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

2006－SLP－64（8）
2006／12／21

予測文と部分単語認識の併用による音声対話システムの検討

河上 まきほ† 西田 昌史‡ 堀内 靖雄‡ 市川 熹‡

†‡千葉大学工学部 〒263-8522 千葉県千葉市稲毛区弥生町 1-33

E-mail: †z3t0756@students.chiba-u.jp, ‡{nishida,hory,ichikawa}@faculty.chiba-u.jp

あらまし これまで我々は，対話状態を予測しその状態ごとに発話内容を文単位で予測することで，認識候補数の抑制を行ってきた．しかし，予測文での認識では複合語など比較的長い単語を含む発話において，単語の一部が正しく認識されていたとしても，文全体を再度確認する処理を行っており，誤認識を繰り返すおそれがあった．そこで，本研究では，予測文に含まれる複合語などの比較的長い単語を分割した部分単語認識と予測文認識を並列処理することで，認識候補を絞り込んで認識結果を確定する手法を提案する．カーナビゲーションシステムの目的地設定の場面を想定し評価実験を行った結果，完全一致したときは確認応答を省略し，発話内容を含む部分一致のときは，正しく認識候補を絞り込んで確認応答処理ができる可能性を示した．

キーワード 音声対話システム，予測文認識，部分単語認識，認識候補の絞込み，信頼度

A study on spoken dialogue system based on parallel recognition of prediction sentence and partial word

Makiho KAWAKAMI† Masafumi NISHIDA‡ Yasuo HORIUCHI‡ Akira ICHIKAWA‡

†‡Faculty of Engineering, Chiba University 1-33 Yayoi-cho, Inage-ku, Chiba-shi, Chiba, 263-8522 Japan

E-mail: †z3t0756@students.chiba-u.jp, ‡{nishida,hory,ichikawa}@faculty.chiba-u.jp

Abstract We have proposed a method that predicts user's utterances in spoken dialogue systems by recognizing prediction sentences of each dialogue state, thereby decreasing recognition errors. However, the conventional method might repeat recognition errors because it confirms the whole sentence even if it recognizes only a part of a long word, such as a compound word, correctly. In this study, we propose a method using decoders based on prediction sentences and partial words obtained by dividing long words. The proposed method can confirm user's utterances by selecting a candidate to a partial matched word using recognition results of the prediction sentence and a divided partial word. We conducted experiments in setting a destination of a car navigation system. We demonstrated that it is possible to confirm user's utterances effectively using the proposed method by selecting recognition candidates.

Keyword Spoken dialogue system, Prediction sentence recognition, Partial word recognition, Selection of recognition candidates, Confidence measure

1. はじめに

近年，カーナビゲーションシステムなどで音声対話の技術は実用化されつつある．しかしながら，音声対話においては，認識対象語彙サイズが大きくなると誤認識が生じやすくなる問題があり，音声認識の精度の向上とともに，早期に正しい認識結果を得ることが重要な課題となっている．

こういった問題に対し，これまで，認識結果の信頼度に関する研究として，N-best 文とその尤度を用いた手法[1]や，発話の種類や対話履歴といった文脈情報を用いた手法[2]が提案されている．また，誤認識からの回復に関する研究として，複数の理解候補を保持して対話を進めていき，その過程で候補を絞り誤認識から回復する手法[3][4]や，非文や未知語を含む語の部分的な誤認識を減らすために，ワードスポッティングでキーワードを絞り込み，N-best 候補を更新する手法[5]，未知語を含む認識対象のカバー率などの向上のために，高頻度単語と短い基本単語を併用した音声認識を用いた組織名入力インタフェースの提案[6]がされている．

それらの手法に対して，我々は発話状況に応じて対話状態と各々の状態での発話内容を予測することで，認識対象語彙サイズを抑制する，予測文認識の手法を提案した[7][8][9]．この手法は，発話内容を文単位で登録し，文節単位でモデル化して認識を行い，大語彙認識を併用することで，発話の予測成否判定を行う．また，予測文認識と大語彙認識の認識結果を比較し，一致した場合，信頼度が高いと判断し確認応答を省略することで，円滑

な対話制御を実現した．しかし，予測文に含まれる語彙が大語彙認識の語彙に含まれているとは限らない．しかも，予測文認識では文節単位でモデル化しているために，複合語などの比較的長い単語の認識において，部分的に正しく認識できていても全部を誤認識と判断して文全体を再度確認する処理を行っており，誤認識を繰り返す可能性があった．

そこで，本研究ではより少ない対話数で正しい認識結果を確定するために，予測文認識内の複合語などの比較的長い単語を部分単語に分割してモデル化した部分単語認識と，従来の予測文認識を併用する手法を提案する．本手法により，それぞれの認識結果を比較することで，認識候補の絞込みを行う．カーナビゲーションシステムにおける目的地点名を確定する場面を想定して，提案手法の評価実験を行う．

2. 発話予測に基づく音声対話

我々は，次発話の予測情報を音声対話に利用する方法として，道案内をタスクとして状態遷移モデルを構築し，それに基づいた対話制御を行っている．状態遷移モデルは，「未知情報要求」「確認」「肯定・否定」などの発話単位タグをもとに対話状態を定義し，発話単位タグに基づいた対話状態の予測モデルを構築した．対話制御モデルの概略図を図 1 に示す．

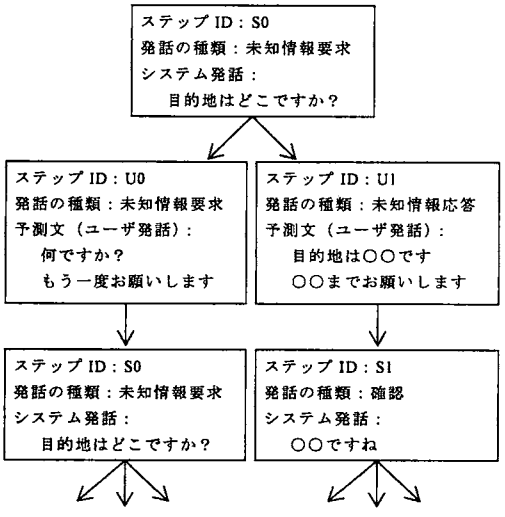

図 1 対話状態遷移図

システム発話ステップには，発話の種類（発話単位タグ）や，その状態でのシステム発話，予測されるユーザ発話ステップ（予測状態）などが登録されていて，保持した複数の予測状態から，ユーザ発話の認識結果によって，予測外の場合を除き，その認識尤度が最尤となった予測文を持つ予測状態へと遷移する．ユーザ発話ステップには，発話の種類（発話単位タグ）や，その状態で予測される文単位のユーザ発話，1 対 1 で対応した次システム発話ステップなどが登録されていて，必ずそのシステム発話ステップへ遷移するようになっている．

音声認識処理では，次発話の予測文候補から認識単位を文節単位に分けて行う．システムがユーザに目的地の発話を要求する状態を例に，認識処理方法の概略を図 2 に示す．図 2 のように，予測されるユーザ発話の予測文はいくつかのスロットを埋める形式となっていて，音声認識というタスクを，予測文を認識するというタスクに置き換えている．図 2 のように，システムがユーザに目的地の発話を要求する状態では，スロットに，予測文候補，フィラー，目的地点名があり，予測文候補列から，予測文の認識が可能な認識用辞書と言語モデルを作成する．言語モデルは，文節単位の bi-gram で構築し，言語尤度は用いていない．つまり，文法ベースの認識手法と等価である．言語モデルを作成する際，自然発話に対応するため，倒置，文節の省略，助詞の欠落，言い淀みの挿入といった処理を行っている．そして，それらを用いて認識を行う．

ID	U1	発話の種類	未知情報応答（地点名）
予測文		・<F>＋目的地は <PN>です ・<F> <PN>まで ＋お願いします ・<F> <PN>に ＋行きたいんですけど	
フィラー <F>		・えっと　・あー ・えー　・(なし)	
地点名 <PN>		地点名辞書： 京葉銀行 西武百貨店 千葉大学 東京駅 札幌ドーム 千葉パルコ 茶畑交番 富士急百貨店 ・・・	
認識結果		えっと 目的地は 千葉大学です	

図 2　予測文認識の認識方法

3. 予測文認識と部分単語認識の併用による音声対話

本研究では，音声対話場面として，カーナビゲーションシステムにおける目的地を設定する対話場面を想定しているため，今回，部分単語認識器での認識は，目的地点名の認識時のみに限定して行っている．

我々が従来から用いている予測文認識の，地点名が登録された辞書には，千葉大学，京葉銀行，西武百貨店といった正式名が登録されている．一方，部分単語認識の辞書には，予測文認識の辞書に含まれている地点名を 2 つに分割した部分単語が登録されていて，前方辞書には，

千葉，京葉，西武といった部分単語が，後方辞書には，大学,銀行,百貨店といった部分単語が登録されている．

部分単語認識器の認識結果は，前方辞書と後方辞書，それぞれの辞書を用いた認識結果を合わせたものとなる．例えば，前方辞書を用いた認識結果が［千葉］，後方辞書を用いた認識結果が［大学］であった場合，部分単語認識器の認識結果は［千葉 大学］となる．

この予測文認識と部分単語認識を併用して地点名の認識を行う．ユーザの発話が入力されると，各認識器はそれぞれの辞書を用いて認識結果を出力する．これら認識結果を比較すると，前方と後方の両方が一致する完全一致，前方もしくは後方のみが一致する部分一致，前方も後方も全く一致しない完全不一致に分類できる．それぞれの場合において，予測文認識と部分単語認識の結果を比較することで，認識候補を絞り込んで認識結果を確定する．

提案手法において，認識結果が完全一致した場合の対話例を図3に，部分一致した場合の対話例を図4に，完全不一致だった場合の対話例を図5に，それぞれ示す．

図3のように完全一致した場合，二つの認識器が同じ認識結果を出力したことから，予測文認識だけを用いるよりも認識結果の信頼性は高くなると考えられる．よって，一致した認識候補に認識結果を確定し，確認応答を省略する．また，図4のように部分一致した場合，一致した部分単語の信頼度が高いとして，その部分単語を含む地点名のみに予測文認識の辞書を変更して認識候補を絞り込む．そして，同じ発話を再認識して得られた結果を認識結果と確定して確認応答をする．この時，認識候補を絞った前後で認識結果が一致すれば部分一致の信頼度は高いと考えられる．また，図5のように完全不一致だった場合，予測文認識の辞書の絞込みを，予測文認識と部分単語認識のどちらかに正解が含まれる可能性を考慮して,予測文認識結果の前方部分単語と後方部分単語，部分単語認識結果の前方部分単語と後方部分単語の，4つの部分単語を含む地点名に拡張して，再認識を行い得られた結果を認識結果と確定して確認応答をする．

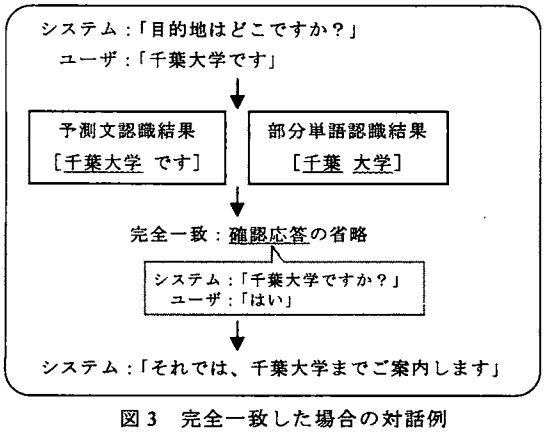

図3 完全一致した場合の対話例

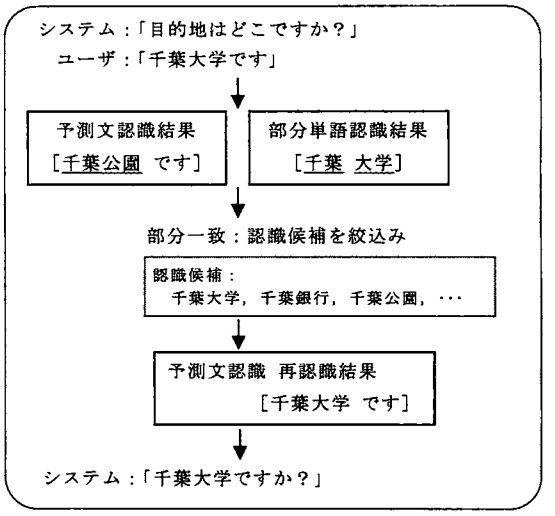

図4 部分一致した場合の対話例

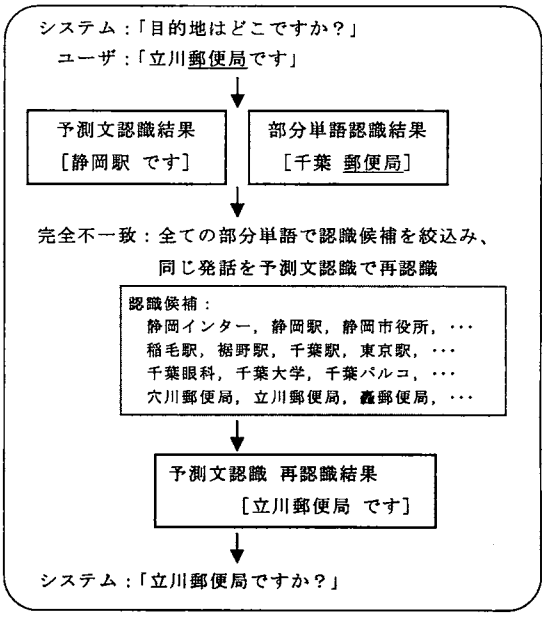

図5 完全不一致だった場合の対話例

4. 評価実験

4.1. 実験条件

本研究では，カーナビゲーションシステムの目的地を設定する対話場面を想定した予測文認識と部分単語認識を併用した音声対話について，事前に収録した音声データを用いたシミュレーションで評価実験を行った．

予測文認識の地点名を登録した辞書には，正式名称250箇所（例：千葉大学，京葉銀行，富士急百貨店，東京駅）が登録されている．部分単語認識の2つの辞書は，前方辞書には77個（例：千葉，京葉，富士急，東京），後方辞書には66個（例：大学，銀行，百貨店，駅）の部分単語がそれぞれ登録されている．

被験者は10名で，予測文認識の辞書に含まれる地点名のうち80箇所を，「千葉大学です」というように5回ずつ発話してもらい，合計4000発話を収録した．

デコーダには，julius3.1を用いている．

予測文認識と部分単語認識のそれぞれを，単独で用いた場合の認識率は，予測文認識 75.1% (2986/3978)，部分単語認識 80.8% (3213/3978)であった．ここで，全4000発話のうち，認識ができなかった音声データが22個あり，これらのデータは実験結果から除いている．また，対話制御においては，次の発話状態へと正しく遷移できれば問題ないので，語尾の「～です」や「～まで」などの言い回しについては認識精度の評価対象から除き，地点名が正しく認識されていれば正解として扱った．

4.2. 予測文と部分単語認識の併用による認識結果

予測文認識と部分単語認識を併用した場合に，それぞれの認識結果が一致した割合を表1に示す．ここで，完全一致における正解とは，ユーザが発話した地点名と2つの認識器の認識結果がすべて一致していたもの（例：ユーザ発話「千葉大学です」，予測文認識結果［千葉大学です］，部分単語認識結果［千葉 大学］）を指し，部分一致における正解とは，ユーザが発話した地点名の一部と，2つの認識結果の一致した箇所が一致していたもの（例：ユーザ発話「千葉大学です」，予測文認識結果［千葉公園 です］，部分単語認識結果［千葉 大学］）を指している．

表1 予測文認識と部分単語認識の併用による認識結果の一致度

	完全一致	部分一致
一致度	72.8% (2897/3978)	9.7% (384/3978)
正解率	99.2% (2874/2897)	44.0% (169/384)

結果から，予測文認識と部分単語認識の認識結果が完全一致したものは，全体のうち72.8%あり，そのうちの99.2%がユーザの発話と同じであった．この結果から，予測文認識だけを用いるよりも認識結果の信頼度は高いと考えられる．よって，従来の予測文認識で，目的地などが発話された際に行っていた確認応答を省略して，次の対話ステップに進めることのできる可能性が示せた．

また，部分一致したものは全体のうち9.7%あり，そのうちの44.0%はユーザが発話した地点名の一部と一致しており，これらは，次発話で使用する予測文認識の辞書を，正解を含む地点名のみに絞り込むことが出来るので，次発話で正解を得ることが出来る可能性が示せたが，若干精度が低い結果となった．

4.3. 部分一致における誤認識傾向の分析

部分一致した発話のうち，前方での部分一致（前方一致）と後方での部分一致（後方一致）のどちらが正しく予測文認識の辞書を絞り込む割合が高かったのか分類した結果を，表2に示す．ここで，正しい部分一致とは，ユーザの発話内容が含まれる部分一致（例：ユーザ発話「千葉大学です」，予測文認識結果［千葉大学 です］，部分単語認識結果［千葉 大学］，部分一致「千葉」）であり，誤った部分一致とは，ユーザの発話内容が全く含まれない部分一致（例：ユーザ発話「千葉大学です」，予測文認識結果[理科会館 です]，部分単語認識結果[理科 眼科]，部分一致「理科」）である．

表2 前方一致と後方一致の割合

	前方	後方	合計
正しい部分一致	64	105	169
誤った部分一致	215	0	215

表2から，今回の実験では誤った部分一致はすべて前方一致であったことが分かる．

部分一致したということは，2つの認識器が同じ認識結果を出力したということで，その認識結果の正解／不正解に関わらず，どちらの認識器も違う認識結果を出力した場合に比べ，認識結果の信頼度が高いと考えられる．そこで，誤って部分一致した語の傾向の分析をした．

誤った部分一致の発話ファイルの音声を聴取して調べた結果，語頭での音声の大きさ（レベル）が小さいために，語頭が認識されなかった可能性の高い発話が177発話あった．表3に，語頭の音量が小さいために認識されなかった可能性の高い発話の認識結果の例を示す．

表3 語頭が認識されなかった可能性の高い誤認識例

正解単語	誤認識結果
すその	さんのう
きさらづ	さつき
たちかわ	ちば
さっぽろ	ポート

また，誤って部分一致した発話のうち，誤認識率の高いものを調べた結果を，表4に示す．表3，表4から誤認識の傾向を見ると，サ行・ハ行・タ行で始まる語の認識率が低いことが分かる．これらサ行・ハ行・タ行の語は，他の音に比べて音声の大きさが小さく，その結果，語頭

の認識に失敗し，その後に続く語の認識も失敗したと考えられる．図6に，誤認識の多かった「裾野消防署」の音声波形の1つを示す．

表4 前方で誤認識率の高かった部分単語

部分単語	人数	誤認識率
裾野	4	38%(19/150)
立川	8	24%(12/ 50)
富士急	7	18%(18/100)
札幌	4	14%(7/ 50)
京葉	7	10%(10/100)

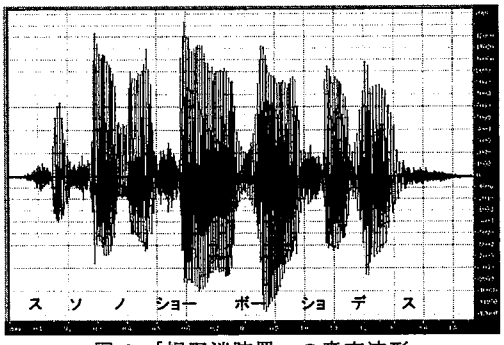

図6 「裾野消防署」の音声波形

また，母音系列が同じ，もしくは似ているものに誤認識しているものも多く，例を表5に示す．表中の()内は，音素の一例であり，下線部が母音系列の同じ箇所である．加えて，表4に示したように「富士急」や「立川」は語頭が認識されづらい単語であり，表5の誤認識結果にもそれが現れていると考えられる．よって，今後，辞書の絞込み手法として，母音系列の類似度を用いる手法が有効である可能性もうかがえる．

以上から，前方に誤った部分一致が多い原因として，発音しづらい音素を持つ単語の語頭での音量が小さいために認識されづらい可能性が高いことが示された．

本研究では，部分単語認識を併用することの有効性を検証したいので，部分一致した384発話中，語頭のレベルが小さかったために誤って部分一致した可能性の高い177発話は除いた207発話を用いて，今後の検討をする．

なお，誤って部分一致した以外にも，語頭の音量の小さいものもあったが，認識率等は，これらの発話を含んで算出している．

表5 母音系列の似ている誤認識例

正解単語	誤認識結果
京葉(keiyou)	西武(seibu)，平和(heiwa)
富士急(fujikyuu)	りんくう(rinkuu)
立川(tatikawa)	千葉(tiba)

4.4. 部分一致における認識結果の確定

部分一致した384発話のうち，語頭のレベルが小さかったために誤って部分一致した可能性の高い177発話を除いた207発話を用いて，絞り込んだ辞書で再認識をした結果を，表6に示す．結果から，部分一致して辞書を絞り込んだ後，同じ発話を再認識すると64.3%が正解を得ることができた．

表6 同じ発話での再認識の結果

	前方一致	後方一致	合計
正解率	41.2% (42/102)	86.7% (91/105)	64.3% (133/207)

また，1回目の予測文認識の認識結果と，部分一致により絞り込んだ辞書を使用した2回目の予測文認識の認識結果を比較して，その結果が一致／不一致だった場合の認識正解率について調べた結果を，表7に示す．

表7 1回目と2回目の予測文認識の認識結果の比較による認識正解の割合

	前方一致	後方一致	合計
一致	63.2% (36/57)	96.6% (28/29)	74.4% (64/86)
不一致	11.1% (5/45)	82.9% (63/76)	56.2% (68/121)

結果から，1回目と2回目の予測文認識の認識結果が一致した場合，74.4%がユーザの発話内容と一致することが分かる．認識候補を絞る前後で結果が同じであれば部分一致の信頼度が高いと考えられるので，1回目と2回目の予測文認識の認識結果を比較して一致した場合，正解である可能性が高く，今後，前方の誤った部分一致の対策ができれば，確認応答を省略できる可能性があることが示された．

4.5. 完全不一致における認識結果の確定

完全不一致になった発話を調べた結果を表8に示す．

表8から，完全不一致の場合でも，予測文認識結果が正解の場合や部分単語認識結果が正解の場合，どちらも不正解だが部分的な正解を含む場合が，全完全不一致中84.1% (586/697)あることが分かった．

そこで，完全不一致だった場合に，予測文認識の辞書の絞込みを，予測文認識結果の前方部分単語と後方部分単語，部分単語認識結果の前方部分単語と後方部分単語の，4つの部分単語を含む地点名に拡張して再認識を行った．例えば図5では，予測文認識結果[静岡駅 です]，部分単語認識結果［千葉 郵便局］から，「静岡」，「駅」，「千葉」，「郵便局」の4つの部分単語を含む地点名を再認識の辞書に登録して，再認識を行う．

表 8 完全不一致だった発話の分類

予測文認識結果	部分単語認識結果	ファイル数[発話数]
正解	不正解	35
不正解	正解	292
どちらかの結果に部分的な正解が含まれる		259
どちらの結果にも部分的な正解が含まれない		111

結果，完全不一致だった 697 個の音声ファイルは，予測文認識のみを用いた場合の認識率が 5.0% (35/697)であったのに対し，絞り込んだ辞書を用いて再認識をした結果，60.5% (422/697)が正解となった．よって，完全不一致だった場合，絞り込み方を拡張した辞書での再認識の結果をユーザに確認することで，ユーザの言い直しの繰り返しを防ぐことができる可能性を示した．

5. おわりに

本研究では，カーナビゲーションシステムにおける目的地を設定する対話場面を想定して，予測文認識と部分単語認識の併用による音声対話手法を提案し，その有効性について検討した．提案手法により，認識結果が完全一致した場合は確認応答を省略し，部分一致と完全不一致の場合は認識候補を絞り込んで再認識することで，より少ない対話数で正しい認識結果を確定できる可能性を示した．

今後は，提案手法と検討事項をもとに，尤度や，例えば母音系列などの類似語彙も考慮に入れた認識結果の一致度による信頼度について検討する．そして，実際のカーナビゲーションシステムを想定して語彙数を増やした辞書を用いて，提案手法を音声対話システムに実装して評価実験を行っていく予定である．

謝　辞

本研究は，富士重工業株式会社との共同研究により実施した．

文　献

[1] 駒谷 和範，河原 達也，“音声認識結果の信頼度を用いた効率的な確認・誘導を行う対話管理，”情報処理学会論文誌，Vol.43,no.10,pp.3078-3086, Oct.2002.

[2] 水谷 誠，伊藤 敏彦，甲斐 充彦，小西 達裕，伊東 幸宏，“音声認識の信頼度と対話履歴を利用した最尤推定型言語理解，” 情報処理学会研究報告，SLP-45-19,pp.113-118,Feb.2003.

[3] 北岡 教英，角谷 直子，中川 聖一，“音声対話システムの誤認識に対するユーザの繰り返し訂正発話の検出と認識，” 電子情報通信学会論文誌（D-II），Vol.J87-D-II，No.3，pp.799-807,2004.

[4] 北岡 教英，矢野 浩利，中川 聖一，“誤認識の修復のための自然で効率的な音声対話戦略，” 情報処理学会研究報告，SLP-61-7, pp.37-42,May2006.

[5] 鈴木 貞之，小暮 悟，伊藤 敏彦，甲斐 充彦，小西 達裕，伊東 幸宏，“頑健な言語理解のための文法とワードスポッティングを併用した音声認識手法の検討，” 電子情報通信学会技術研究報告，SP-105-138,pp.25-30,Dec.2005.

[6] 北岡 教英，押川 洋徳，中川 聖一，“孤立単語認識と連続基本単語認識の併用に基づく組織名の音声入力インタフェース，” 電子情報通信学会技術研究報告，SP2005-110, pp.31-36,Dec.2005.

[7] 玉井孝幸，堀内靖雄，市川熹，“音声対話システムにおける発話予測を利用した音声認識，”情報処理学会研究報告，2002-SLP-43,pp.1-6,Oct.2002.

[8] 西田 昌史，寺師 弘将，堀内 靖雄，市川 熹，“ユーザ発話の予測に基づく音声対話システム，”電子情報通信学会技術研究報告，SP2004-132,pp.61-66, Dec.2004.

[9] 寺師 弘将，西田 昌史，堀内 靖雄，市川 熹，“複数の言語モデルの並列認識に基づく発話の予測判定に関する検討，” 日本音響学会 2006 年春季研究発表会，2-11-11,pp.117-118,Mar.2006.

[10] 西田 昌史，河上 まきほ，寺師 弘将，堀内 靖雄，市川 熹，“予測文と部分単語認識の併用による音声対話，” 日本音響学会 2006 年秋季研究発表会，3-2-9,pp.89-90,Sept.2006.