社団法人　情報処理学会　研究報告
IPSJ SIG Technical Report

2003－ＳＬＰ－49　(23)
2003／12／18

様々な雑音環境での音声対話における文法と認識精度の関係の分析

大庭　隆伸†　　鈴木　基之†　　伊藤　彰則†　　牧野　正三†

† 東北大学大学院工学研究科　〒980-8579 宮城県仙台市青葉区荒巻字青葉 05
E-mail: †{bacchi,moto,aito,makino}@makino.ecei.tohoku.ac.jp

あらまし　音声認識において，雑音下での認識精度の改善は重要な課題の一つとなっている．そのために，音響モデルや雑音除去法の改善など様々な研究が行われているが，本稿では，対話の立場からの精度改善を試みる．具体的には，音声認識にとって不利な雑音環境になるのにあわせ，認識対象とする語彙・侯補数を削減した文法に変更し音声認識を行う．これにより，雑音の影響が小さい場合には，ユーザの自由な発話を認識できる枠組みを残しつつ，雑音下でも一定の認識精度を維持して対話を行うことが可能となる．これを実現するためには，まず，語彙・侯補数を削減した際に，認識側で認識対象としていない語彙や文法を含むユーザ発話が増加してしまうが，そのための対策が必要となる．また，認識文法を環境にあわせて変更させるには，ある雑音下で対話を行った場合に，認識精度がどの程度になるかを推定する必要があり，これをどのように実現するかが課題となる．前者については，システムの質問提示方法を工夫することにより対策を行い，後者については，雑音・文法と認識精度の関係をニューラルネット学習により推定可能か検討する．

キーワード　対話，認識精度，雑音，文法，質問提示方法

Takanobu OBA†, Motoyuki SUZUKI†, Akinori ITO†, and Syozo MAKINO†

† Graduate school of Engineering, Tohoku University
05 Aramaki aza Aoba, Aoba-ku, Sendai, Miyagi, 980-8579 Japan
E-mail: †{bacchi,moto,aito,makino}@makino.ecei.tohoku.ac.jp

Abstract　Speech recognition under noisy environment is one of the hottest topic in the speech recognition research. Noise-tolerant acoustic models or noise reduction techniques are often used to improve the recognition accuracy. In this paper, we propose a method to improve accuracy of spoken dialog system from a dialog strategy point of view. In the proposed method, the dialog system automatically changes its dialog strategy according to the estimated recognition accuracy in noisy environment in order to keep the performance of the system constant. In a noise-free environment, the system accepts any utterance from a user. On the other hand, the system restricts its grammar and vocabulary in a noisy environment. To realize this strategy, we investigated a method to avoid user's out-of-grammar utterances through an instruction given by the system to a user. Furthermore, we developed a method to estimate recognition accuracy from features extracted from noise signal.

Key words　dialogue, word accuracy, noise, grammar, how to ask a question

1. はじめに

音声対話によりユーザ要求を受け取る技術は，ロボットへの指示など，様々な分野での需要がある．音声認識の機能は，最近では，エンターテイメント型のロボットやゲームなどにもみられるほか，目の不自由な方のための音声サービスとして，ATM などにも用いられるようになってきている．

そんな中，我々は，介助ロボット IRIS[1] の製作に参加し，対話システムの搭載を担当した[2]．その対話システムの特長の一つに，話題に応じ認識文法を変え，対話が進む中で認識精度を向上させるというものがある．これにより，８割のタスク達成率を誇る対話システムを作ることができた．

しかし，認識精度は雑音下では低く課題を残した．雑音下での音声認識に関しては，多くの研究機関において雑音処理や，雑音に頑健な音響モデルの研究が行われており，音声認識技術において重要な研究領域となっている．現在では，音声認識の使用環境に特定の雑音しか存在しない場合には，その環境に特化した対策を施すことで，非常に精度よく認識できるようになってきている．しかし，実際の使用環境下では様々な種類の雑音が存在しており，あらゆる雑音下で高い認識精度を達成する

ことが望まれている．多くの雑音への対策方法として，張ら[3]は，区分線形変換による方法を提案しており，雑音重畳音声のHMMパラメータ空間を，雑音の性質とSN比によって木構造に区分化し，入力音声の条件に最も適合した部分空間で，尤度が最大となるような線形変換を行い，雑音へのモデルの適応化を行っている．このような認識性能そのものの改善は，極めて重要である．

しかしながら，雑音下での認識精度が向上したとしても，それでも認識が難しい環境というものは存在する．事実，人間の認識性能を持ってしても，そのような環境がいたる所に存在している．そこで，我々は，雑音下での認識精度改善のため，雑音に応じて文法を変更する．ただし，むやみに語彙を減らすことは，ユーザ要求の受け取りにかかる時間をあまりに長くしてしまう上，音声を利用するメリットを奪ってしまう．そこで，普段は快適に，雑音下でも精度よく利用できる対話システムを構築するため，雑音に応じ，語彙・候補数を削減した文法を選択することを試みる．その際，選択した文法の語彙・候補数が少ないほど，認識できない発話が多くなるといった問題点が懸念される．また，雑音と文法の組合せで，認識精度がどの程度になるか評価する必要がある．本稿では，これらについての検討を行なう．

なお，対話システムが受け付ける要求としては，介助ロボットへの指示や，家電の操作を想定している．一つの用件に対して，ユーザが伝えるべき項目数は，3〜6程度のものとする．

2. 複数文法・対話戦略の選択に基づく雑音環境への適応

人間の音声認識能力は，現行の音声認識システムによるものよりも遥かに優っているものといえる．それでも，SNが非常に悪い環境下では，やはり相当認識能力が低下してしまう．そのような環境でも，人間は一言一言を大きな声ではっきり，かつ，間をやや長めにとって発話するなどの対策をとることで要求を伝え，聞き手もそれにより正しく認識することが可能になる．もちろん，大きな声を出すことはSNを回復させるためであるが，一言一言をはっきり，間をとって発声することは聞き手に選択肢を限定させる効果があると考えられる．我々は，この選択肢限定による効果を，対話システムへ応用させる方法を検討することにした．つまり，認識対象とする語彙・候補数を削減した文法を用いて音声認識を行うのである．しかし，削減され，認識候補から外れてしまった文法は認識が不可能であるので，単純に削減することは好ましくない．また，当然であるが，語彙・候補数を減少させれば，質問形式で対話を行うなどの必要性がでてくるので，ユーザの発話回数を増加させてしまうという問題も生じる．そこで，雑音のない環境下では，語彙数の多い文法を用い，音声認識にとって劣悪な雑音環境になるのにあわせて，認識対象とする語彙数の少ない文法とそれに応じた対話戦略に変更するという方針をとる．これにより，通常はユーザの自由な発話を受理でき，雑音下でも，発話内容の制限と引き換えとなるものの，認識精度を維持したまま対話を行うことができる．ただし，ここでの対話戦略とは，質問形式にするしない，どの程度の項目数を一度に尋ねるかなどを示す言葉として用いている．

対話システムの動作としては，まず対話直前に周辺雑音を収集する．次に，その雑音の分析を行い，所望の認識精度を維持できる文法・対話戦略を選択し，それを用いて対話を開始する．ただし，SN比を知るために，ユーザの一発話目を入手してから，対話戦略の選択を行う必要があるかもしれない．場合によっては，平均音声パワーのようなものを用いても十分であるかも知れない．これについては，本稿では論ぜず，今後の課題とする．

上記のようなシステムの実現を試みるとき，次の二つの課題が挙げられる．一つは，語彙・候補数の少ない文法を選択した際に，システム側で用意している文法以外のユーザ発話は，正しく認識することができないという点である．以降，このようなユーザ発話を辞書外発話と呼び，文法に登録されてあり，正しく認識できる可能性のあるユーザ発話を辞書内発話と呼ぶことにする．課題の二つ目は，それぞれの環境下で，一定の認識精度が維持可能な文法というものを選定する手法についてである．ある環境下で，ある文法を使用して実際に対話を行った際に，認識精度がどの程度になるのかを予測する必要があり，その具体的方法を確立しなければならない．これらの課題について，次節から順に記述する．

3. 辞書外発話への対策

3.1 辞書外発話の認識精度への影響

まず始めに，辞書外発話により，対話全体の認識精度がどの程度低下するか確認しておく．

いくつかの環境下で，語彙数が400語と30語の文法を用意し，実際に対話実験を行った．設定した環境は，雑音無しと，空調機，幹線道路，音楽の各雑音それぞれSN比で20，10dBの場合である．雑音は対話用マイクの周辺$1m$程度にスピーカを置き，そこから再生させた．SN比は，ユーザの平均発声パワーにあわせているので，発話ごとに変化しており，あくまで目安値である．対話戦略に関しては，語彙数が400語の場合は，ユーザに自由に発話をさせ，30語の場合は，同程度の語彙数の文法をいくつか用意し，一問一答式で，文法を切り替えながら対話を進めた．

その結果，辞書内発話のみ集計した認識精度は図1のようになり，辞書外発話を含めて集計した場合は図2のようになった．図1は，確かに語彙・候補数を削減することが雑音下での認識精度向上に貢献していることを表しているといえるが，辞書外発話を含めた場合の図2では，この効果がちょうど失われているのが見てとれる．

なお，一問一答式で対話を進めた際の辞書外発話率は20.2%であった．

3.2 システムの質問提示法とユーザ発話の変化

認識対象とする語彙・候補数の削減による認識精度向上の効果が，辞書外発話により失われたという結果は，逆に考えれば，辞書外発話さえ防ぐことができれば，雑音下でも認識精度の維持が可能であることを示しているということである．

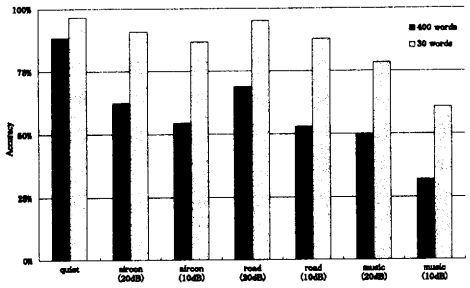

図 1　辞書内発話のみの認識精度

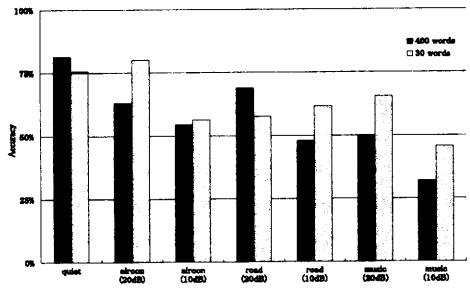

図 2　辞書外発話を含めた認識精度

そこで，システムの質問提示方法を工夫することにより，ユーザの発声に変化を生じさせ，答えの範囲に制限を与え，辞書外発話を防止可能か調べることにした．

3.2.1　システムの質問提示法

まず，ユーザ発話に制約を与えられる可能性のあるシステムの質問提示法として，表 1 のようなものを用意した．これらが，実際にどの程度の制約を与えられるのか，以下の対話実験により検証を行なった．

表 1　ユーザ発話制御のためのシステムの質問提示法案

応答法	応答方法の説明
単項目	1 項目ごとの質問 (例 「操作する家電は何ですか?」)
複数項目	「何をいくつですか?」のように，複数項目同時に質問 語順・解答の範囲を暗黙のうちに指定
質問内指定	対話冒頭で「質問以外に答えないで」と指示 その後は，単項目と同じ
単語指定	対話冒頭で「単語のみで答えて」と指示 ユーザが単語だけで答えてくれるのであれば， 用意すべき語彙・候補数は極めて少なくできる
音声選択	選択肢を設け，選択肢まで (画面表示などだけにせず) 音声で出力 (例 「飲み物はどれですか?1 番，お茶．2 番，紅茶…」)

この後の説明が理解しやすいように，ここで，想定辞書という概念を定義し，表 2 にまとめる．これは，各質問提示法に対して，システムが想定しているであろうユーザ発話の範囲を示したものである．基本的には，尋ねられた質問に対する解答だけを想定辞書の範囲とする．ただし，「はい」，「いいえ」，「違います」，「戻って」なども想定辞書の範囲とする．ある質問提示法に対するユーザ発話が，その想定辞書の範囲におさまっていれば，想定辞書の範囲の文法だけ用意しておくことで，辞書外発話を抑えることができる．

表 2　各質問提示法に対する想定辞書

応答法	想定辞書
単項目	尋ねた質問に対する解答のみ
複数項目	尋ねた複数の項目の語順通りの解答と その複数の項目の一部だけの解答
質問内指定	尋ねた質問に対する解答のみ
単語指定	尋ねた質問に対する解答となる単語のみ
音声選択	尋ねた質問に対する解答 番号だけや，番号と解答語句をあわせたものも含む

3.2.2　実　　験

システムの質問提示法を検証するための実験を行った．被検者は男性 22 名，女性 10 名の計 32 名である．各被検者に対し，各質問提示法を 2 回ずつ試した．このときに，被検者がシステムに伝えるべき内容は，随時，実験者側で指定した．また，試した質問提示法の順番はランダムであり，質問提示法の提示順番がもたらす影響を抑えるようにしている．

これらの実験は Wizard Of Oz 法で行った．ここで，人間相手と機械相手でユーザ応答が変化するとの報告がなされている [4] ことに留意し，被験者には本当に音声認識を行っていると伝え，かつ，応答文をすぐに出せるよう準備し，システム応答を画面表示と合成音声で再生するなどすることで，被験者に本当に音声認識を行っているかのように思い込ませた．

また，「戻って」と言うと，システムが 1 つ前の応答に戻ってくれるという機能もつけておいた．さらに，被検者により認識精度の設定も変えて実験を行った．精度のよい場合は 8〜9 割以上，悪い場合は 5〜7 割程度とした．

3.2.3　結　　果

この実験の結果を図 3 に示す．これは，想定辞書の範囲に納まっているユーザ発話の割合を表したものである．この割合が高いということは，想定辞書の範囲程度からなる認識文法を用意しておけば，辞書外発話を十分抑えられるということである．音声選択では，100%になっているが，被験者の答え方は，番号をいうか，システム応答とほぼ同じ表現をするか，その両方をあわせるか程度となっていて，ユーザ発話に強い制約を与えるものとなった．質問内指定では，質問の該当以外の情報を含んだ発話は 1 例だけであった．一方，単語指定でも，かなり該当情報だけを答えているが，1 回の対話が長引くと，単語のみで答える率が低下してくる．複数項目では，複数項目質問しても，その一部だけ答えるケースがやや多かった．複数個の答えを被験者が答えるときは，必ずシステム応答と同じ語順で答えていた．例えば，「何をいくつどうする？」という問に対して，「紅茶を 2 杯ください」はあっても，「2 杯紅茶をください」といったように，わざわざ語順を入れ換えるといった事例は全く無かった．

このほか，想定認識精度を低くし，誤認識を多く発生させら

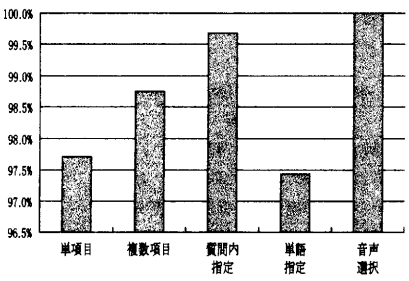

図 3　想定辞書に対する辞書内発話率

れた被験者の方が，質問によく答えるという結果も得られた．誤認識に対しては，システムの確認応答に対し，「いいえ」，「違います」，「戻って」などのような，誤認識された語句に対する修正語句を含まない発話が多く，伊藤ら [4] の報告とも一致した．

今回の実験では，単項目でも想定辞書内の発話率が高いという結果が得られた．この原因は，今回，被験者に対し，終始質問型式の対話戦略のみを提示し続けたことにより，被検者が質問に答えることに慣れてしまったことにあると思われる．

4. 雑音下における認識精度の推定

ここでは，環境に適した文法・対話戦略の選択を行なう際のもう一つの課題となる，ある環境下で，ある文法を使用した際の認識精度の推定方法について考えていく．

4.1 文法選択における認識精度推定の役割

まず，文法・対話戦略の選択を行う際に，認識精度の推定をどのように利用するのか詳しく説明する．

今，対話システムは三種類の文法 A，B，C を持っているものとする．ただし，一度の対話の中で，文法を切り替えながら対話を進めるような場合は，文法集合のうちの平均的な語彙数のもの，もしくは最大の語彙数のものとする．対話戦略は各文法固有のものが用意されている．文法の選択にあたり，まず対話システムは雑音を収録・分析し雑音パラメータを得る．この雑音パラメータと文法のパラメータを用い，認識精度の推定を行う．この推定は，文法 A，B，C 全てに対して行う．その推定結果を a%，b%，c% とする．対話システムは，最低限維持したい認識精度というものを持っており，その値を α% とする．$\alpha > a$，$\alpha < b$，$\alpha < c$ とすると，最終的に選択される可能性のある文法は B か C ということになる．B と C のどちらを使用するかは，対話システムに依存することにしておく．語彙数や対話戦略を基準に選択することが考えられる．したがって，文法選択において，認識精度の推定は α% 以上になるかどうかが重要となる．

4.2 認識精度推定の概略

この推定問題における目的変数は，実際に対話を行った際の認識精度である．一方，説明変数は雑音から計測するパラメータ $X_n = (x_1, x_2, \cdots, x_n)$ と，文法のサイズなどを表すパラメータ $X_g = (x_{n+1}, x_{n+2}, \cdots, x_N)$ からなる．後者としては，語彙数，候補数 (文法から生成される文の数)，のべ語彙数 (文法から生成される全ての文を単語分解したときの，単語ののべ数) をパラメータ候補として用意した．なお，認識に使用した文法は，オートマトンで記述されたもので，全ての認識候補文が等確率で生起するものを使用した．つまり，候補数がそのままパープレキシティとなる．

この推定問題は，説明変数から目的変数への写像 Ψ を求めることに相当する．

$$Accuracy = \Psi(x_1, x_2, \cdots, x_N)$$

この写像をあたえる手法として，いくつか考えられるが，非線形な写像となることが予想されるため，ニューラルネットを用いた方法を試すことにした．そのためにも，まずはこの写像 Ψ を求めるための十分な学習データが必要となる．最初に，雑音のパラメータと学習データの収集方法について記す．

4.3 学習データの収集

4.3.1 雑音のパラメータ候補と計測方法

推定のために必要な雑音パラメータは，認識精度と相関のあるものでなくてはならない．そこで，以下のようなものを採用候補として用意した．

- SN 比
- フレーム間のパワー分散 (標準偏差/平均パワー)
- フレーム間の相関係数の平均・分散
- フレームごとのスペクトル勾配の平均・分散
- 雑音を連続音素認識にかけたときに検出される音素数

相関係数の利用は，音声信号の特徴である調波構造をつかまえるために相関値を利用し，音声区間検出をおこなう研究 [5] に基づいたものである．もし，雑音の相関値が高ければ，認識器が音素と誤認識する可能性がある．雑音を音素認識させるのも，同様の理由からである．

さて，実際に対話システムが認識精度の推定を行なうためには，周辺雑音を入手する必要があるが，実際に利用できる雑音は短いことが予想される．せいぜい，対話開始前の数秒を利用するにとどまるだろう．そこで，今回，学習データを用意するにあたっても，同様に短い区間から雑音パラメータを検出することにする．まず，各雑音データに対し，ランダムに 1 秒区間をいくつか切り取る．以後，この区間を雑音計測区間と呼ぶ．各雑音計測区間について，窓幅 25msec，25msec シフトでパラメータを計測し，その平均・分散を求める．25msec は認識器の窓幅に合わせたたものである．切り出した 1sec 区間は，全雑音トータルで 135 区間である．連続音素認識に関しては，1sec では短すぎ，雑音間に差が出なかったので，2sec を切り抜き使用することにした．

尚，雑音データとして，電子協雑音データベースから，展示会場，駅，板金工場，脱荷シュート，交差点，人ごみ，在来線，空調機，エレベータホール，と，RWC 研究用音楽データベースから，classic，jazz，pop，童謡を用いた．

4.3.2 学習データの生成方法

次に，雑音重畳音声の生成方法について説明する．

音声データに重畳させる雑音データの区間は，雑音計測区間

の後に続く部分とした．これは，実際に対話システムを作成した際に，対話直前に雑音パラメータを計測し，その後，対話が始まるという状況を想定してのものである．また，実際の対話では，システムとユーザの発話が何ターンか繰りかえされるので，音声データに重畳する雑音区間は，雑音計測区間から 1sec おきに 5 区間を用いることにした．

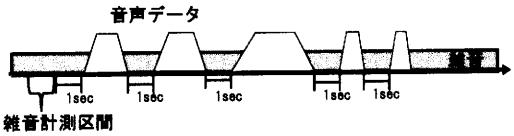

図 4　学習データの生成

認識のため，用意した文法は 52 種類で，語彙数は最小で 30 語，最大で 340 語．認識候補数は最小で 155，最大で 8×10^8 通り．のべ語彙数は，およそ 860〜1.6×10^6 である．

認識は，各文法で，各雑音計測区間に続く 5 つの雑音重畳音声に対し，それぞれ SN 比が，40，30，20，15，10，5，0dB の場合について行なった．結局，認識させた音声データ数は，52(文法) × 135(雑音) × 5(音声) × 7(SN 比) である．また，これらの雑音重畳音声は，認識に使用する文法に対し辞書内発話であるものをランダムに選択した．認識精度の計算は，各雑音計測区間に続く 5 つの雑音重畳音声ごとに行った．

4.4　認識精度の推定

説明変数から目的変数への写像 Ψ を得るために，ここでは，階層型ニューラルネットにより学習する．

4.4.1　実　　験

学習データとして，52(文法)×135(雑音)×5(音声)×7(SN 比) 個の雑音重畳音声を認識させ，52(文法)×135(雑音)×7(SN 比) 個の説明変数-目的変数 (認識精度) セットが得られた．うち，4000 セットはオープン実験用にとっておき，2000 ずつに分け，それぞれ，open1，open2 とした．残りのセットは学習用データとした．また，学習用データのうち，4000 セットをランダムに選択し，オープン実験用データ同様，半分に分け，それぞれ close1，close2 とした．そして，ニューラルネットにより学習をさせ，open1,2，close1,2 により評価を行った．

4.4.2　評 価 方 法

評価方法は，α%以上の認識精度を期待した場合に，実際にそのような状況であることを正しく推定できているかを表す再現率と，α%以上の推定結果に対し，実際に α%以上の認識精度であった割合を表す適合率で示す．評価データ中で，実際の認識精度 (目的変数) が α%以上であるものの総数を t_f，推定結果により α%以上であったものの総数を t_d，そのうち，実際の認識精度も α%以上であったものの総数を t_c とすると，再現率 Re，適合率 Pre はそれぞれ

$$Re = t_c/t_f$$

$$Pre = t_c/t_d$$

である．

一方，文法選択の基準として認識精度の推定を行う場合，推定結果が実際の認識精度よりも大きく上回ってしまうと，認識精度が低くなってしまう雑音下で，語彙数の多い文法を使用してしまうことになり，本来の目的を失ってしまう．そこで，α%以上の推定結果が得られたにも関らず，実際は α より 10%以上低かった場合の割合についても評価することにする．α%以上の推定結果が出たときに，実際は (α-10)%以下であったものの総数を t_l とすると，t_l/t_d である．

4.4.3　結　　果

$\alpha = 50, 60, 70, 80, 90$ についての再現率，適合率をそれぞれ図 5，6 に示す．

オープン，クローズ間で，ほとんど差がなかった．学習データ以外でも，学習データと同様の推定精度が期待できると言える．再現率，適合率は，ともに α の値が大きくなるにつれ低下する傾向がみられた．特に，再現率で急激に低下している．これは，認識精度の推定結果が低い場合でも，実際の認識精度は高かったというケースが多いことを示している．α の値が小さい時に，適合率が高い値を示しているのにはこの影響も含まれている．文法選択にあたっては，推定結果を信用して用いることができるので，適合率が高いことが望まれるが，全体的に高い値を示している．

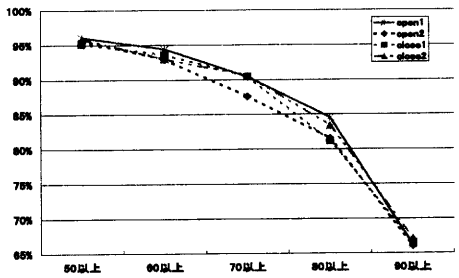

図 5　再　現　率

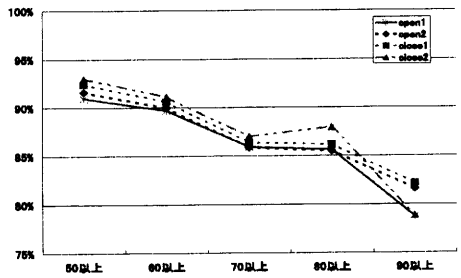

図 6　適　合　率

推定結果が，実際の認識精度を 10%以上上回った割合を図 7 に示す．結果としては，およそ 5〜10%の間にあり，認識精度推定を行った時に，十数回に一度の確率で実際の認識精度が 10%以上，下回ることになる．文法選択の際は，複数の文法に対して認識精度推定を行うことになるので，最終的に選択され

た文法における推定結果が，これに該当するものである確率は極めて低いと考えられる．

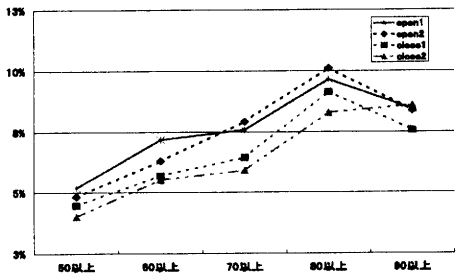

図 7　推定結果が実際より 10%以上高くなった割合

4.5　雑音パラメータの選定

今回，使用した雑音のパラメータは，SN 比，パワーの分散，相関係数の平均と分散，スペクトル勾配の平均と分散，音素認識で検出された音素数の 7 種類である．対話の直前にこれらを計測して認識精度を推定するという枠組みで取扱ってきたが，これらのパラメータ全てとなると，計算量が危惧される．もっと少ないパラメータ数で同等の推定精度を出せないか，推定にほとんど関与しないパラメータはどれなのか検証すべきであろう．もし，新たにパラメータを増やさないのであれば，一つずつパラメータを減らしてニューラルネット学習をし，推定精度の変化をみていく必要がある．今回は推定のためにニューラルネットを用いたが，これは写像 Ψ が非線形であるためであり，厳密な非線形重回帰分析ができれば，同等の結果が得られると予想できる．しかし，これ自体は非常に扱いが困難である．そこで，以下のような回帰式を仮定して分析を行い，得られた回帰係数を用いて認識精度の推定を行ったところ，今回の結果に近いものとなった．

$$Accuracy = (b_0 + \sum_{i=1}^{I-1} b_i f_i(x_i)) + (b_I + \sum_{j=1}^{J-1} b_{j+I} x_j) snr + (b_{J+I} + \sum_{k=1}^{K-1} b_{k+J+I} x_k) snr^2$$

snr は SN 比を，x_i は SN 比以外の説明変数を，b_i は回帰係数を表す．これは，得られた学習データから SN 比-認識精度グラフを作成したときに，緩やかな曲線になっていたので，SN 比の 2 次関数の増加部分で近似できないかという予測のもとにおいた回帰式であり，厳密な重回帰分析により求めた結果ではない．しかし，図 8 のように図 6 と近い結果が得られている．このときに用いた雑音のパラメータは，スペクトル勾配の平均と音素認識で検出された音素数だけである．このことから，ニューラルネットを用いた場合でも，この二つのパラメータと SN 比があれば，今回程度の結果が得られる可能性が高いと考えられる．

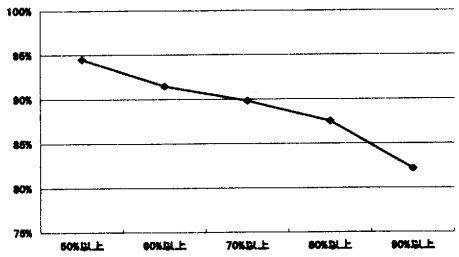

図 8　回帰分析により求めた場合の適合率

5. ま と め

環境に応じ文法・対話戦略の変更を行なうことで，雑音下でも精度よく対話を行えるシステムについて検討した．

語彙・候補数の少ない文法を選択した際に増加する辞書外発話を防ぐため，システムの質問提示法を工夫し，それによりユーザ発話がどのように変化するかを調査した．その結果，質問内指定，音声選択では，高確率でシステムの質問の内容だけをユーザが答えることが確認できた．

後半は，文法選択の基準とするための認識精度の推定にニューラルネットを用いた．この結果，適合率で 80%を示した．今回は，使用した雑音のパラメータ数は多かったが，適切なものだけを選択し最小限で推定できるようにしていく必要がある．

また，認識精度の推定に最も重要な因子となる SN 比は，話者の違いや，マイクとの距離で変化してしまう．それによる推定精度への影響が大きいようであれば，SN 比を見積もる手法について考える必要がでてくる．今後は，このような課題に取り組み，最終的には本機能を対話システムに取り込み評価を行う予定である．

文　献

[1] Yutaka HIROI et al: "A Patient Care Service Robot System Based on a State Transition Architecture", ICMIT,FB1-8(2003)

[2] 大庭他：" マイクロプロセッサを登載した音声認識・合成装置による音声対話システムの構築に関する検討", SICE 講演論文集 (II), pp71-72(2002.12)

[3] 張他：" 区分線形変換による雑音適応法における木構造クラスタリングの検討" , 音講論, pp37-38(2003.3)

[4] 伊藤他：" 目的地設定タスクにおける対話状況の違いによる言語的特徴の分析" , 音講論, pp.65-66(2001.10)

[5] 鈴木他：" フレーム間相関値評価による音声区間検出法を用いた雑音下音声認識", 音講論, pp.143-144(2003.9)

[6] 角谷他：" カーナビの地名入力における誤認識時の訂正発話の分析と検出", 音声言語情報処理 SLP37-11(2001.7.14)

[7] 田熊他：" 種々の音声変動に対応するための並列処理型音声認識システムの研究", 博士論文 (2003.3)

[8] 安田他："2 つの認識文法を用いた主導権混在型対話制御", 音講論, pp.77-78 (2002.9)

[9] 駒谷他：" 音声認識結果の信頼度を用いた効率的な確認・誘導を行う対話監理", 情報処理学会論文誌 Vol43 No10, pp3078-3086 (2002.10)

[10] Yoav Freund,Robert E.Schapire : "A decision-theoretic generalization of on-line learning and an application to boosting", In Europrean Conference on Computational Learning Theory, pp.23-37 (1995)