特集 音声情報処理技術の最先端

話し言葉による音声対話システム

河原 達也
京都大学 学術情報メディアセンター
kawahara@i.kyoto-u.ac.jp

音声を認識するだけでなく，発話の意図を理解・推論して，適切な応答をするのが音声対話システムである．音声対話システムの研究は，音声認識技術の進展[1]に伴って活発に進められてきた．1990年代には，多くの研究機関でプロトタイプシステムの開発が行われ，その一部は自動電話応答（IVR）システムとして実用化された．ただし，これらは限定されたタスクドメインを前提とし，かなりの人手による作業を要するものであった．近年になって，Web等を対象とした汎用的な情報検索や質問応答と一体化した対話システムの設計・構築が行われるようになった．本稿では，こうした音声対話システムの構成論について概観し，著者らが開発・一般公開している音声対話システムの事例を紹介する．

■音声対話システムの構成

音声対話システムの典型的な構成を図-1に示す．ユーザの発話に対して音声認識を行い，その結果を理解・解釈モジュールにより処理し，バックエンドシステムで情報検索を行いながら，ユーザに対する応答や質問を生成する．その際に，対話の履歴や状況を保持し，解釈や応答生成に利用する．

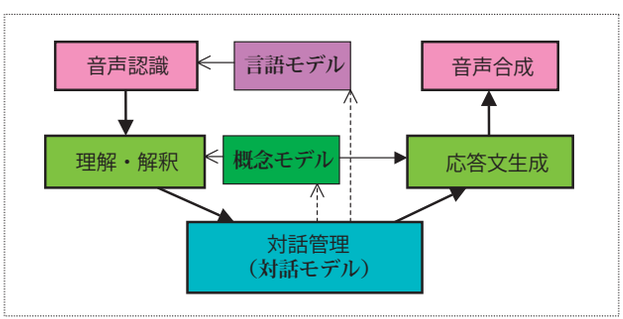

図-1　音声対話システムの構成

テキスト入力の場合と比較して，音声入力を扱うことによる困難な点は，音声認識の誤りが不可避なことに加えて，ユーザの言葉がより自発的で非定型になることである．すなわち，より口語的な表現が使用され，省略や言い淀みなどの現象が頻繁に生じるので，これらへの対処・頑健性が重要である．これらの問題とアプローチについては，文献2）で詳細に述べられている．

■柔軟な音声理解・対話への展開

まず，音声認識の際に用いる言語モデルと，理解・解釈の際に用いる概念モデルについて考察する．古典的な方法論としては，限定されたタスクドメイン（たとえば列車の切符予約など）を仮定し，それに応じて音声認識に用いる有限状態オートマトン（FSA）などの形式文法と意味理解に用いる規則・文法を人手で記述するのが一般的であり，少し前まで（あるいは今でも）ほとんどの音声対話システムで採用されていた．このような方法論は，（デモシステムとしてはOKでも），かなり拘束の強

	発話パターンが形式言語？（FSA）	検索パターンがSQLタイプ？（RDB）	必須スロットのみで検索可能？（システム主導可）
バス案内システム	○	○	○
ホテル検索システム	△	○	×
レストラン検索システム	×	×	×

表-1　各タスクの分類

	音声認識	理解・解釈	対話戦略
バス案内システム	有限状態文法	キーワードからSQLスロットへの変換	必須項目を質問
ホテル検索システム	文法＋*N*-gram（スポッティング）		不確実な項目を確認
レストラン検索システム	単語*N*-gram	文のマッチング	候補を絞込み

表-2　各システムで採用された方法

いタスクにおいて，人手で多くのチューニングを行って初めて実用に供することができるものである．もちろんこれに対して，より頑健な言語理解方式についても盛んに研究されてきた[2]．

音声対話システムの典型的な事例として，著者の研究室においてこれまで開発してきた主なものを以下に示す．

(1) バス案内システム[3]

「京大正門前から京都駅まで」のような発話に対して，該当するバスを検索し，到着までどの程度かかるか案内する．音声認識文法はFSAで記述している．出発地（乗車停留所）と目的地（降車停留所）が決まれば検索できるので，これらを質問・同定するのが理解・対話の基本戦略である．

(2) ホテル検索システム[4]

「京都市内で温泉のある宿を教えて」のような発話に対して，該当するホテルや旅館を検索し，一覧を表示する．値段や付帯施設などの条件の指定は任意で追加・修正でき，そのたびに検索を行う．条件項目の同定が理解・対話の基本戦略であるが，バス案内タスクと異なり，任意項目が多いので，一方的に質問することはできず，音声認識の確信度が低い場合に確認を行うのみである．また，不適格な発話が多いので，文法は記述するものの，キーフレーズを抽出する方式を採用する．

(3) レストラン検索システム[5]

「新宿にあるおしゃれな焼肉屋を教えて」などのような発話に対して，該当するレストランを検索し，一覧を表示する．ホテル検索タスクと異なり，条件の指定自体に曖昧な表現が多いので，文法を記述することが事実上不可能である．音声認識には当該ドメインに適応した単語*N*-gramモデルを用い，音声認識結果とレストランごとに用意された多数の想定質問文をマッチングすることで検索を行う．

以上をまとめると，(a) 発話のパターンがFSAなどの形式文法でカバーできるか，(b) 検索のパターンがSQLなどの関係データベース（RDB）を対象としたコマンドに変換できるか，(c) 完全にシステム主導の対話が可能か，の観点によって表-1のように分類することができる．この分類に従って，音声対話システムの各モジュールで適当と考えられるアプローチをまとめると表-2のようになる．

■データベース検索から文書検索へ

これまで研究・開発されてきた大半の音声対話システムが，天気案内やフライト検索などのように関係データベース（RDB）に対する検索として定式化できるものであった．これに対して，より汎用的な情報検索を指向した研究が行われつつある．近年情報検索といえば，Webでの検索を連想される方が多いと思われるが，これはWebページのような文書を検索する問題として定式化できる．具体的な例として，マニュアルや新聞記事の検索が考えられる．前記のレストラン検索もレストランのWebページの検索とすることもできる．

この場合，検索発話の理解・解釈は，音声認識結果の文と文書とのマッチングのためのものとなり，出現単語を要素とするベクトル空間モデルが採用されるのが一般的である．たとえば2万単語の語彙を想定すると，各単語を要素とする2万次元のベクトルを構成し，各単語の出現回数が代入される．各文書に対しても同様のベクトルを求めて，ベクトル間の距離（コサイン距離など）を求めることでマッチングを行う．ここで，機能語（stop word）などの検索にあまり寄与しない単語を除去したり，ベクトル自体をSVD（Singular Vector Decomposition）などで圧縮することもある．また，単語の2-gramカウントなどの連鎖情報や，構文解析・係り受け解析の情報を

	情報の構造	音声認識	理解・解釈	対話戦略
データベース検索	RDB	タスク限定の語彙・文法	SQLスロットへの変換	スロットの質問・確認
文書検索	自然言語テキスト	*N*-gramによる大語彙連続音声認識	ベクトル空間モデル	質問の明確化・検索結果の絞込み

表-3　データベース検索と文書検索の音声対話システムの比較

利用することもある．

音声認識においては，発話パターンが非常に多様になるため，ディクテーションと同様の大語彙連続音声認識を実行するが，ある程度ドメインが定まっている場合はドメイン適応することが望ましい．特に専門用語に対応することは必要不可欠である．対話戦略においては，質問が曖昧な場合に聞き返したり，検索結果が多すぎる場合に絞り込みを行うことが要求されるが，このような方法については今後一層の研究が必要である．たとえば著者らは，家電製品の操作マニュアルのように，文書が目次で階層構造化されている場合に，その構造を利用して効率的に絞込みを行う方法を提案している[6]．

以上をまとめて，データベース検索と文書検索を比較したのが**表-3**である．

なお，この文書検索の特殊な例として，コールルーティングタスクがある．これは，代表番号に寄せられた問合せを適当な担当・部門に割り振るものであり，たとえば航空会社であれば，「来月北海道に行くのですが」という問合せの顧客を「国内線予約」担当につなぐものである．やはり，ベクトル空間モデルを用いて各担当のモデルとマッチングを行うことにより実現される．

文書検索の発展として，質問応答（Q-A）タスクがある．これは，「中国の外相は？」「2010年のサッカーのワールドカップの開催地は？」などといった質問に対する直接的な答えを文書検索の結果から見つけるものである．このシステムの例として，NTTのSAIQAの音声対話版のSPIQA[7]がある．また，NTCIRワークショップなど情報検索のコンテストにおいても，音声による新聞記事検索や質問応答の試みも模索されている．

■京都市バス運行情報案内システム

古典的なデータベース検索の方法論で実装された例であるが，現在稼動中のシステムとして，京都市バス運行情報案内システム「音声ポケロケ」[3]を紹介する．これは，京都大学と，オムロン（株），京都高度技術研究所の共同研究によるものである．電話による音声対話システムで，電話番号は075-326-3116である．京都市交通局のWebサイトからも紹介されており，試験運用のかたちで一般市民に公開している．

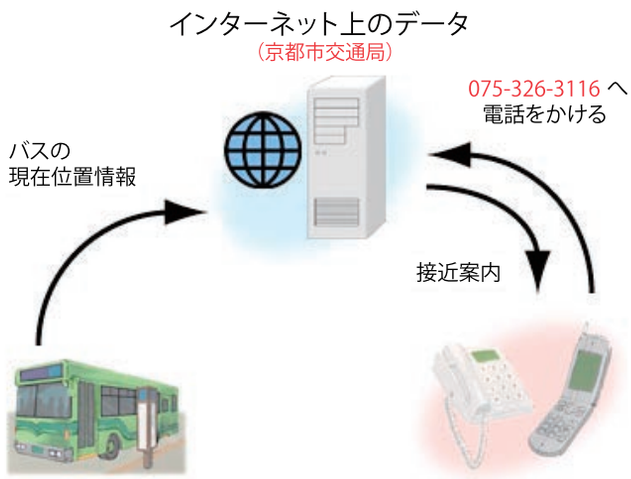

図-2　京都市バス運行情報案内システム「音声ポケロケ」の概念

このシステムは，市バスのリアルタイムの位置情報を管理しているデータベースにアクセスし，接近情報を案内するものである．この概念を**図-2**に示す．なお，同様の情報はインターネット（携帯電話のも含む）でもアクセスできるが，音声で照会した方が手っ取り早く情報を取得できる．対話例を以下に示す．

S:「こちらは音声ポケロケ実験サービスです．ご利用になる停留所または系統番号をおっしゃって下さい．」
U:「京大正門前から京都駅まで．」
S:「京大正門前から京都駅までによろしいですか．」
U:「はい．」
S:「206系統東山通京都駅行きのバスは2つ前の飛鳥井町を出発しました．…」

2004年1月から5月までの間に計1,372件の利用があり，このうち87.3%に対して案内を行うことができた．

バスの接近情報は一刻を争う場合もあるので，音声による情報アクセスは有効と考えられる．特に急いでいるユーザは，システムのプロンプトの途中で割り込んで発話することが多く，またこのような性急度の高いユーザに対しては確認を省略するなど，できるだけ迅速に対応するのが望ましい．そこでユーザモデルとして，システムに対する習熟度，タスクドメインに関する知識レベ

ル，性急度の3つを設定し，発話内容に加えて割り込みや無音時間などの音声の特徴からユーザを自動判別し，その結果に応じて対話戦略や応答内容を切り替える方法を提案した．その結果，使い慣れたユーザに対する対話を冗長にすることなく，初めて使うユーザに対しては適切な誘導を行うことで，平均対話時間（85.4秒→51.9秒）と平均ターン数（8.23→4.03）を大幅に減らし，主観的な満足度も高める効果を確認している[3]（ただし公開システムには未実装）．

■ソフトウェアサポートマニュアルの検索システム

文書検索タスクの例として，ソフトウェアサポートマニュアルの検索を行うシステム「音声版ダイアログナビ」[8]を紹介する．これは，京都大学と，東京大学（黒橋研究室），マイクロソフト（株）の共同研究によるものである．知識ベースとして，マイクロソフト社のソフトウェアサポート技術情報など約4万件のテキストを用いている．

このシステムの処理の概要を図-3に示す．検索発話に対して，このドメインを指向して作成された*N*-gram言語モデルを用いて音声認識を行う．次に，形態素解析・構文解析を行い，文節単位に区切る．文節ごとに，検索対象の文書集合から学習した言語モデルを用いて文書集合との整合度を計算し，検索の際の重み付けとして用いる．その際に，検索に決定的に重要なキーワード（製品名など）を含む文節の整合度が低い場合は，認識誤りの可能性が高く，その場合は検索が意味をなさないので，検索前にユーザに確認を行う．検索は，文節間の係り受け関係も考慮したマッチングにより行う．さらに，音声認識結果の第3候補までそれぞれについて検索を実行し，得られた結果を比較して大きな違いがあれば，認識結果が食い違う区間についてユーザに確認を行う．この様子を図-4に示す．このような方法により，検索結果に影響を及ぼす部分のみを効率的に確認する戦略を実現している．被験者実験により，音声認識結果をそのまま検索に用いる場合に比べて検索成功率を大幅（64.7%→70.2%）に改善し，音声認識の信頼度を用いた確認手法に比べても確認回数を約半分（1回の検索において0.74回→0.34回）に減らす効果を確認している[8]．

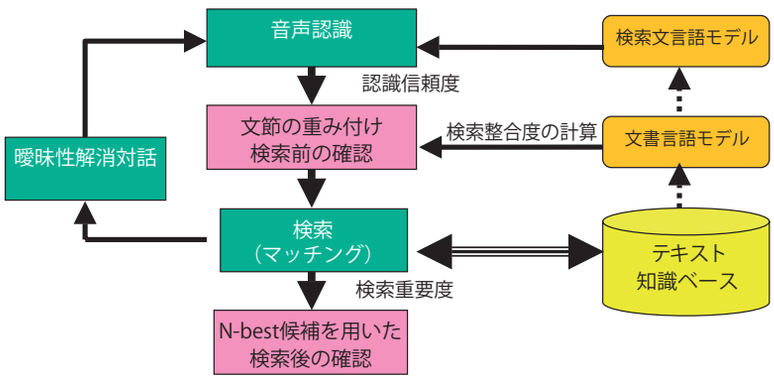

図-3　サポートマニュアル検索システム「音声版ダイアログナビ」の処理の概要

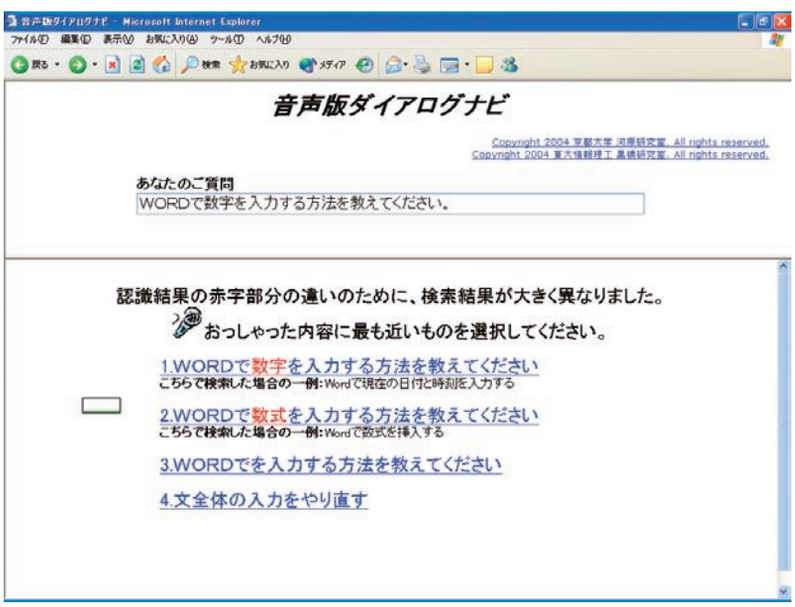

図-4　サポートマニュアル検索システム「音声版ダイアログナビ」の動作画面

本システムはhttp://www.ar.media.kyoto-u.ac.jp/msnavi/からダウンロードできるほか，京都大学学術情報メディアセンター南館のオープンスペースラボラトリ（OSL）に設置し，学生が自由に使えるように試験運用している．

■音声対話システムが使われるために

このように著者らのみならず，多くの大学・企業等でさまざまな音声対話システムの開発が行われてきた．しかしながら，これらの多くは残念ながらデモシステムの域を出ないものであった．前述のバス案内システムのように一般公開・試験運用にこぎつけた事例もあるが，大半がきわめて簡単なタスクのものであり，本格的な音声対話システムとは言いがたい．

システムを実際にエンドユーザに使ってもらうと，音声認識が発話のバリエーションに対応できない，バック

タスク外の発話	41.1%
検索項目・内容がデータベースに存在しない （「料理がおいしい」など）	(29.6%)
曖昧な表現を含んでいる （「料金が安い」,「琵琶湖周辺」など）	(8.0%)
検索要求以外 （「○○旅館に決めました」,「もしもし」など）	(3.4%)
文法が対応していない発話	2.2%
語彙が対応していない発話	18.5%
句末・文末表現のバリエーション	(10.5%)
固有名詞の省略形・別名などのバリエーション	(9.1%)

（発話数：910, 内訳部分は該当するものを各々計数しているため合計は一致しない）

表 -4　ホテル検索プロトタイプシステムにおける想定外発話の分類

エンドの検索が要求に応えられない，GUI（携帯端末を含む）と比べて音声入力を使うメリットが感じられない，といった問題に遭遇する．実際に前述のホテル検索システムをプロトタイプの段階で被験者（28名）に試してもらったところ，**表 -4** に示すように半数以上の発話に対応できなかった．このうち，語彙や文法の問題については大語彙の統計的言語モデルを用いることである程度対処できるが，最も大きな割合を占めている検索不可能な発話については，適切なガイダンスやフィードバックが重要であると考えられる．実際に，検索可能項目や発話パターンをGUIで提示すると，タスク外発話を11.0%まで減らすことができた[4]．ただし，GUIを用いずに音声のみでガイダンスを行うことは容易でなく，逆にGUIがあれば音声入力を用いる優位性が問われるというジレンマに陥る恐れがある．特に日本では，インターネットにアクセスできる携帯電話が普及しており，たいていの情報検索がそれでできるという状況が存在する．

音声対話システムが本格的に利用されるためには，本稿で解説したような自然言語音声を柔軟に理解する枠組みの高精度化が鍵になると考えられるが，それ以外に，「話しかけやすい」対話システムにするためのポイントを以下に挙げる．

(1) 会話エージェントやロボットの インタフェース

明確な（できれば個性を持った）相手が存在し，音声で話しかけられれば，自然と音声でやりとりができると期待される．そのためには，合成音声の自然性・明瞭性を含めたインタフェース全体の検討とともに，以下の問題も解決する必要がある．

(2) 完全に自由な話し言葉への対応

システムからのプロンプトに対して応答するというインタフェースは，ユーザに明確な目的がある場合はよいが，そうでない場合は「何をどう発話してよいか」戸惑う現象を引き起こす．任意のタイミングで，文としてまとまりのない発話ができればよいが，そのためには音声認識・理解や対話管理を現在のものから見直す必要がある．また，大語彙の話し言葉音声認識や，言い直しなどの不適格な事象への対応などの課題がある．

(3) ハンズフリー入力・遠隔発話への対応

会話エージェントやロボットに対して，接話マイクを用いて発話するのは不自然であり，遠隔発話に対応できることが必要である．また，自動車内や家の中などプライベートな空間では機械に話しかけることにそれほど抵抗がないと考えられるが，そのような場所でも接話マイクを用いないハンズフリー音声認識が必要となる．

自然な話し言葉音声による対話システムは，人工知能の究極的なテーマの1つと考えられ，本来深い理解を伴うものであるが，現在の音声認識や本稿で述べた枠組みはかなり表層的な処理が中心であり，一問一答形式であればまだよいが，人間と対話を進めていくには心もとない．確率統計的なモデルにより人間を満足させる（だませる?!）対話がどこまで実現できるかは，さらなる挑戦といえよう．

参考文献

1）河原達也：ここまできた音声認識技術，情報処理，Vol.41, No.4, pp.436-439 (Apr. 2000).
2）河原達也，松本裕治：音声言語処理における頑健性，情報処理，Vol.36, No.11, pp.1027-1032 (Nov. 1995).
3）駒谷和範，上野晋一，河原達也，奥乃　博：ユーザモデルを導入したバス運行情報案内システムの実験的評価，情報処理学会研究報告，SLP-47-12 (2003).
4）駒谷和範，鹿島博晶，田中克明，河原達也：複合的言語制約に基づくキーフレーズ検出を用いた汎用的なデータベース検索音声対話プラットフォーム，情報処理学会論文誌，Vol.44, No.5, pp.1333-1342 (May 2003).
5）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，SLP-39-30 (2001).
6）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 (July 2002).
7）堀　智織，堀　貴明，磯崎秀樹，前田英作，古井貞熙：音声インタラクティブODQAの構築とその評価，日本音響学会春季講演論文集，2-4-7, pp.71-72 (2003).
8）翠　輝久，駒谷和範，清田陽司，河原達也，木戸冬子：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－，情報処理学会研究報告，SLP-52-4 (2004). http://www.ar.media.kyoto-u.ac.jp/msnavi/

（平成16年7月13日受付）