複数の音声認識器に基づく効果的な音声理解手法

嶋田和孝　　堀口怜美　　遠藤　勉

九州工業大学 情報工学部 知能情報工学科

概要

本論文では，複数の音声認識器を併用することで，ロバストでかつ柔軟な発話にも対応可能な音声理解手法について提案する．具体的には，特定のタスクに依存した認識器と一般的な言語モデルに基づく認識器の性質の異なる 2 つの認識器を用意し，その認識結果を分別もしくは統合する枠組みについて実験的に考察する．提案手法では，2 つの音声認識器の出力する結果の信頼度や出力数，および出力結果の類似性を利用して，どちらの認識器の結果を信頼すべきかを判定する．実験結果により，提案手法は，想定した発話について高い認識率を実現し，さらに想定された発話かそうでないかについて高い判別能力を持つことを確認した．

A simple and effective speech understanding approach with several speech recognizers

Kazutaka SHIMADA, Satomi HORIGUCHI, Tsutomu ENDO

Department of Artificial Intelligence, Kyushu Institute of Technology

Abstract

This paper proposes a simple and effective method for speech understanding. The method incorporates some speech recognizers. In this paper, we use two recognizers, a large vocabulary continuous speech recognizer (LVCSR) and a domain-specific speech recognizer. We use the LVCSR for recognition of spontaneous utterances and the domain-specific recognizer for robust recognition of utterances for a domain. The integrated recognizer is a robust and flexible method for speech understanding. We use the following criteria for the integration process: (1) a confidence measure computed from each recognizer, (2) the number of outputs in each n-best list, (3) particular information in the outputs and (4) edit distance of each output sentence. The experimental results show the effectiveness of the proposed method.

1　はじめに

近年の音声認識技術の向上により，音声入力を用いた実用的な対話システムの実現を目指した研究が進められている．しかしながら，実用的な音声対話システムの構築には，音声認識誤りへの対応や音声認識そのものの精度向上が不可欠な状況である．

音声認識の精度向上のための一つのアプローチは，キーワードやキーフレーズの抽出に基づく発話理解である [4, 12]．もう一つのアプローチとしては，対象となるタスクやドメインにかなり限定した文法や言語モデルを利用することである．しかしながら，このようなアプローチは音声認識・理解の精度向上には繋がっても，対象外の発話や予期しない形の発話には対応できず，十分とはいえない．

後藤 [9] は，従来は無視されることの多かった非言語情報（言い淀みや韻律情報など）を積極的に利用することで,「音声スタータ」や「音声スポッタ」といった音声の持つ潜在能力を引き出した音声インターフェースが実現できることを報告している．また，Ogata ら [3] は，音声認識の際に生じる競合候補をユーザに視覚的に提示することで，効率的に訂正する手法の有効性を報告している．しかしながら，これらのアプローチはそのルールの習得は容易であっても，人間側の意識的な音声発話や補助が必要であり，どのような音声対話にも適用できるというわけではない．

音声認識の精度を上げるための効果的な手法の一つは，複数の音声認識器を統合的に，もしくは選択的に用いることである．磯部ら [5] は，話題に依存した複数の音声認識器を利用し，その結果を選択的に用いることで，複数の話題に対応可能な音声認識システムについて報告している．この手法により，単語正解率は若干向上している．さらに，別の話題を追加する場合，その追加する話題に関してのみ音声認識器を構築するだけで良いため，システムの拡張が容易という利点もある．しかし，この手法は，それぞれのタスク依存の認識器が基本的に排他的な状態であり，相互の情報を統合的には扱えない．さらに，タスクに依存しないような自由発話部分についても柔軟に処理できないという問題点がある．一方で，統合的に扱う手法では，複数の音声認識器を併用し，その一致部分を検出することで，信頼度の高い単語の特定や，音声認識誤り箇所を推定している [6, 7, 11, 15]．これらの研究は単語レベルでは精度が向上しているが，発話単位では処理をしていない．また，汎用的な認識器を用いた場合，ドメインに依存する言語モデルを利用した場合と比べて，認識精度が低くなる傾向もある．

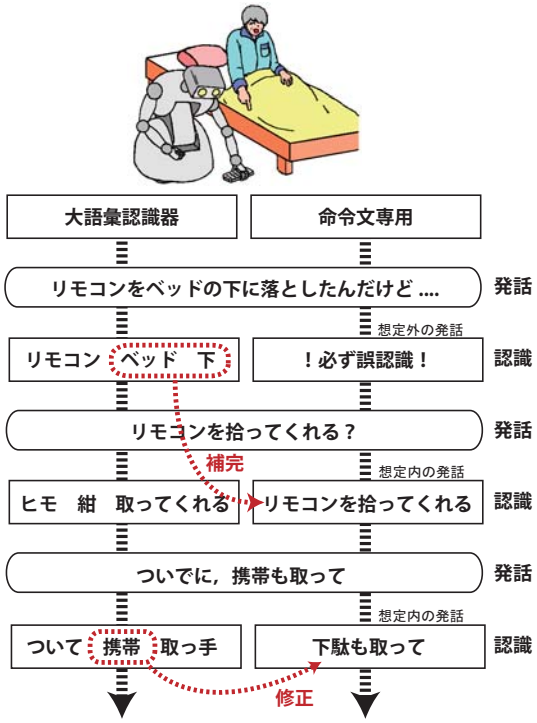

図 1: 提案システムのイメージ．

本論文では，複数の音声認識器を併用することで，特定のタスクやドメインにロバストでかつ自由発話にも柔軟に対応可能な音声理解手法について提案する．我々は現在，介護施設や病院などで暮らしている施設入居者の日常生活や施設スタッフの作業を支援する施設内生活支援ロボットの構築を進めている．本研究はその入力部の 1 つである音声認識部にあたる．この論文では，タスク依存の認識器と一般的な言語モデルに基づく認識器の 2 つを利用する場合について議論する．タスク依存の認識器は，主にユーザからの指示を理解する役割を持ち，正確な認識が可能なことが望まれる．一方で，タスク依存の認識器では十分に認識できないような発話に対しては，大語彙認識器を用いて，認識を行う．それぞれの認識結果を選択的にもしくは統合的に扱うことができれば，柔軟でかつロバストな認識器を実現できる．図 1 に提案システムのイメージを示す．提案手法は，それぞれの認識器の出力する結果の信頼度や出力数，および出力結果の類似性を利用して，認識器の選択や統合などを行う．提案手法については様々な点で評価が必要となるが，本論文では特に，(1) 命令文専用の認識器の認識率と，(2) 入力発話に対してどちらの認識器の結果の信頼度が高いか[1]，の 2 点について実験的に評価す

[1] すなわち，どちらの認識結果が最終的な出力としてふさわしいかを意味する．

る．(1) の評価尺度は，ロボットへの入力として，命令発話に対して認識精度が高いことは必須であることに起因する．(2) については，命令発話に対して高い認識率が得られている場合はその認識結果を大語彙認識器が改悪しないことが重要なためである．

2　認識器と文法

本論文では，2 つの音声認識器として，Julius とその派生形である Julian を利用する．Julius は数万語の語彙を対象とした音声認識ソフトウェアである [8]．本論文では，Julius を前節で述べた大語彙音声認識器として利用する．音響モデル，言語モデルとも同ソフトウェアに添付されているオリジナルのモデルを利用する．

タスク依存の音声認識器としては，Julian を利用する．Julian は Julius を基にした記述文法による音声認識器である．言語モデルとして有限状態文法（DFA）を用い，ユーザが BNF 風の記述で，認識用の構文規則を作成可能である．語彙辞書として，文法ファイル中で使用される単語カテゴリとそれに属する単語およびその読み（音素列）を記述する．音響モデルは Julius と同一のものを利用する．本研究では，ユーザからの命令発話を想定し，下記のような文を受理可能な文法を作成した．

- 「〜を〜して（ください | くれる）？」
  e.g.,「携帯を取って」
- 「〜を〜したい」
  e.g.,「お菓子が食べたい」
- 「〜を〜（ください | くれる）？」
  e.g.,「それをください」
- 「〜にある〜を〜して（ください | くれる）？」
  e.g.,「机の上のリモコンを取ってくれる？」
- 名詞のみ

定義した実際の文法の例と遷移関係を図 2 に示す．ここで，文法は，精度向上のために若干の意味的要素を踏まえて作成されている[2]．例えば,「飲みたい」や「食べたい」という動詞の前には，食べられるものもしくは飲むことができるものに関する名詞のみを受理できるように作成されている．現在の語彙数は 168 語（単語カテゴリ 57）で，文法のルールは 108 個[3]である．

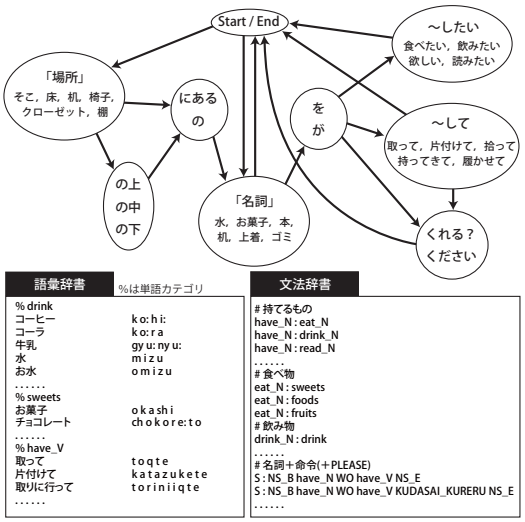

図 2: Julian に実装した文法.

3　提案手法

3.1　問題定義と位置づけ

複数の特性の異なる認識器を組み合わせて処理を行う場合，どの認識器の結果を最終的に利用するか，という問題が発生する．今回のシステムでは，入力された発話がロボットへの命令発話なのか，それとも雑談やそれ以外の発話なのかを分別する必要がある．これは，大語彙認識器（Julius）とタスク依存認識器（Julian）のどちらの認識結果を採用するか，という問題になる[4]．

前述の磯部ら [5] は，最適な認識結果を選択するために各認識器の言語モデルのエントロピーを用いてスコア補正を行っている．しかし，磯部らのシステムは，各認識器が異なる特性を持っており，本システムにおける大語彙認識器とタスク依存認識器とはその関係が異なる．我々の研究と同様にシステムへの問い合わせと雑談部分の判別を行う研究も存在する [10, 14]．佐古ら [10] は音声認識器から得られた言語情報に基づき，AdaBoost によって問い合わせと雑談の判別を行っている．山形ら [14] は，様々な音響特徴や言語特徴を利用して，SVM で分類器を作成し，判別を

[2]図中での%drink や%sweets がそれにあたる．どのレベルまでこのような意味素性を細分化するかは，精度と管理コストの関係を考え，十分議論する必要があるが，今回は最低限必要なレベルを主観的に判断し作成した．

[3]ただし，このうち 61 個は非終端記号と単語カテゴリの対応付け相当である．例えば，図 2 の「eat_N:sweets」のようなもの．

[4]例えば，図 1 で,「リモコンを拾ってくれる？」という入力について，どちらの認識器の結果を採用するかは大きな問題である．

行っている．一般に，機械学習のアルゴリズムを使用する場合，高精度な分類器を作成するためには，十分な訓練データが必要となる．しかしながら，訓練データの収集は，本質的に人間にとってコストの高い作業である．次節において，我々は認識結果から容易に得られる特徴を用いて，シンプルでかつ効果的な手法を提案する．

3.2 分別手法

提案手法では，(1) 音声認識器の出力する信頼度，(2) 出力される候補の数，(3) 出力結果に含まれるショートポーズ記号の有無，(4) 出力された結果の差異，の 4 つの指標を利用して，2 つの認識器のどちらの結果を採用するかを判別する．

3.2.1 信頼度

1 つめの指標は Julius/Julian が出力する単語信頼度である．単語信頼度は 0.0〜1.0 の範囲で各単語ごとに算出され，1.0 に近いほど，その単語に似た競合候補がなかったことを表している．単語信頼度算出アルゴリズムは，単語事後確率に基づく手法の一種で，探索中の部分文仮説のスコアから事後確率を近似的に算出している [2]．

この信頼度はあくまで，認識エンジン自身が与えられた言語モデル・音響モデルの元で算出する確信度ではあるが，有効な特徴の 1 つとして考えられる．

3.2.2 候補数

2 つめの指標は，タスク依存認識器の出力する結果の数である．タスク依存認識器は，想定される命令発話について，ロバストな認識が可能なように少ない語彙・文法で設計されている．そのため，想定している命令発話が入力された場合，正しい認識が行われ，競合候補そのものの数が少なくなる傾向がある．

一方で，想定外の発話，すなわち命令発話以外の雑談部分などが入力されると，そもそも保持している語彙や文法では対応できないため，音響モデルなどを基に，数多くの誤認識結果が出力されるはずである．よって，タスク依存認識器の出力数は，入力が命令発話か否かの有効な特徴になると考えられる．

3.2.3 ショートポーズ記号

Julius/Julian は発話中に生じる短い無音区間（ショートポーズ）をモデル化している．入力に該当する部分があると判断した場合は定義したショートポーズ記号を出力する．実験的に調査したところ，このショートポーズ記号が，タスク依存認識器に想定外の発話を入力した場合に頻出するという傾向が得られた[5]．そのため，このショートポーズ記号の有無を特徴の 1 つとして採用した．

3.2.4 編集距離

最後の指標は，2 つの認識器の出力結果の差異である．人間でさえも似たような音を持つ単語は誤認識することがあり [1]，音声認識器にも同様の傾向があることは自明である．

前述のように，タスク依存認識器は，もし入力が想定された発話であれば，高い認識精度を得られるはずである．一方，大語彙認識器は，自身の音響モデル・言語モデルに基づいて，最適と思われる単語列を出力する．この場合，2 つの認識器の結果は，少なくとも音素レベルでは比較的類似していると考えられる．逆に，想定外の発話の場合，大語彙認識器は命令発話を認識したときと同様に，最適と思われる単語列を出力するが，タスク依存認識器は持ち合わせている語彙と文法を基にあまり適切でない単語列を出力すると考えられる．すなわち，この場合は，2 つの認識結果は，音素レベルでも必ずしも一致しない可能性がある．

そこで，提案手法では，2 つの認識結果の編集距離を算出し，それを分別のための特徴とする．編集距離のような一致度は，1 節でも示したように，複数の認識器の出力の高信頼度部分の推定 [7] や誤認識部分の抽出 [6] などによく使用される指標である．Komatani ら [1] も音声対話システムにおいて，人間が聞き取りにくいと思われる音的に類似した単語ペアの抽出に編集距離を基にした距離値を計算している．

本論文では，それぞれの認識器の出力全体（すなわち発話単位）での編集距離と，各単語ごとの編集距離の 2 つの距離値（ともに音素レベル）を DP マッチングにより算出し，分別処理に利用する．単語単位で編集距離を求める場合，2 つの認識結果から編集距離を計算するペアを決定する必要がある．本論文では，まず完全一致する単語を検索し，それを編集距離計算から除外した後，残りの単語群について総当たりで編集距離を求める．その計算の結果，最も編集距離の低い組み合わせの距離値をそれぞれ採用している[6]．具体的な例を図 3 に示す．図において，破線部は完全一致した箇所を表し，矢印に付属する数値は

[5] タスク依存認識器が自身の語彙と文法で取り扱えない発話が入力されたときに，解析できない場所をショートポーズだとすることで解析可能だと認識した場合に，ショートポーズがあると解釈し，出力したのではないかと考えられる．

[6] ただし，組み合わせを求める際には隣接する 3 つの単語までという制約を設けている．

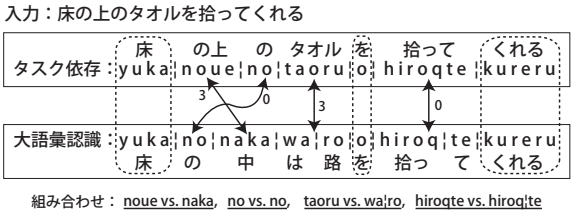

図 3: 編集距離の計算例.

そのペアの編集距離に相当する.

3.2.5 ルール

使用する特徴を段階的に適用し，入力発話が命令発話かそうでないかを分別する．具体的な処理の流れを以下に示す.

1. タスク依存認識器の結果にショートポーズ記号が存在する，もしくは信頼度が算出されていない（すなわち 0 になる）
   →大語彙認識器の結果を採用

2. タスク依存認識器の出力候補数が閾値以下（現在の閾値は 9）
   →タスク依存認識器の結果を採用

3. 発話単位の編集距離[7]が th_{utter} 未満
   →タスク依存認識器の結果を採用

4. 単語単位の編集距離[8]の平均値で比較
   平均値 $\geq th_{word}$ →大語彙認識器
   平均値 $< th_{word}$ →タスク依存認識器

4 実験

本論文では，(1) タスク依存認識器の認識精度と (2) 認識器の選択精度の 2 点について検証した．まず，命令発話の認識精度について述べる．実験に使用する命令文は，タスク依存認識器に実装した文法からランダムに生成された 200 文から，意味的に問題がないものを 50 文選択して利用した[9]．この 50 文を 2 名の被験者（男女 1 名ずつ）に発話してもらい（計 100 発話），その認識率を評価した．実験結果を表 1 に示す．ここで，一部不正解とは，例えば「〜ください」の部分の欠落などを指し，意味的な解釈のレベルでは問題のない事例を指す．不正解とは，名詞などの部分を間違えた場合を指す．すなわち，実質的な認識精度としては 97%であり，タスク依存認識器は非常にロバストな認識器であることが確認された.

表 1: タスク依存認識器の認識精度

分類	発話数
完全一致	94 発話
一部不正解	3 発話
不正解	3 発話

表 2: 分別精度

種類	適合率	再現率	F 値
命令発話	0.888	0.965	0.924
非命令発話	0.962	0.877	0.917

次に，認識器の選択精度について述べる．実験には先ほどタスク依存認識器の評価に用いた 50 文と，タスク外の発話（すなわち非命令発話）として，英会話のテキスト中の日本語訳部分からランダムに選んだ会話文相当を 50 文の計 100 文を用意した．被験者は 4 名（男女 2 名ずつ）で，命令発話およびそれ以外の発話で合計 400 発話を評価対象とした．実験は 2 名分を訓練事例（編集距離のための閾値設定に利用），残りの 2 名分をテスト事例とし，これを全ての組み合わせ，すなわち 6 通りのデータで交差検定的に評価した．実験結果を表 2 に示す．それぞれの値は，交差検定の結果の平均値である．実験において，th_{utter} は 0.24〜0.26 の範囲で，th_{word} は 0.08〜0.13 の範囲であった．これは，編集距離のルールで使用する閾値は事例によって大きく揺れないこと表している．表からわかるように，提案手法は高い分別能力があることがわかる．さらに，各ルールでの正否の事例について検証した．その内訳を表 3 に示す．その結果，編集距離に比べ，その他の指標は若干精度が悪いことが判明した．そこで，これらのルールを削除し，編集距離のみを使用した場合の実験を行った．その結果を表 4 に示す．信頼度などの指標で間違っていた事例の殆どが，発話単位の編集距離によって適切に分別されていることを確認した．実験結果より，提案手法は，シンプルな手法だが，極めて効果的であることがわかる.

[7] 正確には得られた編集距離を発話中の音素数で割ったもの.

[8] 3. と同様に，得られた編集距離を単語中の音素数で割ったもの.

[9] 文生成には，Julian に装備された機能を使っている．構築した文法には前述のように若干の意味的制約が組み込まれているが，その制約は緩いため，必ずしも意味的に正しい文が生成されるとは限らない．例えば,「冷蔵庫の中のアイロンを取って」のような文が生成されることがあるため，意味的に正しいもののみを著者らが選択した.

表 3: 各ルールの正解率

種類	正解数	不正解数
信頼度	48	15
ショートポーズ	93	6
候補数	384	57
編集距離（発話）	192	14
編集距離（単語）	388	3

表 4: 分別精度（2 種類の編集距離のみ）

種類	適合率	再現率	F 値
命令発話	0.963	0.985	0.974
非命令発話	0.985	0.963	0.974

5 考察

本節では，提案手法について考察する．まず，認識器の精度について議論する．現在は語彙が少ないこと，受理できる文法が制限されていることなどにより，かなり高い精度を得ている．しかし，状況によっては，語彙や文法を拡充する必要が出てくると考えられ，その結果，精度の低下は否めない．そこで，タスク依存認識器のさらなる精度向上について考える必要がある．最もシンプルなアプローチは，タスクに準じた言語モデルの使用である．Julian はタスク用の文法を記述できるが，それは必ずしもタスク用の言語モデルを使用することと等価ではない．我々は，他のタスクで，対話コーパスから得られた言語モデル相当を Julian の認識結果に適用し，出力された n-best 候補の中から最適な組み合わせを抽出することで精度が向上することを確認している [4]．提案手法のさらなる精度向上には，対話コーパスの収集と解析，そしてその結果の適用が不可欠であると考えられる．

現在の文法では，単文相当の発話しか扱っていない．複数の命令を含む発話や従属節を含むような発話には対応できない．Julian の文法を改良し，それらを受理できるようにすることも可能だが，複雑な文法を構築し，それを利用することは，タスク依存認識器を使用するメリットを失うことに繋がる[10]．さらに，長い文章の場合，人間の発話では，文の切れ目で無音区間が生じることが多く，それを受理できる文法を作成しても，入力の段階で複数の発話とみなされ，適切に扱えない場合もある．このような状況については，提案手法のように，それぞれの目的に沿ったタスク依存認識器を複数構築し，それを併用して，選択もしくは統合する方が効果的である．また，状況や話者に応じてタスク依存認識器を切り替えるという手法も効果的であり，提案手法はそのような場面にも有効であると考えられる．現在我々は，プロジェクトの一環として，発話者や発話区間の推定に関する研究も行っている [13]．この研究と提案手法を組み合わせ，状況に応じて認識器や言語モデルを適宜選択することができれば，さらなる精度向上が見込める．

本論文では，2 つの認識器の選択方法については実験的に検証したが，統合的に扱う手法については十分議論されていない．誤認識箇所の推定や補完（図 1 の「携帯→下駄」）や大語彙認識器を利用した文脈形成（「ベッドの下」の拡充）については今後十分な考察が必要となる．これらのアプローチについては，併用する大語彙認識器の認識精度も問題になる．内容語について大語彙認識器の認識精度を確認したところ，命令発話については 65%[11]，それ以外の発話については，32%程度の認識率だった．この認識精度では，文脈形成や誤認識箇所の補正には十分とはいえず，提案手法をより効果的に機能させるためには，大語彙認識器の精度向上も一つの重要な課題となる．文脈形成，誤り補正の枠組みについては，先行研究 [6, 7] の考え方などを参考にし，大規模認識器の精度を向上させつつ，提案手法に組み込んでいく予定である．

6 おわりに

本論文では，施設内生活支援ロボットの音声認識部として，複数の音声認識器を併用することで，ロバストでかつ自由発話にも柔軟に対応可能な音声理解手法について提案した．語彙数は少ないが，想定される命令発話に対して，9 割強の極めて高い認識精度を実現した．さらに，提案手法は，音声認識器から得られる信頼度や出力結果数，2 つの認識器の出力結果の編集距離などを用いて，入力発話が命令発話かそうでないかの分別を行った．提案手法は，シンプルでかつ効果的であることを実験的に確認した．実験の結果，編集距離が分別のために最も有効であり，分別精度も極めて高いものとなった．これは，認識器の精度が高く，さらに認識器の選択も適切であることを意味し，

[10] 複雑になれば高い精度が保証されなくなる．一般に，複雑な文法を扱いたいのなら大語彙認識器に専用の言語モデルを適用する方が効果的だと考えられる．

[11] これは単語単位の認識率である．タスク依存認識器の場合の評価基準である発話単位での認識率で考えるとさらに精度は悪くなる．

大語彙認識器の認識率に問題は残るものの，提案手法は特定のタスクに関する発話にはロバストで，かつそれ以外の発話についても対応可能であることを示している．

今後の課題としては，(1) 対話コーパスの収集とそれに基づく言語モデルの適用，(2) 複雑な命令発話の認識，(3) 大語彙認識器の信頼性向上，(4) 実環境での音声認識実験，(5) カメラ画像など，他の入力モダリティとの協調によるマルチモーダル化と精度向上などが挙げられる．

謝辞

本研究は次世代ロボット知能化技術開発プロジェクト（独立行政法人新エネルギー・産業技術総合開発機構）における「施設内生活支援ロボット知能の研究開発」の成果の一部である．

参考文献

[1] Kazunori Komatani, Ryoji Hamabe, Tetsuya Ogata, and Hiroshi G. Okuno. Generating confirmation to distinguish phonologically confusing word pairs in spoken dialogue systems. In *Proceedings of 4th IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, pp. 40–45, 2005.

[2] Akinobu Lee, Kiyohiso Shikano, , and Tatsuya Kawahara. Real-time word confidence scoring using local posterior probabilities on tree trellis search. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2004)*, Vol. I, pp. 793–796, 2004.

[3] Jun Ogata and Masataka Goto. Speech repair: Quick error correction just by using selection operation for speech input interfaces. In *Proceedings of Interspeech 2005*, pp. 133–136, 2005.

[4] Kazutaka Shimada, Tsutomu Endo, and Sayaka Minewaki. Speech understanding based on keyword extraction and relations between words. *Computational Intelligence*, Vol. 23, No. 1, pp. 45–60, 2007.

[5] 磯部俊洋, 伊藤克亘, 武田一哉. 複数の認識器を選択的に用いる音声認識システムのためのスコア補正法. 電子情報通信学会論文誌 D, Vol. J90-D, No. 7, pp. 1773–1780, 2007.

[6] 伊藤友裕, 西崎博光, 関口芳廣, 中川聖一. 音声文書インデキシングのための web 文書を利用した自動誤り訂正. 第 3 回情報科学技術フォーラム講演論文集, 第 2 巻, pp. 343–344, 2004.

[7] 宇津呂武仁, 西崎博光, 小玉康広, 中川聖一. 複数の大語彙連続音声認識モデルの出力の共通部分を用いた高信頼度部分の推定. 電子情報通信学会論文誌, Vol. 7, pp. 974–987, 2003.

[8] 河原達也, 李晃伸. 連続音声認識ソフトウエア julius. 人工知能学会誌, Vol. 20, No. 1, pp. 41–49, 2005.

[9] 後藤真孝. 非言語情報を活用した音声インタフェース. 情報処理学会 音声言語情報処理研究会研究報告 2004-SLP-52-7, pp. 41–46, 2004.

[10] 佐古淳, 滝口哲也, 有木康雄. Adaboost を用いたシステムへの問い合わせと雑談の判別. 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション, pp. 19–24, 2006.

[11] 西崎博光, 中川聖一. 音声認識誤りと未知語に頑健な音声文書検索手法. 電子情報通信学会論文誌 D-II, Vol. J86-D2, No. 10, pp. 1369–1381, 2003.

[12] 宮崎昇, 中野幹生, 相川清明. 逐次発話理解法による対話音声理解. 電子情報通信学会論文誌 D-II, Vol. J87-D-II, No. 2, pp. 456–463, 2004.

[13] 元吉大介, 嶋田和孝, 榎田修一, 江島俊朗, 遠藤勉. 対話型ロボットのための口領域動画像に基づく発話推定. 人工知能学会第 22 回全国大会, 2008.

[14] 山形知行, 佐古淳, 滝口哲也, 有木康雄. SVM を用いたシステムへの問い合わせと雑談の判別. 日本音響学会 2007 年春季研究発表会, pp. 185–186, 2007.

[15] 山口辰彦, 酒向慎司, 山本博史, 菊井玄一郎. 信頼度尺度に基づく音声認識誤りの検出および誤り訂正. 電子情報通信学会技術研究報告 音声 SP2003-65, pp. 7–12, 2003.