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

2004－ＨＩ－109　(4)
2004－SLP－ 52　(4)
2004／7／16

音声対話による大規模知識ベース検索システム -音声版ダイアログナビ-

翠 輝久† 駒谷 和範† 清田 陽司‡ 河原 達也†‡ 木戸 冬子††

† 京都大学 情報学研究科 知能情報学専攻
‡ 科学技術振興機構 さきがけ研究 21
†† マイクロソフト株式会社

あらまし　本稿では，音声対話による大規模知識ベース検索システム「音声版ダイアログナビ」について紹介する．音声対話システムにおいては，話し言葉特有の冗長性や，音声認識誤りに対処する必要がある．音声版ダイアログナビでは，検索整合度と検索重要度という2つの尺度を導入し，検索に決定的な影響を与える箇所は検索を実行する前に確認し，結果として検索に影響を及ぼす箇所は検索結果の違いに基づいて効率的な確認を行う．30名の被験者による実験により，単純に音声認識結果を用いる場合より検索成功率が向上し，また音声認識の信頼度を用いる確認戦略よりも効率的に確認が行えることが示された．また，本システムは，2004年5月から京都大学学術情報メディアセンターにおいて，一般学生を対象として試験運用を開始している．

-Speech Dialog Navigator- Large Scale Document Retrieval System with Spoken Dialog

Teruhisa Misu† Kazunori Komatani† Yoji Kiyota‡
Tatsuya Kawahara†‡ Fuyuko Kido††

† School of Informatics, Kyoto University
‡ PRESTO, Japan Science and Technology Agency
†† Microsoft Co., Ltd.

Abstract　The paper describes Speech Dialog Navigator, which can retrieve from large scale document with spoken dialog. Adequate confirmation is indispensable in spoken dialog systems to eliminate misunderstandings caused by speech recognition errors. Spoken language also inherently includes redundant expressions such as disfluency and out-of-domain phrases, which do not contribute to task achievement. The system can cope with these problems by making confirmation prior and posterior to the retrieval. An experimental evaluation using 651 sentences by 30 users shows that the system generates confirmation more efficiently for better task achievement compared with the method using the conventional confidence measure of automatic speech recigntion. The system is running since May 2004 at Academic Center for Computing and Media Studies in Kyoto University.

1 はじめに

大語彙音声認識技術の高精度化に伴い，音声対話システムの研究対象は関係データベースの検索から，一般的な文書の検索へと広がりつつある [1]．音声対話システムにおいて，発話からユーザの意図を解釈する手続きは不可欠である．従来のバス案内タスクなどのスロットフィリング型のタスクでは，発話の中から検索に必要なキーワードを抽出することでユーザの意図を解釈し，それが同定できなければ確認するといった方法論を用いることができた．しかし，マニュアル [2] や Web ページなど，テキストで記述された大規模知識ベースを検索する際には，キーワードの集合を明確に定義することが不可能であり，音声認識結果全体を自然言語文として解釈する必要がある [3]．

しかし，音声で自然言語を入力する場合に，単純に音声認識エンジンの出力結果をそのまま用いて検索すればよいわけではない．この理由として，以下の 2 つが考えられる．

- **音声認識誤り**
  大語彙連続音声認識において，音声認識誤りは不可避である．従来のデータベース検索タスクでは，検索に必要なキーワード集合があらかじめ与えられているため，そのようなキーワードに関して音声認識の信頼度を計算することで頑健な解釈・対話を行うことができた [4]．しかし，テキストで記述された知識ベースを検索する場合にはキーワードの定義が明確でないため，このような手法を用いることは難しい．

- **話し言葉音声に含まれる冗長性**
  話し言葉音声にはフィラーや多様な文末表現など，冗長性が多い．これらの冗長な情報を検索に利用しても，検索に貢献しないばかりか，知識ベースとのマッチングを困難にする要因となる．

これらの問題に対処するためには，音声認識結果から検索に有用な部分を自動的に判別する枠組が必要になる．

本研究では，検索に用いる知識ベースのみから求められる統計量と，音声認識の N-best 候補に対する検索結果を用いて，音声認識結果の各文節が検索に有用かどうかを判定する．音声認識の言語モデルと検索文書の言語モデルを使い分けることにより，音声認識の頑健性を向上させながら検索に有用でない部分を検出する．さらに複数の候補を求めて検索結果を得ることで，検索結果に違いを与える部分を同定する．これにより，検索結果に影響を与えない確認を削減することができる．この確認戦略に基づいて，自然言語音声によりソフトウェアサポート用の知識ベースを検索するシステム「音声版ダイアログナビ」を構築した．

今回，システムの評価のために 30 人の被験者による実験評価を行った結果を報告する．さらに，音声版ダイアログナビの検索対象の知識ベースを拡張して，京都大学の教育用計算機端末の一般利用者を対象としたヘルプシステムを構築した．このシステムは，2004 年 5 月より学術情報メディアセンターにおいて試験運用を開始している．本稿では，これにより収集した発話データの分析結果についても報告する．

表 1: ソフトウェアサポート用知識ベース

知識ベースの種類	件数	文字数
用語集	4,707	約 70 万
ヘルプ集	11,306	約 600 万
サポート技術情報	23,323	約 2200 万
合計	39,336	約 4000 万

[HOWTO] Windows XP で音声認識を使用する方法

この資料は以下の製品について記述したものです．

- Microsoft Windows XP Professional
- Microsoft Windows XP Home Edition

概要 この資料では，Windows XP で音声認識を使用する方法について説明しています．Microsoft Office XP の音声認識をインストールしているか，または，Office XP がインストールされたコンピュータを新たに購入した場合は，すべての Office アプリケーションや，音声認識が利用可能なその他のアプリケーションで音声認識を使用できます．

詳細 音声認識は，音声をテキストに変換するオペレーティング システムの機能です．音声認識エンジンと呼ばれる内部ドライバによって，単語が認識され，テキストに変換されます．音声認識エンジンは，‥

図 1: ソフトウェアサポート用知識ベースの例

2 対象とする知識ベース

検索対象とする文書は，マイクロソフト社のソフトウェアサポート用知識ベースであり，この概要は表 1 の通りである．これらは自然言語によって記述されている．サポート技術情報の例を図 1 に示す．

この知識ベースに対して，ユーザのテキスト入力文により検索を行うシステムとして，ダイアログナビ [5] が東京大学で開発されている．ダイアログナビでは，自然言語入力文と知識ベースを柔軟にマッチングするために，係り受け関係や同義表現を考慮して解釈している．

このダイアログナビをバックエンドとしてとして使用し，音声入力によりこの知識ベースを検索するシステムとして，音声版ダイアログナビを実装した．

3 音声版ダイアログナビの確認戦略

音声認識結果を入力文として扱う場合に，音声認識誤りの可能性が高い部分全てを一つずつ確認するのは非効率的である．また，音声認識誤り箇所が常に検索に悪影響を与えるとは限らない．そこで単語ごとの音声認識誤りによる損失を考慮して，確認の方法を切り替える．

ダイアログナビは，Web の検索エンジンのように複数の検索結果を提示するため，発話の一部が正確に認識されなくても，検索結果に違いがないことがある．そこで，音声認識結果の N-best 候補を用いて検索を行い，候補間で検索結果に違いがないかを調べる．検索結果の違い (=検索重要度) が大きい場合には，検索結果に影響を与える原因となった N-best 候補間の相違箇所を提示してユーザに確認する．これが検索後の確認である．一方，検索に決定的な影響を与える語句 (本タスクではプロダクト名などがそれに該当する) が誤認識された場合，その後の検索が意味をなさない可能性が高い．そのため，これらの重要語句に対しては検索前に確認する．この確認を行う際の基準として，検索整合度を導入する．

以上の確認を組み込んだシステムの処理の流れは以下の通りである．

1. 認識結果に対して文節単位で検索整合度を計算する．
2. 検索整合度が低い文節中の重要語句をユーザに確認する．
3. 認識結果の N-best 候補それぞれについて検索を行う．
4. 検索重要度を計算し，それが高い場合にはユーザへの確認対話を生成する．
5. 最終結果をユーザに提示する．

これらの全体の処理の流れを図 2 に示す．また，以下の節でその詳細について述べる．

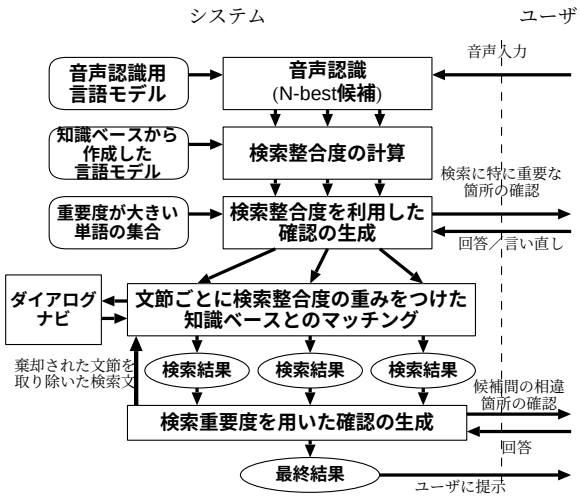

図 2: 音声版ダイアログナビの処理の概要

3.1 検索整合度を用いた確認と重み付きマッチング

検索整合度の計算には，音声認識の際に用いる言語モデルとは別に，検索対象である知識ベースのみから学習した単語 N-gram モデルによる単語パープレキシティを使用する．音声認識結果中の認識誤りである箇所は文脈的に不自然である場合が多く，また検索に直接関係がない語句は知識ベース内での出現確率が低いため，パープレキシティは高くなる．このように，音声認識時と異なる言語モデルによりパープレキシティを計算することで，認識結果中の誤り箇所や，正しく認識されたが検索には有用でない箇所を検出する．パープレキシティ PP を検索整合度 RS(relevance score) に変換するには以下のシグモイド関数を使用する．

$$RS = \frac{1}{1 + \exp(\alpha * (\log PP - \beta))}$$

部分的な認識誤りを棄却するために文節単位で検索整合度を計算する．実際の認識結果に対して検索整合度を計算した例を図 3 に示す．この例では,「不要になった」という検索に直接関係がない文節と，文末の誤認識した文節のパープレキシティ PP が高くなり，検索整合度 RS は低くなっている．

検索に決定的な影響を与える語句が誤認識された場合，検索が失敗する可能性が高い．そのため，こうした語句が不自然に出現している場合は，検索を実行する前にユーザに確認する．これらの語句は，知識ベースにおいて計算した *tf·idf* 値により規定する．まず，各文書で *tf·idf* 値が最も高い単語をその文書

ユーザ発話：「ＷＩＮＤＯＷＳ９８で不要になったＩＭＥ２０００を削除したいのですがどうしたらいいでしょうか？」

音声認識結果：「ＷＩＮＤＯＷＳ９８で不要になったＩＭＥ２０００を削除したいのですがどうしたいでしょか」

構文解析により文節単位に分割：「ＷＩＮＤＯＷＳ９８で／不要になった／ＩＭＥ２０００を／削除したいのですが／どうしたいでしょ／か」

検索整合度の計算：

文節 (コンテクスト)	PP	RS
(<S>)WINDOWS98 で (不要)	423.16	0.99
(で) 不要になった (ＩＭＥ)	22854.70	0.00
(なった)IME2000 を (削除)	349.24	0.99
(を) 削除したいのですが (どう)	10.56	1.00
(が) どうしたいでしょ(か)	2459.01	0.36
(でしょ) か (</S>)	3044.04	0.23

<S>，</S>はそれぞれ始端記号，終端記号

図 3: 検索整合度の計算例

の代表とする．その上で，全文書集合で代表となった回数の多い,「Word」や,「セットアップ」などの，35 単語を選択した．

検索整合度が閾値以下である文節にこれらの語句が含まれる場合には，誤認識の可能性が高いと考えられるので，ユーザに認識結果を提示し，確認を行う．ユーザは提示された文節が認識誤りであると判断した場合には，その文節を認識結果から取り除くか，その文節のみを言い直すかを選択できる．

知識ベースとマッチングを行う際には，音声認識誤りを含む文節や，検索に有用でない文節を除外することが望ましい．そこで，マッチングを行う際に，各文節の検索整合度 RS をその文節に対する重みとして用いる．これにより，認識誤りや無関係な部分による検索への悪影響を抑制する．

3.2 検索重要度による確認

ユーザ発話の認識結果の第 1 候補が誤りであっても，N-best 候補の中に正解が含まれる可能性がある．しかし，検索に影響が少ない単語の置換も多いため，これら全てを確認するのは非効率的である．そこで，音声認識結果の N-best 候補それぞれに対する検索結果を用いて検索重要度を求める．

まず，音声認識結果の N-best 候補間の相違箇所を同定する．次に，この N-best 候補それぞれについて実際に検索を行い，検索結果の相違の大きさを検索重要度 SS(significance score) として定義する．検索重要度 SS は，第 n 候補に対する検索結果を $res(n)$，

図 4: システムの生成した確認の例

その数を $|res(n)|$ として，以下のように定義する．

$$SS = 1 - \frac{|res(n) \cap res(m)|^2}{|res(n)||res(m)|}$$

検索重要度が閾値を越えている場合には，その相違部分をユーザに提示し確認する．逆に，検索重要度が閾値以下の場合には確認を行わず，第 1 候補による検索結果をそのまま表示する．なお，今回音声認識の結果として出力する候補数 N は 3 とした．提示された候補の中からユーザが適切なものを選択すると，対応する検索結果が表示される．

4 実装と評価実験

ユーザの音声による質問によりソフトウェアサポート用知識ベースを検索するシステム「音声版ダイアログナビ」をマイクロソフト社の Web ブラウザ Internet Explorer 6.0 上で動作するシステムとして作成した．音声認識は，我々の研究室で開発された Julius for SAPI[1] [6] によりクライアント PC 上で行う．また，ユーザに対する確認は図 4 のように画面に出力し，ユーザは番号を音声により読み上げるか，選択肢をクリックすることにより回答する．

4.1 評価用データの収集

評価用データは，音声対話システムを利用したことのない 30 名の被験者により収集した．各人に，設定した想定場面に基づいて 11 課題，これとは別に自由に 3 課題について検索を行ってもらった．ただし，質問の回答としてふさわしい検索結果が得られなかった場合には，被験者の判断で各課題につき 3 度までの言い直しを許した．

[1] http://julius.sourceforge.jp/sapi/

表 2: 検索成功率 (被験者実験)

発話数	書き起こし入力	認識結果入力	提案手法
651	520 (79.9%)	421 (64.7%)	457 (70.2%)

その結果，合計 420 課題，651 発話のデータを得た．全発話に対する音声認識の単語認識精度は平均で 76.8%である．

4.2 検索成功率による評価

まず，提案手法の評価尺度として，全 651 発話に対する検索成功率を調べた．ここでは，システムが最終的に提示した候補の中に，最初の質問に対する正しい回答が含まれていた場合を検索成功としている．収集した音声データに対して，以下の 3 つの条件で検索実験を行った．

1. ユーザ発話の正確な書き起こし (人手で作成) を用いて検索した場合 **［書き起こし入力］**
2. 音声認識結果の第 1 候補を用いて検索した場合 **［認識結果入力］**
3. 検索整合度と検索重要度の両方を用いて確認及び検索を行い，生成する確認に対してユーザが適切に応答する場合 **［提案手法］**

これらの条件での検索成功率を表 2 に示す．提案手法により検索を行った場合は音声認識結果の第 1 候補をそのまま用いて検索を行った場合よりも検索成功率が 5.5%向上している．

4.3 確認の効率性の評価

もう一つの評価尺度として，生成した確認の回数に関して評価を行った．提案手法により生成された確認回数は 221 回である．これは，おおよそ 3 発話に 1 回強，確認が行われたことになる．このうち，検索整合度を用いた事前確認の回数は 66 回あり，検索重要度を用いた事後確認が 155 回であった．

比較対象として，音声認識結果の N-best 候補から計算される信頼度 [4] を用いて確認を行う場合との確認回数，検索成功率を比較した．確認を行うための信頼度の閾値 θ_1 として，0.4，0.6，0.8 の 3 通りを用いた．信頼度が閾値以下の自立語を確認するものとし，それが誤認識されたものであった場合には，その単語を含む文節を棄却して検索した．

この結果を表 3 に示す．提案手法は，従来手法 ($\theta_1 = 0.8$) に比べて確認回数を半分以下に抑えながら，より高い検索成功率を得ている．従来手法の信頼度 [4] は，音声認識の音響的・言語的尤度のみを反映したものであるのに対して，本手法での確認は検索に関する有用性がより直接的に反映されている．

表 3: 音声認識の信頼度を用いた確認戦略との確認回数・検索成功率の比較

	確認回数	検索成功数 (成功率)
提案手法	221	457 (70.2%)
信頼度 ($\theta_1 = 0.4$)	77	427 (65.6%)
信頼度 ($\theta_1 = 0.6$)	254	435 (66.8%)
信頼度 ($\theta_1 = 0.8$)	484	445 (68.4%)

5 ユーザインターフェースの改良

被験者実験を行った際に，被験者から以下のような問題点を指摘された．

- 発話を開始してよいタイミングがわかりづらい．
- 質問をしてから，システムが検索結果を提示するまでに時間かかる．

まず，音声入力のタイミングの問題について検討を行った．本システムの最初の実装では，発話が可能なタイミングを明示的に提示していなかった．また，システムがユーザの質問に対して確認画面や検索結果を提示している時でも，それに関わりなく新たな質問を入力することができた．被験者からの指摘を受けて，発話が可能なタイミングをわかりやすくするために，合成音声によるプロンプトを導入し，発話入力可能であることを表すアイコンを画面に提示することにした．さらにシステムが出力する確認画面には,「質問をやり直す」という選択肢を，また検索結果を提示する際には,「次の質問を行う」という選択肢をそれぞれ新たに追加した．このような改良により，発話を開始してよいタイミングがわからないという指摘は減った．

次に，検索時間の問題について検討を行った．本システムは，検索対象の知識ベースが約 4 万件と膨大であるため，発話が入力されてからシステムが最終結果を提示するまでに時間がかかる．最初は，システムが検索している時に，ユーザに対して特に何も情報を提示しなかったため，システムがフリーズしたと感じるユーザが多いことがわかった．そこで，音声入力の検出，発話終了の検出，検索開始，出力画面生成のそれぞれの段階で，画面上に対話の進行状況を提示することにした．

図 5: 試験運用中のシステムの利用風景

6 京都大学学術情報メディアセンターでの試験運用

本システムは，京都大学学術情報メディアセンターのオープンスペースラボにおいて，教育用計算機を利用する一般学生を対象として試験運用を開始している (図 5)．運用に先立って，センターの TA にヒアリングを行った結果，メールシステム等の教育用計算機システムに関する質問が多いことがわかった．そのため，メディアセンター固有の質問に対応できるように，一般利用者向けに用意されている「よくある質問とその答え」(FAQ)77 件を知識ベースに追加した．本システムは，2004 年 5 月 18 日より 6 月 12 日までの 19 日間の運用で，合計 81 回の利用があった．これらの発話データの内容と発話スタイルに関する分析を行った．

まず，ユーザの発話内容を調べた．その結果,「明日の天気が知りたい」といった，ドメイン外の発話が多いことがわかった (30%)．このようなドメイン外発話は，システムの運用を開始した当初は全発話の約 5 割を占めていた．そこで，システムの利用方法の中に，パソコンに関する質問以外は音声認識されないことを明記したところ，ドメイン外発話は大幅に減少した．残りのドメイン内発話のうち，システムが適切な候補を提示できたのは，51%であった．このうち,「アカウント名を忘れた」のようなメディアセンターの教育用計算機システムに関する質問が 38%あった．また，全般的に「Word の起動方法」といった，アプリケーションの基本的な使用方法・操作方法をたずねる質問が多かった．

次に発話のスタイルを分析した．まず，利用者が大語彙連続音声認識を利用したシステムを利用したことがないためか,「印刷」といった単語だけの発話が多かった．逆に文で質問している発話でも,「えーと」のようなフィラーが含まれることはほとんどなかった．さらに，周囲の利用者を意識するためか，ささやき声による発話が多かった．

7 むすび

音声対話によりソフトウェアサポート用知識ベースを検索するシステム「音声版ダイアログナビ」を実装した．

音声認識結果に含まれる音声認識誤り・話し言葉音声に含まれる冗長性の問題に対して，検索整合度と検索重要度の 2 つの尺度を用いて，検索の前後で確認を行う戦略を考案し，被験者実験によりその有効性を確認した．

さらに，知識ベースに京都大学学術情報メディアセンターの一般ユーザ向けの項目を追加し，オープンスペースで試験運用を開始した．本システムは以下の URL からダウンロードすることもできる．
http://www.ar.media.kyoto-u.ac.jp/msnavi/speech/

謝辞

本研究に対し，多大な協力を頂いた東京大学の黒橋禎夫助教授に深く感謝します．

参考文献

[1] C. Hori, T. Hori, H. Isozaki, E. Maeda, S. Katagiri, and S. Furui. Deriving disambiguous queries in a spoken interactive ODQA system. In *Proc. ICASSP*, Vol. 1, pp. 624–627, 2003.

[2] 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.

[3] 駒谷和範, 河原達也, 清田陽司, 黒橋禎夫, Pascale Fung. 柔軟な言語モデルとマッチングを用いた音声によるレストラン検索システム. 情報処理学会研究報告, 2001-SLP-39-30, 2001.

[4] 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, 2002.

[5] 清田陽司, 黒橋禎夫, 木戸冬子. 大規模テキスト知識ベースに基づく自動質問応答 ーダイアログナビー. 自然言語処理, Vol. 10, No. 4, pp. 145–175, 2003.

[6] 住吉貴志, 李晃伸, 河原達也. 音声認識エンジン Julius/Julian の API 実装. 情報処理学会研究報告, 2001-SLP-37-16, 2001.