性値正解人数(人)	属性値認識率(%)	姓正解人数(人)	姓正解率(%)
かとう	0.58	5	25.0	3	60.0
たかはし	0.68	6	30.0	4	66.7
おおもり	0.65	9	45.0	5	55.6
なかむら	0.75	7	35.0	4	57.1
ますだ	0.77	6	30.0	4	66.7
かげやま	0.58	5	25.0	3	60.0
くぼ	1.78	6	30.0	5	83.3
いとい	0.62	5	25.0	3	60.0
たけばやし	0.68	12	60.0	8	66.7
さこ	0.76	8	40.0	5	62.5
きたまる	0.80	8	40.0	5	62.5
おざ	0.65	8	40.0	5	62.5
しげり	0.67	7	35.0	4	57.1
なしろざわ	0.75	9	45.0	5	55.6
から	0.58	4	20.0	3	75.0
いかりの	0.61	7	35.0	4	57.1
うわ	0.73	6	30.0	3	50.0
ますむね	0.77	7	35.0	3	42.9
かみつじ	0.58	8	40.0	5	62.5
わさの	0.97	6	30.0	3	50.0
平均	－	7.0	34.8	4.2	60.7

表 5.10：先頭漢字の読み仮名を用いた有効度評価結果

姓	有効度	属性値認識結果		姓認識結果	
		属性値正解人数 (人)	属性値認識率 (%)	姓正解人数 (人)	姓正解率 (%)
かとう	1.26	14	70.0	12	85.7
たかはし	1.20	14	70.0	12	85.7
おおもり	1.04	15	75.0	13	86.7
なかむら	1.45	15	75.0	13	86.7
ますだ	1.39	14	70.0	12	85.7
かげやま	1.38	14	70.0	11	78.6
くぼ	1.19	13	65.0	12	92.3
いとい	1.48	12	60.0	11	91.7
たけばやし	1.29	12	60.0	9	75.0
さこ	1.15	11	55.0	9	81.8
きたまる	1.26	12	60.0	10	83.3
おざ	1.04	13	65.0	10	76.9
しげり	1.37	13	65.0	11	84.6
なしろざわ	1.52	14	70.0	9	64.3
から	1.01	11	55.0	9	81.8
いかりの	1.67	15	75.0	13	86.7
うわ	1.09	11	55.0	9	81.8
ますむね	1.42	7	35.0	5	71.4
かみつじ	1.09	15	75.0	12	80.0
わさの	1.50	12	60.0	11	91.7
平均	―	12.9	64.3	10.7	82.6

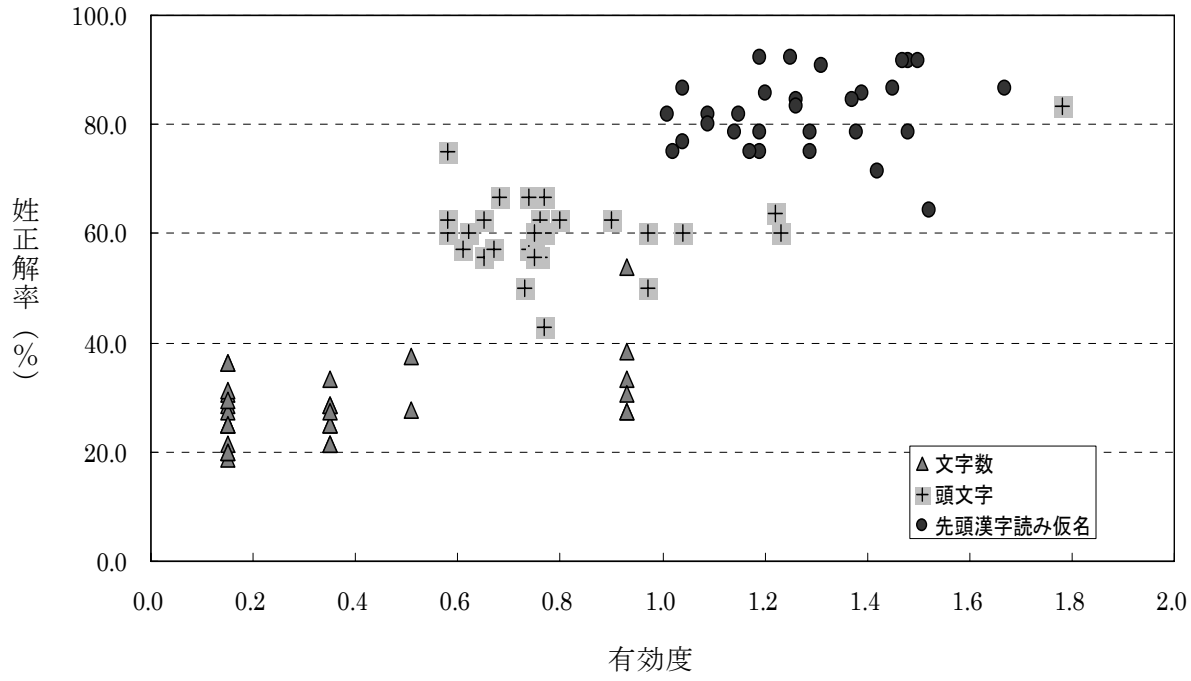

図 5.1：有効度と姓正解率の関係

5.3.4　精度向上のための改善策

本節では，属性値の認識率向上のための 2 種の改善策を提案し，評価を通して，改善策の有効性を検証する．

5.3.4.1　属性値の採用範囲拡大

5.3.3 節の実験結果において，正しい属性値が認識結果 1 位に出現しなかった被験者の入力に着目する．Nuance が最大尤度を算出した属性値が誤りであっても，認識尤度が大きい上位 10 種の属性値の中に正解が出現している場合が多い．いずれも，1 位の属性値と正解属性値の尤度は僅差である．

そこで，評価プログラムにおいて，認識結果 1 位の属性値のみを採用するのではなく，認識結果 1 位との尤度差に閾値 X を設定して，尤度差が X 以内の属性値を採用し再実験した．すなわち，1 位との尤度差が X 以内の属性値を持つ姓のみを対象として，入力姓を認識し，認識結果 1 位の姓の正誤を判断した．3 種の属性について，属性値正解人数と属性値正解率，姓正解人数と姓正解率を求め，表 5.11 に文字数を用いた場合の結果，表 5.12 に頭文字を用いた場合の結果，表 5.13 に先頭漢字の読み仮名を用いた場合の結果を示す．図 5.2 に，3 種の属性について有効度と正解率の関係をグラフに示す．予備実験に基づいて，実験では X= 3.0 とした．

再実験の結果，認識結果 1 位のみを属性値として採用した場合と比較して，すべての属性において，平均属性値正解人数の増加がみられ，それに伴い，姓正解人数の増加，及び姓正解率の向上も確認できる．特に，頭文字の平均属性値正解人数は，7.0 人から 15.6 人と大幅に増加した．この結果から，属性値の採用範囲拡大は，有効度の値を反映した絞り込みの実現に有用であることが確認できた．

表 5.11：尤度差 X 以内の文字数を採用した場合の有効度評価結果

姓	有効度	属性値認識結果		姓認識結果	
		属性値正解人数 (人)	属性値認識率 (%)	姓正解人数 (人)	姓正解率 (%)
かとう	0.35	18	90.0	7	38.9
たかはし	0.15	18	90.0	8	44.4
おおもり	0.15	20	100.0	7	35.0
なかむら	0.15	18	90.0	7	38.9
ますだ	0.35	18	90.0	8	44.4
かげやま	0.15	20	100.0	7	35.0
くぼ	0.93	20	100.0	8	40.0
いとい	0.35	20	100.0	6	30.0
たけばやし	0.51	20	100.0	8	40.0
さこ	0.93	20	100.0	8	40.0
きたまる	0.15	18	90.0	5	27.8
おざ	0.93	20	100.0	8	40.0
しげり	0.35	19	95.0	7	36.8
なしろざわ	0.51	20	100.0	7	35.0
から	0.93	20	100.0	7	35.0
いかりの	0.15	19	95.0	6	31.6
うわ	0.93	19	95.0	6	31.6
ますむね	0.15	18	90.0	7	38.9
かみつじ	0.15	20	100.0	5	25.0
わさの	0.35	20	100.0	6	30.0
平均	－	19.3	96.3	6.9	35.9

表 5.12：尤度差 X 以内の頭文字を採用した場合の有効度評価結果

姓	有効度	属性値認識結果		姓認識結果	
		属性値正解人数 (人)	属性値認識率 (%)	姓正解人数 (人)	姓正解率 (%)
かとう	0.58	12	60.0	10	83.3
たかはし	0.68	18	90.0	14	77.8
おおもり	0.65	15	75.0	11	73.3
なかむら	0.75	18	90.0	13	72.2
ますだ	0.77	16	80.0	12	75.0
かげやま	0.58	16	80.0	12	75.0
くぼ	1.78	15	75.0	12	80.0
いとい	0.62	15	75.0	10	66.7
たけばやし	0.68	18	90.0	12	66.7
さこ	0.76	15	75.0	11	73.3
きたまる	0.80	17	85.0	12	70.6
おざ	0.65	17	85.0	10	58.8
しげり	0.67	15	75.0	11	73.3
なしろざわ	0.75	15	75.0	11	73.3
から	0.58	13	65.0	9	69.2
いかりの	0.61	14	70.0	10	71.4
うわ	0.73	16	80.0	11	68.8
ますむね	0.77	17	85.0	12	70.6
かみつじ	0.58	15	75.0	12	80.0
わさの	0.97	15	75.0	12	80.0
平均	－	15.6	78.0	11.4	73.0

表 5.13：尤度差 X 以内の先頭漢字の読み仮名を採用した場合の有効度評価結果

姓	有効度	属性値認識結果		姓認識結果	
		属性値正解人数 (人)	属性値認識率 (%)	姓正解人数 (人)	姓正解率 (%)
かとう	1.26	17	60.0	15	88.2
たかはし	1.20	18	70.0	16	88.9
おおもり	1.04	17	55.0	16	94.1
なかむら	1.45	20	65.0	20	100.0
ますだ	1.39	17	85.0	15	88.2
かげやま	1.38	15	75.0	15	100.0
くぼ	1.19	17	85.0	15	88.2
いとい	1.48	16	80.0	15	93.8
たけばやし	1.29	16	80.0	14	87.5
さこ	1.15	12	60.0	11	91.7
きたまる	1.26	17	85.0	15	88.2
おざ	1.04	17	85.0	14	82.4
しげり	1.37	16	80.0	15	93.8
なしろざわ	1.52	18	90.0	16	88.9
から	1.01	16	80.0	14	87.5
いかりの	1.67	17	85.0	16	94.1
うわ	1.09	14	70.0	11	78.6
ますむね	1.42	15	75.0	13	86.7
かみつじ	1.09	17	85.0	15	88.2
わさの	1.50	17	85.0	16	94.1
平均	－	16.5	76.6	14.9	90.2

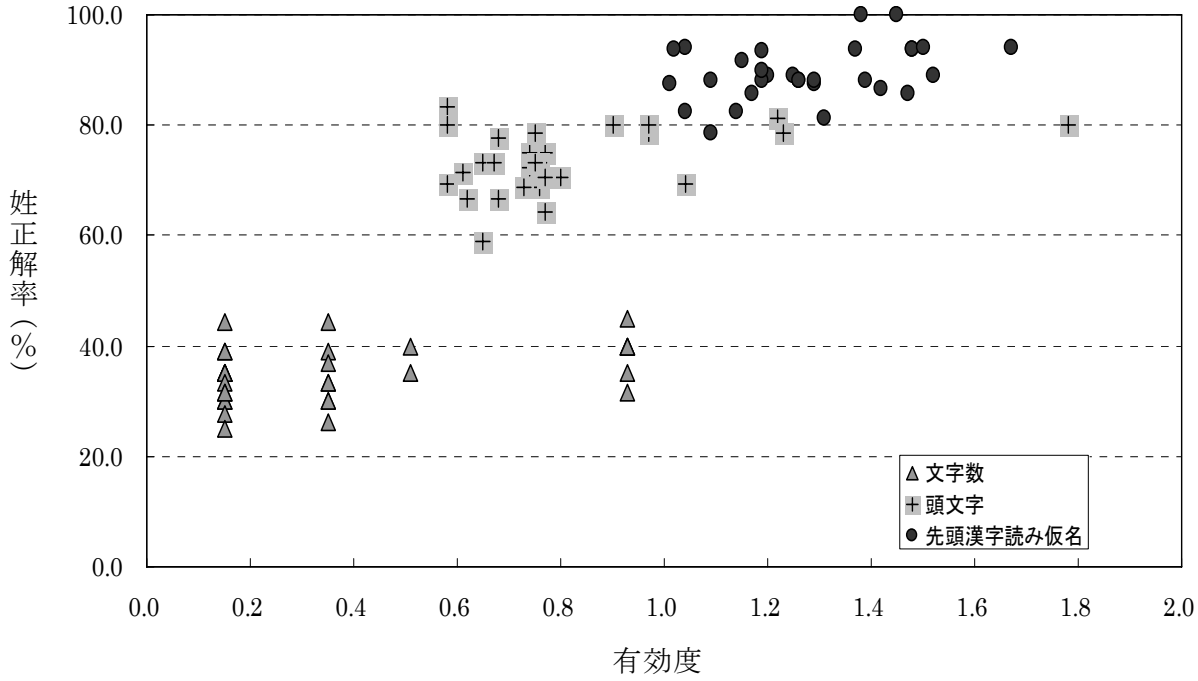

図 5.2：尤度差 X 以内の属性値を正解とみなした場合の有効度と姓正解率の関係

5.3.4.2　属性値の入力方法の改善

前節では，属性値の採用範囲拡大が属性値正解人数の増加に有効であることを確認したが，認識尤度は，利用者の発話状況や周囲の雑音などによって変化するため，どのような環境下でも，認識結果 1 位との尤度差が X の範囲内に正解属性値が出現するとは限らない．属性値正解人数増加のためには，属性値そのものの認識率を向上させる必要がある．そこで，本節では，5.3.2 節の実験において，属性値認識率が低い頭文字に焦点を当て，認識率向上のための改善策を提案する．

5.2.1 節において，属性の有効度は，確定難易度$D(A)$と曖昧性減少度合い$I(A_i)$を用いて定義した．複雑度の逆数である確定難易度$D(A)$は，式（5.1）から分かるように，0.1 の相違でも対象の複雑さに大きく影響する．複雑度は，3.1.2 節で述べたように，認識対象語の音韻数を考慮して，対象語を構成するすべての音韻の複雑さの平均値で定義する．しかし，孤立単語音声認識では，類似性を判断できる要素

が多い音韻数の多い単語ほど，高精度に認識できることが知られている．特に，頭文字のような平仮名 1 文字は，発話の始端に当たる子音の認識が困難であるため，母音のみが結果に出現する場合が多く，誤認識となる場合が多い[15，16]．

5.1 節で述べたコールセンタにおけるオペレータと利用者との対話では，オペレータがなかなか聞き取れなかった姓については，「さしすせそのすの“すだ（須田)”です」，「たちつてとのちの“ちはら（千原)”です」のように，利用者がオペレータに対して，頭文字を用いて自分の姓を説明する場面がみられた．そこで，音韻数の少ない単語は認識が困難であるという孤立単語音声認識の性質と，オペレータと利用者との対話分析に基づいて，頭文字を，“あいうえおのあ”，“かきくけこのか”のように，平仮名の行を前に付けた形で入力するように被験者に依頼し再実験した．被験者は 5.3.2 節と同じ 20 名である．頭文字と区別するために，この属性を行付頭文字と呼ぶ．行付頭文字の確定難易度D (A)は 0.09，各属性値の曖昧性減少度合いI (A_i)は，表 5.5 の頭文字の曖昧性減少度合いと同じになる．や行については“やゆよのや”，“やゆよのゆ”，“やゆよのよ”，拗音も同様に“きゃきゅきょのきゃ”，“じゃじゅじょのじょ”のように，被験者に発声を指示し，被験者の発声の録音を評価プログラムに入力した．入力姓は，5.3.2 節の評価実験で用いた被験者 20 名の録音音声を再利用した．表 5.14 に，認識結果 1 位との尤度差がX 以内の行付頭文字を採用したプログラムの実行結果を示す．実験において尤度差は，前節と同様にX = 3.0 とした．

行付頭文字を用いた結果，頭文字と比較して，平均属性値正解人数が約 2 倍の 16.4 人に増加し，それに伴い，姓正解人数の増加，及び姓正解率の向上も確認できる．この結果から，行付頭文字は，有効度の値を反映した絞り込みの実現に有用な属性であることが確認できた．

表 5.14：尤度差 X 以内を採用した場合の行付頭文字の有効度評価結果

姓	有効度	属性値認識結果		姓認識結果	
		属性値正解人数 (人)	属性値認識率 (%)	姓正解人数 (人)	姓正解率 (%)
かとう	0.35	17	85.0	13	76.5
たかはし	0.41	18	90.0	14	77.8
おおもり	0.39	18	90.0	14	77.8
なかむら	0.46	17	85.0	13	76.5
ますだ	0.47	14	70.0	11	78.6
かげやま	0.35	15	75.0	11	73.3
くぼ	0.47	17	85.0	12	70.6
いとい	0.38	15	75.0	11	73.3
たけばやし	0.41	18	90.0	14	77.8
さこ	0.46	17	85.0	13	76.5
きたまる	0.49	16	80.0	12	75.0
おざ	0.39	18	90.0	12	66.7
しげり	0.41	16	80.0	13	81.3
なしろざわ	0.46	17	85.0	12	70.6
から	0.35	16	80.0	13	81.3
いかりの	0.38	15	75.0	11	73.3
うわ	0.44	16	80.0	12	75.0
ますむね	0.47	16	80.0	11	68.8
かみつじ	0.35	16	80.0	12	75.0
わさの	0.59	16	80.0	12	75.0
平均	－	16.4	82.0	12.3	75.0

5.4　音声対話システムへの属性を利用した対話の適用

5.3 節の評価実験を通して，属性の有効度は個人姓に対する絞り込み効果を表す尺度として有用であることが確認できた．複数の属性が存在する場合，有効度が大きい属性ほど，正解姓を導き出せる可能性が大きい．

4 章で提案した思い込み応答では，思い込み設定外の対象が発話された場合は誤認識となる．本研究では，誤認識修正のために，属性を尋ねる対話の適用を提案する．すなわち，思い込み応答の結果，誤認識の場合は，有効度が大きい属性を尋ね，尋ねた結果に基づいて大語彙の中から正解候補を選択し，選択した候補のみを対象として認識する．こうすることで，利用者に負担をかけずに誤認識が修正できる．

属性を尋ねる対話の特徴は，5.3.1 節の評価プログラムで示したように，利用者に属性値の正誤を確認せずに姓を絞り込む点にある．正しい属性値を確定してから認識対象姓を選択すれば，属性値誤認識による姓の絞り込み失敗は回避できる．しかし，属性値が誤認識の場合は，属性値の誤認識修正のための対話が必要になる．2.2.3 節で述べたように，タスク達成までの対話の回数や誤提示の回数，及び誤認識修正のための同じ質問の繰返しが，利用者満足度獲得に影響を与えるという見解から，本研究では，属性値の正誤は確認せずに，認識結果 1 位との尤度差 X 以内の属性値を採用する対話を提案する．

第 6 章
大語彙を対象とした音声対話システムの構築

本章では，思い込み応答と属性の有効度を利用して，大語彙を対象とした音声対話システム実現のための対話制御手法を提案する．提案手法を，個人姓 87,944 種の確定をタスクとした対話システムに適用する．実装，評価を通して，提案対話制御手法の有効性を検証する．評価では，大語彙を一度に認識対象とする現状の音声対話システムにおける対話手法，及び人間オペレータの対応との比較を行う．

6.1　個人姓確定対話制御手法

前章までに述べたように，思い込み応答は，人間同士の対話と同様の聞き取り傾向を利用者に提供し，思い込み設定外の対象が発話された場合は，誤認識となるが，属性を尋ね正解を導き出す対話によって，利用者に入力要求，提示確認を繰返さずに誤認識が修正できる．本節では，対象が大語彙であっても，利用者に負担を与えないシステムを実現するために，思い込み応答，及び属性を尋ねる対話を適用した対話制御手法を提案する．

6.1.1　思い込みが外れた場合への対応

音声対話システムでは，利用者の入力を 100%の精度で認識できないため，システムが意図した通りに対話を制御できず，タスクが達成できない場合もある．本節では，思い込み応答，及び属性を尋ねる対話制御が失敗に終わった場合に備え，思い込み設定外の語彙を対象とした認識処理をバックグラウンドで進め，その結果を併用することを提案する．

思い込み応答や属性を利用した対話の進行過程で，利用者に対する入力要求や提示確認の時間，及び利用者の回答時間など，認識エンジンが使用されていない時間を利用して，バックグラウンドでの認識処理を実行する．リアルタイムな応答を利

用者に提供するために，思い込み応答，及び属性を尋ねる対話が終了するまでに，バックグラウンドでの処理は完了させる必要がある．一方，精度の観点から考えると，思い込みに設定する対象は一部であり，設定外の大語彙を一度に認識対象としても高精度な結果は期待できない．そこで，バックグラウンド処理では，思い込み設定外の対象を高精度な応答が期待できる数ごとに分割し，複数回，認識することで，各分割に対する精度向上を実現する．1 つの分割に含まれる姓を小語彙に抑えるほど精度向上が期待できるが，思い込み応答，及び属性を尋ねる対話が終了するまでの認識エンジンが使用できる時間内に，処理が終了する分割数でなければならない．バックグラウンド処理の入力には，利用者の最初の姓入力をシステム内部に録音したものを用いる．利用者にはバックグラウンドでの処理は知らせない．

6.1.2 個人姓確定対話制御フロー

本節では，個人姓確定のための対話制御手法の詳細を述べる．個人姓を絞り込むための属性として，属性値数 8 の文字数，属性値数 100 の行付頭文字，属性値数 1,190 の先頭漢字の読み仮名を用いる．最初に実在頻度順位上位の姓に対して思い込み応答を実行する．思い込み応答の対象として選択した頻度上位の姓を思い込み対象と呼ぶ．

提案対話制御手法では，誤提示を可能な限り回避するために，認識エンジンが出力する尤度を用いて出力結果の信頼度を判断する．思い込み応答では，信頼度が高いと判断できる尤度を持つ候補が得られた場合は，思い込み対象が発話されたと判断し利用者に提示確認する．信頼度が高い候補が得られなかった場合は，思い込みが外れたと判断し属性を尋ねる対話を進める．属性を尋ね正解姓を導き出す対話では，採用した属性値を持つ個人姓のみに対象を限定して，入力姓を認識した際の尤度に着目し，信頼度が高い候補が得られた場合は提示確認，信頼度が高い候補が得られなかった場合は，別の属性を尋ねる対話を実行する．

以下，対話制御の手順を説明する．図 6.1 に，フローの概要を示す．

(Step1) 利用者に姓の入力を要求する．

(Step2) 頻度順位上位の姓を思い込み対象として認識エンジンに設定し，入力された姓を認識する．

(Step3) (Step2) の結果，1 位候補の尤度が閾値 α 以上の場合，利用者に 1 位候補を提示して正誤確認する．

(Step4) (Step3) において，利用者から提示が否定された場合，又は 1 位候補の尤度が閾値 α 未満の場合は，上位 n 候補各々について，文字数，行付頭文字，先頭漢字の読み仮名の有効度を計算し，3 種の属性の有効度の平均値を求める．n は (Step2) の 1 位候補との尤度差が X 以内の候補数を表す．

(1) $n \neq 0$ の場合
文字数，行付頭文字，先頭漢字の読み仮名の中で，平均有効度が最大の属性の入力を利用者に要求する．

(2) $n = 0$ の場合
平均有効度が最大の先頭漢字の読み仮名の入力を利用者に要求する．

(Step5) (Step4) で入力された属性値を認識し，出力結果 1 位からの尤度差が X 以内の属性値を持つ姓のみを認識対象として，最初に入力された姓の録音を再認識する．

(Step6) (Step5) の再認識の結果，1 位候補の尤度が閾値 β 以上ならば，利用者に 1 位候補を提示して正誤確認する．

(Step7) (Step6) において，利用者から提示が否定された場合，又は 1 位候補の尤度が閾値 β 未満の場合は，残った 2 種の属性のうち平均有効度が大きい属性の入力を利用者に要求する．

(Step8) (Step7) で入力された属性値を認識し，(Step5) と同様に，出力結果 1 位

からの尤度差が X 以内の属性値を持つ姓のみを認識対象として，最初に入力された姓の録音を再認識する．

(Step9) (Step8) の再認識の結果，1 位候補の尤度が閾値 γ 以上ならば，利用者に 1 位候補を提示して正誤確認する．

(Step10) (Step9) において，利用者から提示が否定された場合，又は 1 位候補の尤度が閾値 γ 未満の場合は，残りの属性の入力を利用者に要求する．

(Step11) (Step10) で入力された属性値を認識し，(Step5)，(Step8) と同様に，出力結果 1 位からの尤度差が X 以内の属性値を持つ姓のみを認識対象として，最初に入力された姓の録音を再認識する．再認識の結果，1 位候補の尤度が閾値 δ 以上ならば，利用者に 1 位候補を提示して正誤確認する．

(Step12) (Step3) から (Step11) までの処理と並行して，(Step2) で思い込み対象外とした姓を対象として，バックグラウンドで認識を実行する．(Step11) において，利用者が提示を否定した場合，又は 1 位候補の尤度が δ 未満の場合は，バックグラウンドでの認識結果を統合し，最大の尤度を持つ候補を利用者に提示して正誤確認する．その際，それまでのフローにおいて，利用者から否定された姓は候補から除外する．

(Step4)，(Step7) で，n 候補の各属性の平均有効度が同値の場合は，先頭漢字の読み仮名，行付頭文字，文字数の順に優先する．これは，全個人姓を対象とした場合の平均有効度[13]が大きい順である[14]．(Step12) において，バックグラウンドでの認識結果も誤提示の場合は，未確定のまま対話を終了する．実装の際は，87,944 種のすべての姓について，3 種の属性の有効度を予め計算しておく．先頭漢字の読み仮名の有効度とは，その姓について考えられるすべての先頭漢字候補が持つ読み仮名の平均有効度を意味する．そして，3 種の属性の各属性値を持つ姓のリストも予め用意する．例えば文字数の場合は，1 文字の姓から 8 文字の姓まで，8 個の属性値

[13] 各属性において，すべての属性値の有効度の和を属性値数で除算した値．
[14] 文字数の平均有効度=0.76, 行付頭文字の平均有効度=0.83, 先頭漢字の読み仮名の平均有効度=0.94 である．

について各属性値を持つ姓のリストを用意する．表 6.1 に，各姓の属性 3 種の有効度を計算したリストの一部を示す.

表 6.1：個人姓 87,944 種の属性 3 種の有効度

頻度順位	姓	属性値 (有効度)		
		文字数 (有効度)	行付頭文字 (有効度)	先頭漢字の読み仮名 (有効度)
1	さとう	3 文字 0.35	さ 0.46	佐，里，左 0.76
2	すずき	3 文字 0.35	す 0.54	鈴，鱸，須･･･ 0.82
3	たなか	3 文字 0.35	た 0.41	田，棚，太･･･ 0.87
4	たかはし	4 文字 0.15	た 0.41	高，多，鷹･･･ 0.98
5	わたなべ	4 文字 0.15	わ 0.59	渡，渉，航･･･ 1.09
6	いとう	3 文字 0.35	い 0.38	伊，井，夷･･･ 1.29
7	やまもと	4 文字 0.15	や 1.78	山，耶，野･･･ 0.89
:	:	:	:	:
:	:	:	:	:
87,943	うしまぎ	4 文字 0.15	う 0.44	牛 1.98
87,944	ぎゅうき	3 文字 0.35	ぎゅ 1.32	牛 1.98
平均	―	0.76	0.83	0.94

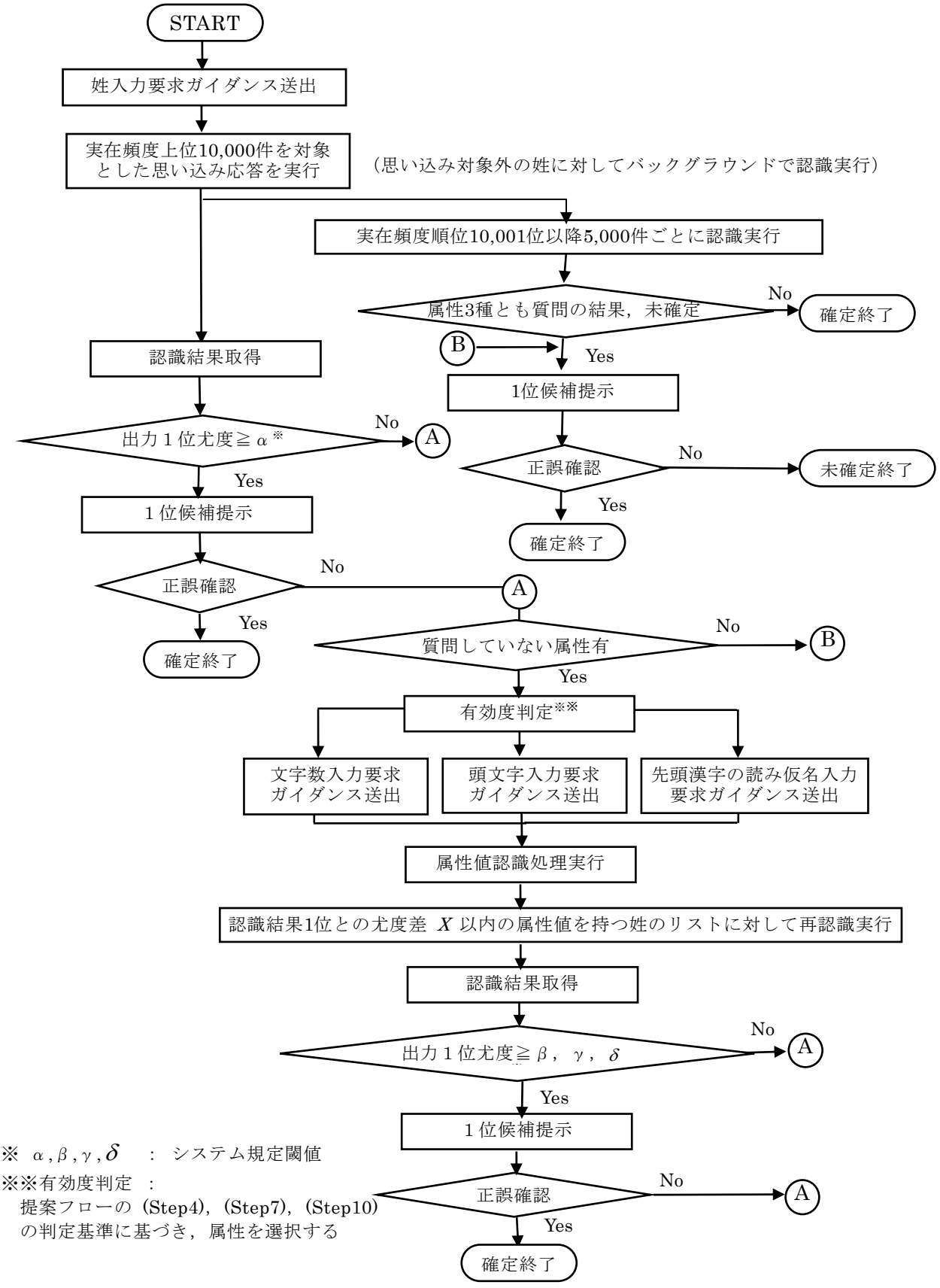

図 6.1：個人姓確定対話制御フロー

6.1.3 実装

本節では，6.1.2 節で提案した対話制御手法を用いて，個人姓 87,944 種の確定をタスクとした音声対話システムを実装する．以下，実装環境について述べる．

実装にはNuance7.03 を使用する．開発環境はMicrosoft Visual Basic 6.0，実行環境はWindows NT 4.0 Workstation が動作するDELL Poweredge2400 であり，回線ボードDialogicD / 41E-PCIとDSPボードDialogic AnteresIIを使用する．実装プログラムに対して，擬似回線エミュレーターTL1010[15]経由で電話機TEL-282[16]を接続した．4.2.1 節の分析に基づき，実在頻度順位上位 10,000 件の姓を思い込み対象として選択し，Nuance Grammar Builderを用いて認識対象辞書を作成する．N-best 値は 10 とする．予備実験に基づき，提案対話制御フローにおける閾値は表 6.2 のように設定する．

表 6.2：提案対話制御フローにおける閾値

閾値	値
α	70.0
β	70.0
γ	70.0
δ	70.0
X	3.0

実装では，(Step2) の終了と同時に，頻度 10,001 位以降の思い込み対象外の姓 77,944 種を対象とした認識処理をバックグラウンドで開始する．4.2.1 節で述べたように，Nuanceが実用レベルの精度を提供できる認識対象語彙数は，複雑度から考えると，約 10,000 語程度が限界である．仮に，バックグラウンドでの認識処理において，思い込み対象外の姓を 10,000 件ずつ 8 分割した場合，Nuanceは各分割の認識

[15] テクノシステム（株）製．
[16] (株) サンヨー製．

に 0.58 秒[17]，8 回の認識に計 4.64 秒を要する．5,000 件ずつ 16 分割した場合は，各分割に対して 0.42 秒，計 6.72 秒を要する．実装において，(Step11) が終了するまでに，バックグラウンド処理で認識エンジンが使用できる時間は，システムの入力要求や提示確認のガイダンス,及び利用者の回答時間を合計すると 7.42 秒である．この分析から，提案対話制御フローでは，リアルタイムかつ高精度な認識結果を提示するためには,思い込み対象外の姓を 5,000 件ずつ 16 分割することが効果的であると判断した．すなわち，提案対話制御フローの (Step12) では，16 回のバックグラウンドでの認識結果を尤度順に統合し，尤度が最大の姓を利用者に提示して正誤確認する．

6.2 評価実験

本節では，6.1.3 節で実装した個人姓確定対話制御フローに対する評価実験について述べる．実験では，大語彙一括認識フローと人間オペレータフローも実装し，比較を通して，提案対話制御手法の有効性を検証する．

6.2.1 大語彙一括認識フロー

大語彙一括認識フローとは，大語彙を一度に対象として認識し，誤認識の場合，利用者に対して入力要求，提示確認のみを交互に繰返す現状の音声対話システムにおける対話手法である．提案対話制御方式と同様に，Nuance7.03 を使用して，個人姓 87,944 種を対象とした大語彙一括認識フローを実装する．実装環境も 6.1.3 節と同様とする．予備実験に基づき，閾値 $\alpha = 70$ とする．

以下，大語彙一括認識フローの詳細を述べる．

(Step1) 利用者に姓の入力を要求する．

(Step2) 87,944 種の姓を対象として，入力された姓を認識する．

[17] 10,000 回の認識試験の平均CPU使用時間.

(Step3) (Step2) の結果，1 位候補の尤度が閾値 α 以上の場合は，利用者に 1 位候補を提示して正誤確認する．

(Step4) (Step3) において，利用者から提示が否定された場合，及び 1 位候補の尤度が閾値 α 未満の場合は，姓の再入力を利用者に要求する．

(Step5) 87,944 種の姓を対象として，(Step4) において再入力された姓を認識する．

(Step6) 再認識の結果，1 位候補を利用者に提示して正誤確認する．

(Step7) 提示が否定された場合，再々度，姓の入力を利用者に要求する．正解が提示できるまで (Step6)，(Step7) を繰返す．

6.2.2 人間オペレータフロー

人間オペレータフローとは，コールセンタにおいてオペレータ経験を持つ女性に被験者が入力する姓の聞き取りを依頼したものである．オペレータ役は，被験者に対して，姓を提示して正解であるという確認が得られるまで，一問一答の質疑応答を続ける．オペレータ役には，以下の 3 種の質問のみを許可する．同じ質問は連続してもよい．

(1) 「～様ですか?」という提示による正誤確認質問．

(2) 「もう一度，苗字をおっしゃって下さい」という姓の再入力要求．

(3) 姓の文字数，行付頭文字，先頭漢字の読み仮名を尋ねる質問．

6.2.3 評価実験概要

提案対話制御フロー，大語彙一括認識フロー，人間オペレータフローに対して 100 件ずつ，計 300 件の姓の入力を被験者に依頼する．被験者は，5.3.2 節における属性の有効度の有用性評価とは異なる男女 10 名ずつの計 20 名であり，音声認識技術や対話システムに関する知識は持っていない．各フローに対して，入力を依頼した姓は，頻度順位 1 位から 10,000 位までの思い込み対象内の姓 50 件と，10,001 位以降の思い込み対象外の姓 50 件の計 100 件である．各被験者には，フロー3 種分の思い込み対象内の姓 150 件と，対象外の姓 150 件のリストを予め渡す．リストには，“松本（まつもと)”，“菅原（すがわら)”のように，漢字表記と読み仮名のみを対に記し，思い込み対象 150 件，対象外 150 件の順に並べる．

被験者には，拗音は前の平仮名と併せて 1 文字と扱うことのみを事前に説明する．これは，選択した属性が，被験者にとって容易に回答可能であるかを評価するためであり，各姓の 3 種の属性については，システムからの入力要求ガイダンスを聞いて被験者自身が考える．各フローにおいて，姓の入力要求と正誤を問う提示確認に対する回答は必須とし，提案対話制御フロー，人間オペレータフローの属性を尋ねる質問に対してのみ「分かりません」という回答を許容した．また，被験者には，繰返しの入力要求や長時間の対話の継続に負担を感じるなどの理由から，システムとの対話を中止したいと感じた場合は，挙手による途中放棄を認めた．被験者が挙手をした時点，及び属性を尋ねる質問に対して「分かりません」と回答した時点で，その姓は未確定終了とし，次の姓の入力要求を開始する．すなわち，未確定終了する可能性のある場面は，以下のいずれかになる．

(1) 3 種のフローにおいて，被験者が挙手をして途中放棄を申し出た場合．

(2) 提案対話制御フロー，人間オペレータフローの属性を尋ねる質問に対して，被験者が「分かりません」と回答した場合．

(3) 提案対話制御フローにおいて，(Step12) のバックグラウンド処理の結果も誤提示の場合．

3 種のフローの実行を 1 セットとして，被験者 1 人あたり 100 セット，計 300 件の姓入力を実行する．各セットにおいて，3 種のフローはランダムに起動する．被験者には，起動したフローが 3 種のどれに当たるのかは知らせず，1 セット終了ごとに，3 種のうち最も人間に近いと感じたフローを 1 つ選択してもらう．各フローのシステムガイダンスは，人間オペレータフローで協力を依頼したオペレータ役の女性の録音音声を用いて作成した．すなわち，3 種のフローのシステムガイダンスの声質に差異はない．また，被験者には，リアルタイムに人間オペレータが対応しているフローの存在は知らせない．すべてのフローにおいて，姓の提示確認に対する被験者の正誤回答は 100%の精度で認識できるものとする．

6.3 評価実験結果と考察

表 6.3 から表 6.9 に，被験者 20 名の 300 件の姓入力に対する結果を示す．以下，実験結果の分析を通して，思い込み応答，及び誤認識修正のための属性を尋ねる対話の有効性について考察する．

6.3.1 思い込み応答の効果

被験者の入力に対する各フローの第一応答に着目する．第一応答とは，被験者の入力に対するシステムの最初の応答を意味する．提案対話制御フローでは，思い込み応答の結果，閾値以上の候補が得られた場合は，第一応答で姓を提示し，閾値以上の候補が得られなかった場合は，有効度が最大の属性の入力を要求する．一方，大語彙一括認識フローでは，認識の結果，閾値以上の候補が得られた場合は姓を提示し，得られなかった場合は姓の再入力を要求する．表 6.3 に，被験者 20 名の思い込み対象内の延べ 1,000 件の姓入力について，各フローの第一応答の内容を示す．同様に，思い込み対象外の延べ 1,000 件の姓についての結果を表 6.4 に示す．

思い込み応答の有効性を評価するために，各入力に対するシステムの第一応答について，受入れ可能か否かの判断を被験者に依頼した．実験では，3 種のフローの実行を 1 セットとして，被験者に人間の対話との類似性評価を依頼しているため，受入れ可否評価を各フローの第一応答が終了した時点で行うことは，類似性評価の妨げになると考えた．そこで，被験者の入力とシステムの第一応答を録音し，実験

後に第一応答のみを再現して，受入れ可否評価を依頼した．表 6.3，表 6.4 に，各フローの第一応答に対して，受入れ可能と判断された対話数を応答内容別に示す．

表 6.3：思い込み対象内の姓 1,000 件に対する第一応答

フロー	正解 受入れ可 (件)	誤提示 受入れ可 (件)	再入力要求 受入れ可 (件)	属性入力要求 受入れ可 (件)
提案対話制御フロー	692 692	0 —	—	308 101
大語彙一括認識フロー	212 212	621 161	167 19	—
人間オペレータフロー	713 713	89 13	32 7	166 55

表 6.4：思い込み対象外の姓 1,000 件に対する第一応答

フロー	正解 受入れ可 (件)	誤提示 受入れ可 (件)	再入力要求 受入れ可 (件)	属性入力要求 受入れ可 (件)
提案対話制御フロー	0 —	487 354	—	513 478
大語彙一括認識フロー	109 109	834 501	57 34	—
人間オペレータフロー	302 302	111 89	435 379	152 146

表 6.3 から, 思い込み対象内の姓 1,000 件に対する第一応答の精度に着目すると，提案対話制御フローは，大語彙一括認識フローよりも高精度であり，かつ人間オペレータフローとほぼ同じ精度であることが分かる．また，思い込み対象内の姓に対して，提案対話制御フローでは，第一応答が誤提示の対話は 0 件であり，第一応答で正解を提示できなかった 308 件は，すべて尤度が閾値以下であったため，被験者に対して有効度が最大の属性入力を要求している．一方，大語彙一括認識フローでは全体の約 6 割, 人間オペレータフローでも全体の約 1 割に該当する 89 件の第一応答が誤提示であった．この結果から，大語彙を一括で認識する場合と比較して，思い込み応答の適用によって，思い込み対象内の姓に対する精度向上が確認できる．

一方，思い込み対象外の姓 1,000 件に着目すると，提案対話制御フローでは，思い込み対象外の姓は，いずれも最初の認識対象に含まれていないため，第一応答で正解が出現することはない．487 件の第一応答が誤提示，513 件が属性入力を要求している．これに対して，大語彙一括認識フローでは，第一応答で正解が提示できたのは全体の僅か 1 割の 109 件のみであり，約 8 割は誤提示であった．また，人間オペレータフローでも，第一応答で正解が提示できたのは 3 割に当たる 302 件のみであり，111 件が誤提示，残り 587 件は再入力，又は属性入力を要求している．この結果から，大語彙を一度に認識対象とした場合，利用者に対して高精度な応答の提供は困難であること，また人間も思い込み対象外の姓が発話された場合，聞き間違いを起こしやすいという傾向が再確認できる．

また，第一応答に対する受入れ可否評価の結果に着目すると，思い込み対象内の姓の誤提示に対して，被験者が受入れ可能と判断した対話数は，大語彙一括認識フローでは 621 件のうち 26%に当たる 161 件，人間オペレータフローでは 89 件のうち 15%に当たる 13 件のみであり，思い込み対象内の姓に対する誤認識は利用者に負担を与えることが分かる．一方，思い込み対象外の姓の誤提示に対しては，提案対話制御フローでは，487 件の 73%に当たる 354 件が受入れ可能と判断され，誤認識であっても利用者の負担になっていない．大語彙一括認識フローでも 834 件のうち 60%に当たる 501 件，人間オペレータフローでも 111 件のうち 80%に当たる 89 件が受入れ可能と判断され, 提案対話制御フローと同様に, 思い込み対象外の姓は, 誤提示しても利用者に受入れられることが分かる. また, すべてのフローにおいて, 思い込み対象外の姓は，対象内の姓と比較して，正解が提示できずに再入力，又は属性入力を要求しても，受入れ可能と判断される割合が高いことが確認できる．

これらの結果から，思い込み応答は，思い込み対象内の姓に対しては高精度な第一応答が提供できること，かつ思い込み対象外の姓に対する誤認識は利用者に負担

を与えないことが検証できた.

6.3.2 属性を利用した対話の効果

次に，思い込み対象内の姓 1,000 件について，確定終了した対話数とそれらの確定までに要した平均ターン数，被験者から途中放棄された対話数と平均放棄箇所を表 6.5 に示す. 同様に, 思い込み対象外の姓についての結果を表 6.6 に示す. 表 6.5，表 6.6 では，被験者の途中放棄以外の未確定終了[18]対話はカウントしていない．本実験においても，入力要求や提示確認，属性を尋ねる質問など，システムから被験者への入力要求と，それに対する被験者の回答の組をターンと呼ぶ．すべてのフローにおいて，最初の姓の入力要求とそれに対する被験者の姓入力はターンに含めず，次のシステムの要求から 1 ターン目をカウントする.

大語彙一括認識フローにおいて，確定終了した対話の確定までのターン数に着目する．思い込み対象内の姓 1,000 件のうち，1 ターン目で正解が提示できた対話は 212 件，対象外の姓は 109 件のみであり，約半数に当たる 937 件の対話は被験者から途中放棄されている．このことから，入力要求と提示確認の繰返しは，被験者に負担を与えることが確認できる．大語彙一括認識フローにおいて，確定終了した対話の確定までの平均ターン数は，思い込み対象内の姓が 2.75 ターン，対象外の姓が 2.97 ターンであることから，誤提示の後，被験者に再入力を要求し，再認識した結果，正解が提示できた対話が多いことが分かる．途中放棄された対話の平均放棄箇所は，思い込み内外ともに 5 ターン目以降である．これは，誤提示，姓の再入力要求，誤提示，姓の再々入力を要求した結果，それでも誤提示であったため，姓の入力を再々々度要求した時点での放棄と予想できる．したがって，被験者にとって，姓の再入力要求は連続 2 回までが許容範囲と言えるであろう.

一方，人間オペレータフローでは，思い込み対象外の姓は，対象内と比較して，確定までに約 2 倍のターンを要している．そして，人間オペレータが対応しても，思い込み対象外の姓については 125 件の途中放棄がみられる．途中放棄された平均箇所は，人間オペレータフローは 8 ターン目以降であり，大語彙一括認識フローと比較して，人間オペレータフローでは被験者との対話が継続している．つまり，大語彙一括認識フローでは，入力要求，提示確認のみを交互に繰返すのに対して，人

[18] 6.2.3 節の未確定終了となりうる場面の (2)，(3) に相当.

間オペレータは，被験者を飽きさせないように対話を継続させることができる．人間オペレータフローにおいて，途中放棄された思い込み対象外の姓 125 件の中には，先頭漢字の読み仮名を用いた絞り込みは困難であるのに，オペレータは，先頭漢字の読み仮名を尋ねたために，かえって混乱してしまい対話が継続した例もみられた[19]．

これに対して提案対話制御フローでは，最後までフローは実行されたが未確定のまま終了した対話が，思い込み対象内に 51 件，対象外に 159 件の計 210 件存在する．この 210 件は，先頭漢字の読み仮名を尋ねたところ，平仮名 1 文字が答えられた姓であり，孤立単語音声認識の性質から，正しい先頭漢字の読み仮名が認識結果 1 位との尤度差 X の範囲内に出現していない．これらは，“佐藤（さとう)”，“佐伯(さえき)”，“佐田（さだ)”，“伊達（だて)”，“伊藤（いとう)”，“伊土（いど)”，“須田（すだ)”，“須藤（すどう)”など，平仮名 1 文字の音読みしか持たない漢字で始まる姓であり，正しい属性値が採用できずに正解姓を導き出すことができなかった．そして，文字数，行付頭文字を用いた絞り込みも，ともに失敗に終わっている．そこで，これら 210 件を人間オペレータフローに入力し，人間の聞き取り傾向を調べた．入力は，被験者の 1 人に依頼した．その結果，約半数に当たる 103 件は平均 3.41 ターンで確定終了し，残りの 107 件は被験者の途中放棄によって未確定のまま終了した．107 件の平均放棄箇所は 7.21 ターン目であった．したがって，これら 210 件は，人間でも確定までにターンを要する姓であり，約半数については未確定のまま終了するほど，聞き取りが困難であることが分かる．

また，思い込み対象内の姓 7 件，対象外の姓 5 件の計 12 件は，先頭漢字の読み仮名を尋ねたところ，「分かりません」と回答されたため，その時点で未確定終了した．これらは，“服部（はっとり)”，“和泉（いずみ)”，“東海林（しょうじ)”，“九十九(つくも)”，“九石（さららし)”などであり，被験者は，先頭漢字の読み仮名の入力要求を，対象姓の中で先頭漢字が何という読みに該当するか[20]を答える質問であると解釈したため，分からないという回答に繋がったと予想できる．

提案対話制御フローにおいて注目すべき結果は，被験者から途中放棄された対話が，思い込み対象内，対象外ともに 1 件も存在しない点である．さらに，思い込み対象内の姓については，確定までの平均所要ターン数は，人間オペレータフローと

---

[19] “小鳥遊（たかなし)”（小鳥が自由に飛んで遊べる空には鷹がいないという意味が由来)，“月見里(やまなし)”（高い山がないところでは月が見えるという意味が由来)，“四月一日（わたぬき)”，“八月一日（ほづみ)”（暦の上での行事が由来）など．

[20] 例えば，和泉（いずみ）の中で，和は“い”，“いず”，“いずみ”のどれに該当するのか．

ほぼ同じである．一方，思い込み対象外の姓については，1 ターン目で正解が提示できた対話は存在しないにもかかわらず，属性を利用した対話によって，平均 3.30 ターンですべての姓が確定できている．また，思い込み対象外の姓の中には，最長 7 ターンまで継続した対話が 19 件存在する．この結果から，提案対話制御フローは，思い込み対象内の頻出姓よりも，対象外の希少姓に対しての方が確定までにターンを要する点，かつ被験者を飽きさせない方向へ対話を誘導する点で，人間オペレータフローと同様である．

表 6.5：思い込み対象内の姓 1,000 件の結果

フロー	確定終了 (件)	確定 平均ターン数	途中放棄 (件)	途中放棄 平均ターン数
提案対話制御フロー	942	1.53	0	－
大語彙一括認識フロー	611	2.75	389	5.06
人間オペレータフロー	1,000	1.30	0	－

被験者からの途中放棄以外の未確定終了はカウントしていない．

表 6.6：思い込み対象外の姓 1,000 件の結果

フロー	確定終了 (件)	確定 平均ターン数	途中放棄 (件)	途中放棄 平均ターン数
提案対話制御フロー	836	3.30	0	－
大語彙一括認識フロー	452	2.97	548	5.27
人間オペレータフロー	875	2.44	125	8.07

被験者からの途中放棄以外の未確定終了はカウントしていない．

次に，途中放棄された対話の特徴を分析するために，途中放棄された箇所に着目する．被験者にとっては，少ないターンで正解が提示されることが好ましいが，提案対話制御フローでは，最長 7 ターンの対話が存在することから，ターンの長さのみが途中放棄の要因ではないと予測した．表 6.7，表 6.8 に，各フローにおいて，姓の入力を連続して要求した対話数を連続回数別に示す．連続回数が 2 回とは，被験者の姓の入力に対して認識した結果，閾値以上の侯補が出力されない，あるいは誤提示であったため，姓の再入力を要求した場合に相当する．表中の括弧は，姓の再入力を要求した時点で，被験者が放棄した対話数を示す．なお，提案対話制御フローについては，アルゴリズム上，姓の入力を連続して要求することはないので割愛する．

表 6.7：思い込み対象内の姓 1,000 件の再入力の連続要求

フロー	2 回 (件)	3 回 (件)	4 回 (件)	5 回 (件)	6 回 (件)	計 (件)
大語彙一括認識フロー	275 (0)	345 (162)	220 (198)	29 (29)	0 －	－ (389)
人間オペレータフロー	0	0	0	0	0	－

括弧内の数字は，姓の再入力を要求した時点で，被験者が放棄した対話数を表す．

表 6.8：思い込み対象外の姓 1,000 件の再入力の連続要求

フロー	2 回 (件)	3 回 (件)	4 回 (件)	5 回 (件)	6 回 (件)	計 (件)
大語彙一括認識フロー	172 (0)	389 (281)	220 (202)	54 (52)	13 (13)	－ (548)
人間オペレータフロー	354 (40)	130 (30)	55 (55)	0 －	0 －	－ (125)

括弧内の数字は，姓の再入力を要求した時点で，被験者が放棄した対話数を表す．

表 6.7，表 6.8 から，人間オペレータフローでは，提案対話制御フローと同様に，思い込み対象内の姓については，姓の入力を連続して要求した対話は 1 件も存在しない．つまり，オペレータは聞き取れなかった姓については，再入力を要求するのではなく，属性を尋ねる方向へ対話を誘導する．大語彙一括認識フロー，及び人間オペレータフローの思い込み対象外の姓について，再入力を要求した時点で放棄された対話数を合計すると，表 6.5，表 6.6 の途中放棄された対話数に一致する．さらに，繰返し回数が多くなるほど，放棄される対話の割合が増えている．この結果から，タスク達成までの対話の回数のみではなく，同じ質問の繰返しが被験者の満足度に大きく影響していることが分かる．

これらの結果から，提案対話制御手法における属性を尋ねる対話は，姓の入力要求を繰返さずに正解を導き出せることから，利用者を飽きさせず，ターンが長くなっても利用者に受入れられることが確認できた．

6.3.3 人間の対応との類似性

表 6.9 に，3 種のフローの実行を 1 セットとした計 100 セットについて，人間の対話との類似性評価の結果を示す．思い込み対象内の姓 1,000 件については，人間オペレータフローが最も人間に近いという評価が最多であり，提案対話制御フローが，最も人間らしいと評価された対話も 187 件存在する．この 187 件は，1 ターン目で正解姓が提示できた対話が約半数であり，残りの半数は，閾値以上の候補が出現しなかったため，先頭漢字の読み仮名を尋ねた結果，正解が提示できた対話である．この結果から，先頭漢字の読み仮名を尋ねる対話は，被験者にとって違和感なく自然に受入れられることが分かる．一方，思い込み対象外の希少姓については，提案対話制御フローが最も人間に近いという評価が約半数得られた．これは，提案対話制御フローの約半数は，正誤確認，再入力要求，及び 3 種の属性を尋ねる質問のみに発話を制限した人間の対応と区別がつかないと評価されたことを意味する．

実験を通して，提案対話制御手法は，特に思い込みが外れた場合に有効であり，属性を尋ね正解を導き出す対話は，利用者満足度が獲得できることが確認できた．

表 6.9：最も人間に近いと思われたフローの選択結果

フロー	思い込み対象内の姓（件）	思い込み対象外の姓（件）
提案対話制御フロー	187	407
大語彙一括認識フロー	0	0
人間オペレータフロー	813	593

提案対話制御手法では，先頭漢字として常用漢字のみを対象としている．そのため，先頭が常用漢字ではない 4,990 種の姓については，漢字の読み仮名を利用しても絞り込めない．評価実験では，これら 4,990 種の姓は入力対象としていない．この 4,990 種の姓のうち，約 8 割は旧字体，異字体であり，同じ読みを持つ代替漢字が常用漢字に存在するため，漢字の読み仮名を用いた絞り込みが可能であるが，残りの姓については対応が必要になる．同じ読みを持つが常用漢字が存在しない漢字を先頭に持つ姓について，文字数と行付頭文字のみを用いて評価したところ，行付頭文字を用いて 92%の姓が確定できた．したがって，これらについては，文字数，又は行付頭文字を尋ねる対話を継続し，2 種の属性を尋ねても確定できない場合は，予め先頭が常用漢字ではない姓のリストを作成しておき，それらを対象として，再認識を試みる手法の適用が考えられる．

第 7 章
住所への対話制御手法の適用

本章では，6 章で提案した大語彙確定のための対話制御手法を住所確定対話システムに適用する．提案手法では，人間と同様の聞き取り傾向を実現するための思い込み対象の選択，及び利用者が容易に回答可能であり，大語彙に対する絞り込み効果が大きい属性の選択が必要になる．本章では，住所確定に効果的な思い込み対象，及び属性の選択方法について述べる．

7.1　住所確定のための対話制御手法

7.1.1　思い込み対象の選択

日本全国には約 40 万の地名が存在する．全国の住所の確定をタスクとした対話システムを実現する場合，都道府県 47 種，市区郡 4,100 種，町村字 173,600 種の計 177,747 種類の地名が対象になる．個人姓と同様に，現状の音声認識技術では 177,747 種の地名を一度に対象としても，利用者に対して高精度な応答を提供することはできない．

3.2.2 節で述べたコールセンタへの 1 日のアクセスは，東京都内在住の利用者からが 77%，神奈川，千葉，埼玉の 3 県のいずれかに在住の利用者からが 19%と大きな偏りがある．同様に，資料送付を専門とする仙台市にあるコールセンタへの 1 日のアクセスを分析したところ，東北地方在住の利用者からのアクセスが 89%を占めた．コールセンタの受付け業務に住所確定対話システムを適用する場合，この偏りを利用した思い込み対象の選択が有効である．本研究では，東京 23 区在住の利用者は町村字名を最初に発話する場合が多く，それ以外の首都圏在住の利用者は市区郡名を最初に発話する場合が多いという 3.2.2 節で述べた傾向を踏まえ，都道府県 47 種とアクセスの多い市区郡，及び町村字名を思い込み対象として選択する．

7.1.2　住所確定に効果的な属性

本節では，住所確定において思い込みが外れた場合に尋ねる属性について考える．利用者の入力が都道府県名の場合は，思い込み対象として選択されているため，第一応答で正解が提示できる可能性が大きい．本研究では，利用者の入力が市区郡，町村字の場合は，残りの 2 階層の地名のうち有効度が大きい地名を属性として採用する．すなわち，市区郡については所属する都道府県，又はその市区郡内にある町村字, 町村字については所属する都道府県, 又は市区郡を属性として尋ねることで，思い込みが外れても正解が導き出せると考える．次節では，住所確定対話制御手法の詳細について説明する．

7.1.3　住所確定対話制御フロー

本節では，住所確定のための対話制御手法について述べる．都道府県，及びアクセス頻度の高い地名を思い込み対象として選択し，思い込み応答を実行する．思い込み応答の結果，信頼度の高い候補が得られた場合は，思い込み対象内の地名が入力されたと判断して利用者に提示確認する．信頼度の高い候補が得られなかった場合は，思い込みが外れたと判断して属性を尋ねる対話を進める．6.1.2 節の個人姓確定対話制御フローと同様に，誤提示を回避するために，認識エンジンが出力する尤度を用いて，出力結果の信頼度を判断する．

以下，対話制御の手順を説明する．

(Step1)　利用者に地名の入力を要求する．

(Step2)　アクセス頻度の高い地名を思い込み対象として選択し，入力された地名を認識する．

(Step3)　(Step2) の結果，1 位候補の尤度が閾値 α 以上の場合，利用者に 1 位候補を提示して正誤確認する．

(Step4)　(Step3) において，利用者から提示が否定された場合，又は 1 位候補の尤

度が閾値 α 未満の場合は，上位 n 候補各々に対して，都道府県，市区郡，町村字の有効度の平均値を求める．都道府県については，都道府県そのものの有効度は 0 とする．市区郡についても市区郡の有効度は 0，町村字についても町村字の有効度は 0 とする．n は (Step2) の 1 位候補の尤度との尤度差が X 以内の候補数を表す．

(1) $n \neq 0$ の場合
都道府県，市区郡，町村字の中で，平均有効度が最大の属性の入力を利用者に要求する．

(2) $n = 0$ の場合
平均有効度が最も大きい都道府県の入力を利用者に要求する．ただし，入力が都道府県の可能性を考慮して，都道府県を入力した場合は再度，都道府県名を入力するように誘導する．

(Step5) (Step4) で入力された属性値を認識し，出力結果 1 位からの尤度差 X 以内の属性値を持つ地名のみを認識対象として，最初に入力された地名の録音を再認識する．

(Step6) (Step5) の再認識の結果，1 位候補の尤度が閾値 β 以上ならば，利用者に 1 位候補を提示して正誤確認する．

(Step7) (Step6) において，利用者から提示が否定された場合，又は 1 位候補の尤度が閾値 β 未満の場合は，残った 2 種の属性のうち，平均有効度が大きい属性の入力を利用者に要求する．

(Step8) (Step7) で入力された属性値を認識し，(Step5) と同様に，出力結果 1 位からの尤度差が X 以内の属性値を持つ地名のみを認識対象として，最初に入力された地名の録音を再認識する．

(Step9) (Step8) の再認識の結果，1 位候補の尤度が閾値 γ 以上ならば，利用者に 1 位候補を提示して正誤確認する．

(Step10) (Step9) において，利用者から提示が否定された場合，又は 1 位候補の尤度が閾値 γ 未満の場合は，残りの属性の入力を利用者に要求する．

(Step11) (Step10) で入力された属性値を認識し，(Step5)，(Step8) と同様に，出力結果 1 位からの尤度差が X 以内の属性値を持つ地名のみを認識対象として，最初に入力された地名の録音を再認識する．再認識の結果，1 位候補の尤度が閾値 δ 以上ならば，利用者に 1 位候補を提示して正誤確認する．

(Step12) (Step3) から (Step11) までの処理と並行して，(Step2) で思い込み対象外とした地名を対象として，バックグラウンドで認識を実行する．(Step11) において，利用者が提示を否定した場合，又は 1 位候補の尤度が δ 未満の場合は，バックグラウンドでの認識結果を統合し，最大の尤度を持つ候補を利用者に提示して正誤確認する．それまでのフローにおいて，利用者から否定された地名は候補から除外する．

(Step4)，(Step7) で，n 候補の各属性の平均有効度が同値の場合は，3 種の属性のうち平均有効度が最大の都道府県名を尋ねる．(Step12) において，バックグラウンドでの認識結果も誤提示の場合は，未確定のまま対話を終了する．実装の際は，177,747 種のすべての地名について，残りの 2 階層の地名の有効度を予め計算しておく．都道府県を絞り込むための属性である市区郡の有効度とは，その都道府県内のすべての市区郡の有効度の平均値を意味する．同様に，町村字の有効度とは，その都道府県内のすべての町村字の有効度の平均値を意味する．市区郡を絞り込むために用いる都道府県の有効度とは，その市区郡が所属する都道府県の有効度であり，町村字の有効度は，その市区郡内の全町村字の有効度の平均値とする．町村字を絞り込むために用いる都道府県の有効度は，その町村字が所属する都道府県の有効度，市区郡の有効度は，その町村字が所属する市区郡の有効度とする．そして，都道府県，市区郡，町村字について，各属性値を持つ地名リストも予め用意する．都道府県については各都道府県内の市区郡名と町村字名のリスト，市区郡については，各市区郡が所属する都道府県名と各市区郡内の町村字名のリスト，町村字については，各町村字が所属する都道府県名と市区郡名のリストを用意する．

7.1.4 実装

本節では，7.1.3 節で提案した対話制御手法を用いて，日本全国の地名 177,747 種の確定をタスクとした対話システムを実装する．実装では，6.1.3 節の個人姓確定対話制御フローと同様に，認識エンジンは Nuance7.03 を使用する．実装環境も 6.1.3 節と同様であり，Nuance の N-best 値は 10 とする．

3.2.2 節のコールセンタへの適用を想定し，都道府県 47 種とアクセスの多い東京都内の市区郡と町村字，神奈川県，千葉県，埼玉県の 3 県内の市区郡の計 11,200 種を思い込み対象として選択した．提案対話制御フローにおける閾値は，個人姓確定対話制御フローと同様に，$\alpha = \beta = \gamma = \delta = 70.0$，$X = 3.0$ とした．

実装では，(Step2) の終了と同時に，思い込み対象外の地名を対象とした認識処理をバックグラウンドで開始する．リアルタイムかつ高精度な認識結果を利用者に提供するためには，日本全国を，関東の残り，北海道，東北，東海，中部，北陸，関西，四国，中国，九州の 10 個に分割し，各分割について思い込み対象外の地名リストを作り，バックグラウンドで 10 回の認識処理を実行することが効果的であると判断した．すなわち，提案対話制御フローの (Step12) では，10 回のバックグラウンドでの認識結果を尤度順に統合し，尤度が最大の地名を利用者に提示して正誤確認する．

7.2 評価実験

7.2.1 評価実験概要

本節では，住所確定対話制御フローに対する評価実験について述べる．実験では，3.2.2 節のコールセンタへの 1 日のアクセスの中から，オペレータに対する利用者の最初の発話 1,000 件を入力対象として選択した．コールセンタでは，資料を送付するために，住所を番地まで確定するが，本実験では利用者が最初に発話した地名の確定を目的とする．1,000 件はすべて町村字であり，思い込み対象内の町村字 500 件，対象外の町村字 500 件になるように選択した．被験者は，音声認識技術や対話システムに関する知識を持たない 20 代の男女 5 名ずつの計 10 名である．

実験では，比較のために，住所の階層を利用して，都道府県から順に住所を絞り込む従来システムの対話フローも実装した．従来フローでは，最初に都道府県の入

力を要求し，認識結果 1 位を提示確認する．提示の結果，誤認識の場合は，正解が提示できるまで都道府県の入力要求，提示確認を繰返す．そして，都道府県が確定した後，市区郡の入力を要求し，確定した都道府県内の市区郡のみを対象として認識し，確定できるまで市区郡の入力要求，提示確認を繰返す．その後，確定した市区郡内の町村字のみに対象を絞り込み，町村字の入力要求，提示確認を繰返す．

各被験者に対して，思い込み対象内の地名 50 件と対象外の地名 50 件の計 100 件の入力を依頼する．被験者には，確定対象の町村字名と，それらが属する都道府県名，及び市区郡名を記したリストを渡す．実験では，提案対話制御フロー，及び従来フローの実行を 1 セットとして，被験者 1 人あたり 100 セット，計 200 回の地名入力を実行する．各セットにおいて，2 種のフローはランダムに起動する．被験者には，各セットに対して同じ地名を 2 回ずつ入力するように依頼し，1 セット終了ごとに 2 種のフローに対する満足度を尋ねた．満足度は，5 段階で評価され，満足度が高いほど高得点になる．実験において，被験者には，対話の継続に負担を感じた場合は，挙手による途中放棄を認めた．両フローのシステムガイダンスには，個人姓確定対話制御フローと同じオペレータ経験のある女性の録音音声を用いた．フローにおいて,住所の提示確認に対する被験者の正誤回答は 100%の精度で認識できるものとする．

7.2.2 評価実験結果と考察

表 7.1 に，思い込み対象内の地名 500 件について，確定終了した対話数とそれらの確定までに要した平均ターン数，満足度評価の結果を示す．同様に，思い込み対象外の地名についての結果を表 7.2 に示す．本実験では，システムの地名入力要求と，それに対する利用者の入力から 1 ターンと数える．

提案対話制御フローにおいて，思い込み対象内の地名，及び対象外の地名ともに未確定終了した対話は 1 件も存在しない．一方，従来フローでは，思い込み対象内の 95 対話，対象外の 129 対話が未確定のまま終了している．従来フローでは，すべてのターンが入力要求，提示確認の繰返しのみであり，この繰返しが被験者に負担を与え，途中放棄に繋がったと予想できる．

提案対話制御フローにおいて，確定終了した対話の確定までのターン数に着目すると，思い込み対象内の地名については平均 2.79 ターンであり，入力要求に続く最初のターンである 2 ターン目で正解が提示できた対話は，500 件のうち 382 件存在

する．思い込み対象外の地名については，最初の認識対象に含まれていないため，入力に続く 2 ターン目で正解が出現することはなく，確定までに平均 3.98 ターンを要している．これは，入力要求に続く 2 ターン目で属性の入力を要求し，3 ターン目で正解が提示できた対話が多いことを意味する．一方，従来フローは，思い込み対象内，対象外の地名ともに，提案対話制御フローと比較して，確定までに約 2 倍のターンを要している．

次に，両フローに対する満足度に着目すると，提案対話制御フローに対する評価が従来フローを大きく上回っている．思い込み対象内の地名については，提案対話制御フローでは，入力要求に続く 2 ターン目で正解が提示できる対話が多いのに対して，従来フローでは，都道府県からの入力を強制し，都道府県，市区郡，町村字を一度で正しく認識できたとしても確定までに最低 3 ターンを要し，誤認識の場合は，さらに多くのターンが必要になる．従来フローでは，この都道府県からの入力の強制，及び確定までのターンの長さが利用者に負担を与える．また，思い込み対象外の地名についても，提案対話制御フローは，従来フローと比較して，高い満足度が得られている．

実験を通して，思い込み応答，及び誤認識を修正するための属性を利用した対話は，従来の対話システムと比較して，利用者満足度が高い住所確定対話システム実現の一手法になることが確認できた．

表 7.1：思い込み対象内の地名 500 件の結果

フロー	確定終了 (件)	確定 平均ターン数	途中放棄 (件)	平均満足度
提案対話制御フロー	500	2.79	0	4.32
従来フロー	405	5.92	95	1.35

表 7.2：思い込み対象外の地名 500 件の結果

フロー	確定終了 (件)	確定 平均ターン数	途中放棄 (件)	平均満足度
提案対話制御フロー	500	3.98	0	3.89
従来フロー	371	6.32	129	0.81

第 8 章
結論

本研究では，大語彙を対象とした音声対話システム実現のための対話制御手法を提案した．

現状の音声対話システムでは，対象が大語彙になるほど類似音韻候補が増大するため，人間同士の対話では，通常みられないような誤認識が起こりやすくなる．そして，利用者に対して，入力要求，提示確認という修正のためのプロセスを繰返す．再認識を繰返しても，正解が提示できず，利用者から途中放棄される対話も少なくない．本研究では，人間の対話では起こらないような機械特有の誤認識，及び修正のための同じプロセスの繰返しが，大語彙を対象とした対話システムにおける利用者の負担の要因であると考えた．

まず初めに，機械特有の誤認識を解決するために，思い込み応答を提案した．思い込み応答とは，人間同士の対話において聞き間違いが少ない対象のみをシステムに設定することで，人間の対話と同様の聞き取り傾向を実現するものである．思い込み応答によって，設定した対象については，認識精度，及び処理速度が大きく向上する．設定外の対象が発話された場合は誤認識となるが，人間も間違いやすい対象なので，その後の対話制御によって迅速に正解を導き出すことができれば，利用者の負担にはならない．

次に，思い込み設定外の発話に対する誤認識を解決するために，属性を尋ねる対話を提案した．大語彙に対する絞り込み効果を属性の有効度として定義し，評価を通して，有効度は尋ねるべき属性を決定するための尺度として有用であることを確認した．

以上の思い込み応答と属性の有効度を適用した大語彙確定のための音声対話制御手法を提案した．提案対話制御手法は，利用者の入力に対して思い込み応答を実行し，誤認識の場合は，有効度が大きい属性を尋ね，尋ねた結果に基づいて正解を絞り込むことによって，利用者に負担を与えない応答を実現する．

提案対話制御手法を 87,944 種の個人姓の確定をタスクとした対話システムに適用した．実装では，頻度上位の 10,000 件の姓を思い込み対象として選択し，個人姓

を絞り込むために，文字数，行付頭文字，先頭漢字の読み仮名という 3 種の属性を採用した．評価を通して，大語彙を一度に認識対象とする現状の対話システムと比較したところ，提案対話制御手法は，思い込み対象内の姓 1,000 件に対する初回の応答精度は約 8 割と非常に高く，対象外の姓 1,000 件についても，初回応答はすべて誤認識となるものの利用者の負担にはならないことを検証した．現状の対話システムでは，約半数の 937 件の対話が被験者から途中放棄されたのに対して，提案対話制御手法では，途中放棄された対話は 1 件も存在せず，属性を尋ねる対話は，利用者を飽きさせることなく正解を導き出せることを検証した．さらに，思い込み対象外の姓については，提案対話制御手法を用いた対話の約半数は，人間の対応と同等であるという高い評価を得た．

さらに，個人姓以外の大語彙への適用例として，提案対話制御手法を利用して，日本全国の地名 177,747 種を対象とした対話システムを実装した．コールセンタの受付け業務への適用を想定し，アクセス頻度の偏りを利用して，アクセスされやすい地名のみを選択して，思い込み応答を実行する．思い込みが外れた場合は，都道府県に対してはその都道府県内の市区郡と町村字，市区郡に対しては市区郡が所属する都道府県と市区郡内の町村字，町村字については町村字が所属する都道府県と市区郡という，残りの階層の地名を属性として尋ねる対話により住所の絞り込みを実現した．実装，評価を通して，提案対話制御手法は，利用者に都道府県からの入力を強制し，階層を利用して住所を絞り込む従来の住所確定対話手法と比較して，利用者満足度が高いシステムが実現できることを確認した．

最後に今後の課題であるが，提案対話制御フローを用いて大語彙を対象とした対話システムを実現するためには，思い込み対象，及び絞り込みに利用する属性の選択方法の検討が必要になる．個人姓のように実在頻度に偏りがある場合，あるいは住所のようにアクセス頻度に偏りがある場合は，頻度の偏りを利用して思い込み対象を選択することで，発話される可能性が大きい対象に対しては高精度な第一応答が提供できる．例えば，年間約 10 万回開催されるコンサート[21]の特定などでは，前年度の売り上げ実績に基づいて，売り上げの多いアーチストのコンサートを思い込み対象として選択し，思い込みが外れた場合は，会場，開催日を属性として尋ねる対話によって正解が導き出せる．利用者からのアクセス頻度や実在頻度に偏りがない大語彙を対象とする場合，思い込み対象の選択方法，及び利用者が迷わずに回答可能であり，大語彙に対する絞り込み効果が大きい属性の検討が必要になる．

[21] 定期刊行誌“ぴあ”(ぴあ株式会社) 12 ヶ月分の集計．

これまで本研究は，音声入力に対する応答について議論を進めてきた．思い込み応答は，音声入力に限らず，大量の検索空間から利用者が必要とする情報を高速かつ高性能に検索する手段としても有効である．現状，検索空間が広範囲に及ぶ情報検索の分野では，検索キーに対して数多くの検索結果が取得できてしまい，情報を絞り込む手段が存在しない．利用者ごとの行動履歴やアクセス履歴に基づいて，利用者の検索趣向を導き出す手段として思い込み応答を適用することで，同じ検索キーが入力された場合でも，個々人適応型の情報提供が可能になる．

本論文を通して，提案対話制御手法は，対象が大語彙であっても，人間の対応に近い対話システム実現のための一方法となることが確認できた．本提案が，オペレータ業務の代替として適用可能な実用システムの実現に役立つことを期待する．

謝辞

本研究をまとめるにあたり，種々の御指導，御教示を賜りました慶応義塾大学情報工学科の斎藤博昭専任講師に深く感謝致します．

お忙しい中，副査をお引き受け下さり，御指導，御助言を頂きました原田賢一教授，小澤慎治教授，今井倫太専任講師にも深く御礼を申し上げます．

本研究は，NTT 情報通信研究所，第二プロジェクトリーダ東田正信氏（現在，NTT アドバンステクノロジー株式会社）との出会いによって始まりました．そして，NTT 情報流通プラットフォーム研究所において，松本隆明プロジェクトマネージャー(現在，株式会社 NTT データ技術開発本部長）より，研究の継続，発展のために数々の御指導，御支援を頂き，再び大学院での就学の機会を頂きました．松本隆明氏の御支援がなければ，研究を継続することはできなかったと感じます．心からの感謝と御礼を申し上げます．

株式会社 NTT データ技術開発本部，ユーザインタフェースグループの菅原昌平部長，遠藤淳課長，磯部俊洋課長には，社会人と学生生活の両立に当たり，数多くの御支援，御理解を頂きました．

NTT コミュニケーション科学基礎研究所の堂坂浩二氏には，研究当初から多岐にわたる御教示を賜りました．NTT サイバースペース研究所の永田昌明氏，大附克利氏には，音声認識処理，音声対話処理の研究動向について御教示を賜りました．

NTT 情報流通プラットフォーム研究所時代，研究を進めるに当たり，伊土誠一氏(現在，株式会社 NTT ソフトウェア副社長)，菱沼千明氏（現在，東京工科大学教授)には，研究に対して数々の御支援と御理解を頂きました．

本研究を進めるにあたり，御支援を頂きましたすべての皆様に深く感謝の意を表します．そして，本研究内容を，故･中西正和教授に御報告致します．

最後に，社会人と学生生活の両立を支えてくれた両親に，そして学業を優先することを許してくれた主人の優しさに深く感謝します．

2004 年 12 月

参考文献

[1] 赤堀一郎, 加藤利文, 北岡教英. “地名認識システムとその応用” 情報処理学会研究会報告, SLP-7-9 (1995).

[2] 赤塚忠, 阿部吉雄 (編集). “漢和辞典改訂新版” 旺文社 (1986).

[3] Venkataraman, A., Franco, H., Myers, G. “An Architecture for Rapid Retrieval of Structured Information Using Speech with Application to Spoken Address Recognition,” In Proceedings of the ASRU-2003, pp. 459-464 (2003).

[4] Aust, H., Oerder, M., Seide, F., Stenbiss, V. “The Philips automatic train timetable information system,” Speech Communication, Vol. 17, pp. 249-262 (1995).

[5] B. シュナイダーマン著, 東基衛, 井関治訳. “ユーザインタフェースの設計” 日経 BP 社 (1995).

[6] Biermann, A. W., Rodman, R. D., Rubin, D. C., Heidlage, J. F. “Natural Language With Discrete Speech as a Mode for Human-to-Machine Communication,” Communications of the ACM, Vol. 28, No. 6, pp. 628-636 (1985).

[7] Billi, R., Castagneri, G., Danieli, M. “Field trial evaluations of two different information inquiry systems,” Speech Communication, Vol. 23, pp. 83-93 (1997).

[8] Bossemeyer, R. W., Schwab, E. C. “Automated alternate billing services at Ameritech: Speech recognition performance and the human interface,” Speech Technology Mag. Vol. 5, No. 3, pp. 24-30 (1991).

[9] C. シュマント著, 石川泰訳. “コンピュータとのヴォイスコミュニケーション” サイエンス社 (1995).

[10] Chu-Carroll, J. “MIMIC: An adaptive mixed initiative spoken dialogue system for information queries,” In Proceedings of the ANLP-2000, pp. 97-104 (2000).

[11] Córdoba, R., San-Segundo, R., Montero, J. M., Colás, J., Ferreiros, J., Macias-Guarasa, J., Pardo, J. M. “An Interactive Directory Assistance

Service for Spanish with Large-Vocabulary Recognition," In Proceedings of the Eurospeech '01, pp. 1279-1282 (2001).
[12] Cozannet, A., Siroux, J. "Strategies for oral dialogue control," In Proceedings of the ICSLP-94, pp. 963-966 (1994).
[13] 堂坂浩二, 安田宜仁, 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).
[14] Ferguson, G., Allen, J. F. "TRIPS: An integrated intelligent problem-solving assistant," In Proceedings of the AAAI-98, pp.567-572 (1998).
[15] 古井貞熙. "音響・音声工学" 近代科学社 (1992).
[16] 古井貞熙. "音声情報処理" 森北出版 (1998).
[17] Goddeau, D., Brill, E., Glass, J., Pao, C., Phillips, M., Polifroni, J., Seneff, S., Zue, V. "GALAXY: A Human-language Interface to On-line Travel Information," In Proceedings of the ICSLP-94, pp. 707-710 (1994).
[18] Goddeau, D., Meng, H., Polifroni, J., Seneff, S., Busayapongchai, S. "A Form-Based Dialogue Manager For Spoken Language Applications," In Proceedings of the ICSLP-96, pp. 701-704 (1996).
[19] Gerbino, E., Baggia, P., Ciaramella, A., Rullent, C., "Test and evaluation of a spoken dialogue system," In Proceedings of the ICASSP, Vol. 2, pp. 135-138 (1993).
[20] 星野裕, 加藤博子, 永田健児. "30 万人読み方書き方辞典" 日外アソシエーツ (1993).
[21] http://support.jp.dell.com/jp/jp/spm/phone/
[22] http://vcl.vaio.sony.co.jp/info/technical.html
[23] http://www.jmscom.co.jp/bsn/bnsAuto.html
[24] http://www.ntt.com/v-portal/
[25] http://www.ntt.com/contact/index.html
[26] http://www.nomura.co.jp/service/kabukadial.html
[27] http://www.nuance.com/
[28] http://www.ufj-tsubasa.co.jp/contact/index.html
[29] http://www.voizi.net/
[30] Hymes, D. "Models of the interaction of language and social life." Directions in Sociolinguistics, New York, Holt, Rinehart & Winston (1972).
[31] 磯部俊洋, 森島昌俊, 吉谷文徳, 小泉宣夫. "電話音声認識による資金移動サー

ビスと対話の評価” 情報処理学会論文誌, Vol. 39, No. 5, pp. 1258-1265 (1998).
[32] 伊藤敏彦, 小暮悟, 中川聖一. “協調的応答を備えた観光案内対話システムとその評価” 情報処理学会論文誌, Vol. 39, No. 5, pp. 1248-1257 (1998).
[33] 伊藤敏彦, 甲斐充彦, 岩本義行, 水谷誠, 由浅裕規, 小西達裕, 伊東幸宏. “目的地設定タスクにおける対話状況の違いによる言語･音響的特徴の比較” 情報処理学会論文誌, Vol. 43, No. 7, pp. 2118-2129 (2002).
[34] 伊藤亮介, 駒谷和範, 河原達也. “機器操作マニュアルの知識と構造を利用した音声対話ヘルプシステム” 情報処理学会論文誌, Vol. 43, No. 7, pp. 2147-2154 (2002).
[35] 自治省. “国土行政区画総覧” 国土地理協会 (1998).
[36] 亀田弘之, 藤崎博也. “情報検索における音声･言語処理” 情報処理学会研究会報告, SLP-16-11 (1997).
[37] 菊池英明, 白井克彦. “対話効率の向上を目的とした音声対話制御のモデル化” ヒューマンインタフェース学会誌, Vol. 2, No. 2, pp.145-152 (2000).
[38] 北研二. “確率言語モデル” 東京大学出版会 (1999).
[39] 北岡教英, 赤堀一郎, 中川聖一. “認識結果の正解確率に基づく信頼度とリジェクション” 電子情報通信学会論文誌 D-II, Vol. J83-D-II, No. 11, pp. 2160-2170 (2000).
[40] 駒谷和範, 河原達也. “音声認識結果の信頼度を用いた頑健な混合主導対話の実現法” 情報処理学会研究会報告, SLP-30-9 (2000).
[41] 小坂昌宏, 松橋聡, 中山卓郎. “音声認識における対話制御方式の一考察” 電子情報通信学会総合大会, D-14-6, pp. 217 (1998).
[42] Kuroiwa, S., Takeda, K., Naito, M., Inoue, N., Yamamoto, S. “Error analysis of field trial results of a spoken dialogue system for telecommunications applications,” IEICE Trans. E78-D, No. 6, pp. 636-641 (1995).
[43] Litman, D. J., Walker, M. A., Kearns, M. S. “Automatic detection of poor speech recognition at the dialogue level,” In Proceedings of the ACL99, pp. 309-316 (1999).
[44] Litman, D. J., Hirschberg, J. B., Swerts, M. “Predicting automatic speech recognition performance using prosodic cues,” In proceedings of the ANLP-2000, pp. 218-225 (2000).
[45] 村上仁一, 嵯峨山茂樹. “自由発話音声認識における音響的および言語的な問題点の検討” 日本音響学会研究会資料, SP-91-100, pp. 71-78 (1991).

[46] 中川聖一. “音声対話システムの評価法” 情報処理学会研究会報告, SLP-7-19 (1995).
[47] 中川聖一, 堂下修司. “音声言語情報処理研究の動向と研究課題” 情報処理, Vol. 36, No. 11, pp. 1012-1019 (1995).
[48] 中川聖一. “音声対話システム構築の課題” 日本音響学会誌, Vol. 54, No. 11, pp. 783-790 (1998).
[49] 中野幹生, 堂坂浩二. “音声対話システムの言語・対話処理” 人工知能学会誌, Vol. 17, No. 3, pp. 271-278 (2002).
[50] Nakano, M., Dohsaka, K., Miyazaki, N., Hirasawa, J., Tamoto, M., Kawamori, A., Sugiyama, A., Kawabata, T. “Handling rich turn-taking in spoken dialogue systems,” In Proceedings of the Eurospeech ’99, pp. 1167-1170 (1999).
[51] Nielsen, P. B., Baekgaara, A. “Experience with a dialogue description formalism for realistic applications,” In Proceedings of the ICSLP-92, pp. 719-722 (1992).
[52] 新美康永, 小林豊. “音声認識の誤りを考慮した対話制御方式のモデル化” 情報処理学会研究会報告, SLP-5-7 (1995).
[53] 新美康永, 小林豊. “音声認識の信頼性に基づいた対話制御方式” 電子情報通信学会技術研究報告, SP-96-30 (1996).
[54] 新美康永. “音声対話システムの対話制御” 日本音響学会誌, Vol. 54, No. 11, pp. 791-796 (1998).
[55] 新美康永, 西本卓也, 荒木雅弘. “確認対話の制御方式の効率と音声認識システムの性能との関係” 情報処理学会研究会報告, SLP-27-17 (1999).
[56] Niimi, Y., Takigawa, N., Nishimoto, T. “Modeling dialogue strategies to resolve speech recognition errors,” In Proceedings of the Eurospeech ’95, pp. 534-537 (1995).
[57] Niimi, Y., Nishimoto, T., Kobayashi, Y. “Analysis of interactive strategy to recover from misrecognition of utterances including multiple information items,” In Proceedings of the Eurospeech ’97, pp. 2251-2254 (1997).
[58] 竹林洋一. “音声自由対話システム TOSBURGII-ユーザ中心のマルチモーダルインタフェースの実現に向けて-” 電子情報通信学会論文誌 D-II, Vol. J77-D-II, No. 8, pp. 1417-1428 (1994).
[59] Takebayashi, Y., Tsuboi, H., Kanazawa, H., Sasamoto, Y., Hashimoto, H.,

Shinchi, H. “A real-time speech dialogue system using spontaneous speech understanding,” IEICE Trans. E76-D, No. 1, pp. 112-120 (1993).

[60] 竹沢寿幸. “対話音声文法の構築” 日本音響学会誌, Vol. 54, No. 11, pp. 803-806 (1998).

[61] 田中修一, 中里収, 帆足啓一郎, 白井克彦. “複合作業下における音声インタフェースの設計と評価” 電子情報通信学会論文誌 D-II, Vol. J79-D-II, No. 12, pp. 2163-2169 (1996).

[62] 土岡正純. “Computer TELEPHONY,” リックテレコム (2004. 2 月号).

[63] 上野晋一, 駒谷和範, 河原達也, 奥乃博. “バス運行情報案内システムにおけるユーザモデルを用いた適応的応答の生成” 情報処理学会研究会報告, SLP-42-2 (2002).

[64] Watanabe, T., Araki, M., Doshita, S. “Evaluating Dialogue Strategies under Communication Errors using Computer-to-Computer Simulation,” IEICE Trans. E81-D, No. 9, pp. 1025-1033 (1998).

[65] Yamamoto, M., Kobayashi, S., Moriya, Y., Nakagawa, S. “A Spoken dialog system with verification and clarification queries,” IEICE Trans. E76-D, No. 1, pp. 1081-1089 (1993).

[66] 山本幹雄, 伊藤敏彦, 肥田野勝, 中川聖一. “人間の理解手法を用いたロバストな音声対話システム” 情報処理学会論文誌, Vol. 37, No. 4, pp. 471-482 (1996).

[67] 吉岡理, 荒井和博, 菅村昇, 嵯峨山茂樹. “音声認識機能を含むマルチモーダルインタフェースを持つ住所入力システムの開発と評価” 電子情報通信学会論文誌, J80-D-II, No. 5, pp. 1007-1015 (1997).

[68] Zeigler, B., Mazor, B. “Dialogue design for a speech-interactive automation system,” In Proceedings of the Eurospeech ’95, pp. 113-116 (1995).

[69] Zue, V., Seneff, S., Polifroni, J., Phillips, M., Pao, C., Goodine, D., Goddeau, D., Glass, J. “PEGASUS: A spoken dialogue interface for on-line air travel planning,” Speech Communications, Vol. 15, pp. 331-340 (1994).

[70] Zue, V. “Conversational interfaces: advances and challenges,” In Proceedings of the Eurospeech ’97, pp. 9-18 (1997).