特　集　「音声言語インタフェースの実用化と音声言語対話への展開」

音声対話システムの言語・対話処理

Language and Dialogue Processing in Spoken Dialogue Systems

中野　幹生　日本電信電話（株）NTTコミュニケーション科学基礎研究所
Mikio Nakano　NTT Communication Science Laboratories, NTT Corporation.
nakano@atom.brl.ntt.co.jp, http://www.brl.ntt.co.jp/people/nakano/

堂坂　浩二　（同　上）
Kohji Dohsaka　dohsaka@atom.brl.ntt.co.jp, http://www.brl.ntt.co.jp/people/dohsaka/

Keywords: spoken dialogue system, dialogue management, response generation, language processing.

1. はじめに

人間と音声言語によってコミュニケーションを行うシステム，すなわち，音声対話システムは，音声言語処理のさまざまな技術を統合して構築される．一般的に，音声対話システムは，音声認識部，音声生成部，言語・対話処理部の三つのモジュールからなる．言語・対話処理部はユーザの音声認識結果を受け取って，どのようなシステム発話を行うべきかを決定する．本稿では，現状の音声対話システムにおける言語・対話処理技術を概説する．音声対話システムと単なる音声応答システムの違いは，文脈情報，すなわち，対話の履歴を使うかどうかにある．

音声応答システムは，ユーザ発話の音声認識結果を，データベースクエリに変換して，その結果を言語表現に変換した後，音声合成してユーザに伝える．これに対して，音声対話システムは，文脈を考慮して，ユーザ発話を理解し，システム発話を生成する．図1は架空の対話例であるが，現状の最高レベルのシステムが行えるのと同等の複雑さをもった対話である [Seneff to appear]．例えばU2の「京都です」が降車駅を意味していることは，文脈を考慮しなければ理解できない．また，S4で「京都までですね」とシステムが発話するのも，直前のユーザ発話から得られる情報だけでは不可能である．

対話システムの研究は多分野にまたがるが，特に音声処理と人工知能・自然言語処理の二つの分野で活発である．音声処理の分野では，音声処理のアプリケーションとしての対話システムの研究が行われており，音声認識誤りにもかかわらずタスクを達成することや，音声認識の負担を減らすためユーザの発話を制限するように対話を制御する技術が開発されてきた．人工知能・自然言語処理の分野では，主に人間どうしの対話の分析に基づいて，人間の対話モデルの研究が行われてきた（以下AI系対話モデルの研究と呼ぶ）．言語学や心理学との関係も深い．しかしながら，近年の音声認識技術の向上により，人工知能・自然言語処理研究者も，実動する音声対話システムの研究を行うようになってきた．

S1　どのようなご用件でしょうか？
U1　京都まで予約したいのですが．
S2　東京までですね？
U2　京都です．
S3　どちらから京都までですか？
U3　新横浜からです．
S4　新横浜から京都までですね．いつの列車ですか?
U4　水曜日の9時すぎに出るのぞみの指定席が欲しいんですけど．
S5　水曜日の9時すぎののぞみですね．何名様ですか？
U5　1人です．
S6　お1人様ですね．
U6　はい．
S7　9時12分発，12時12分着ののぞみ3号があります．
U7　もう一つ早いのは?
S8　8時30分新横浜発ののぞみ3号がありますが，普通席は満席です．
U8　では9時12分発で良いです．
S9　禁煙席でよろしいですか？
U9　はい．
S10　乗車券はいかがいたしましょう．
U10　特急券だけで結構です．
S11　かしこまりました．確認させていただきます....

図1　新幹線指定券予約の対話例
Sはシステム発話，Uはユーザ発話

現在，一般ユーザと音声対話システムとの対話を収録し，音声対話システムの要素技術を評価することが盛んである．例えば米国では，2001年までDARPA主催のCommunicatorと呼ばれるプロジェクトがあり，飛行機での旅行プランニングアシスタントタスクを共通タスクとし，約10の研究機関でシステムの開発とデータの収録が行われた [Doran 01, Rudnicky 00, Seneff to appear]．また欧州でも，ARISEという鉄道の時刻表案内のプロジェクトがある [Lamel 00, Sturm 99]．MITのCommunicatorシステムは，一般ユーザと対話して，数十ターンの対話を行い，約7割の確率でタスクが遂行できる [Seneff to appear].

本稿では，これらの音声対話システムで用いられている技術に焦点を当て，最新の技術を紹介するが，必要に応じてAI系対話モデル研究にも触れる．

2. 対話システムの言語処理の概略

まず，音声対話システムの言語処理部の概略を説明する．一般的に，音声対話システムは，図2のような構成になっている．音声処理，言語処理という区別とは別に，音声対話システムは，発話理解と発話生成の二つの機能からなるとみることができる．この二つのモジュールは対話状態（dialogue state）と呼ばれるデータを介してつながっている．対話状態には，その時点までの対話の履歴と，ユーザの意図の推定結果やシステムの発話プランなどの情報が書き込まれる．発話理解は，ユーザの発話と現在の対話状態を入力として，新しい対話の状態を出力する関数，発話生成は，現在の対話状態を入力とし，システム発話と新しい対話状態の二つを出力する関数とみなすことができる．

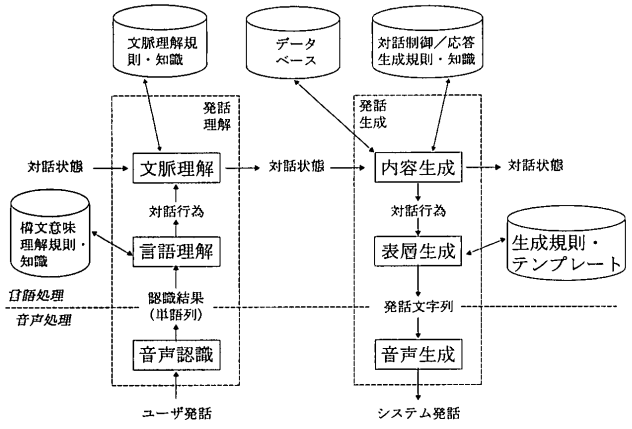

図2 音声対話システムの構成

一般に，対話システムの音声認識部は，システム発話の終了後，ユーザが話し，一定の長さのポーズを置くまでの間（音声区間）の音声を認識し，その結果の単語列（またはNbest単語列）を言語処理部に送る．これをもとに言語処理部で対話状態を変更する．

発話理解の言語処理は，言語理解（文理解）と文脈理解（談話理解）に分けて考えられる．言語理解は，音声認識の結果の単語列を受け取り，対話行為と呼ばれる表現を出力とする．対話行為は，発話の意図を，対話状態（すなわち文脈）と独立な形で表現したものである．例えば，図1U4の「水曜日の9時すぎに出るのぞみの指定席が欲しいんですけど」という発話の対話行為は，

[対話行為タイプ：予約の要求
列車種別：のぞみ
日にち：水曜日
出発時間帯：[時間：9時
時間付加情報：すぎ]]

と表される．同じ発話の意図を，システム開発者によっては，

[対話行為タイプ：要求
要求内容：指定席予約
列車種別：のぞみ
日にち：水曜日
時間帯：[時間：9時
時間付加情報：すぎ]]

と記述する場合もあるだろう．このように，対話行為の表現の仕方は後段の処理の都合の良いように決められる．

この対話行為を，これを文脈の中で解釈することにより，ユーザの意図を理解する．これが文脈理解で，ユーザ発話の対話行為と，ユーザ発話が行われる前の対話状態を入力とし，新しい対話状態を出力する処理である．発話生成は内容生成と表層生成に分けて考えられる．内容生成は，どのような内容をシステムがユーザに伝えるべきか，を決定し，それを対話行為の形で表現する．表層生成は，対話行為を，言語表現に変換する．

内容生成は，対話状態を入力とし，対話行為表現を出力する．まずユーザの要求を聞き，不確定な部分を確認したり不足している情報をユーザに聞いた後，データベース検索をし，その結果をユーザに伝える，といった対話の進行を決定する．データベース検索の結果をユーザに伝えることを応答生成，そのほかの対話の進行の決定を対話制御と呼ぶ（対話制御という言葉は，文脈理解を含むより広い意味で使われることもある．さらに言語理解・生成部も含む場合もある）．

表層生成は，テキスト生成でシステミック文法や単一化文法を用いた生成システムが研究されてきたのに対し，対話システムでは，多くの場合テンプレートベースの手法が用いられてきた．これは，対話システムでは，生成する文は比較的単純であるからである．しかし，最近はコーパスから得られた統計的な情報を用いて，より自然な文を生成する手法が検討されている [Axelrod 00, Oh 00]．

3. 対話状態とその表現法

対話状態に含まれるべき情報として以下のようなものが提案されている．

対話履歴 その時点までのユーザ発話の理解結果とシステム発話の内容．

ユーザ意図の推定結果 ユーザがこのアプリケーションでどのようなことをしたいかをシステムが推定した結果．例えば，新幹線の座席予約タスクなら，旅行日，時間，乗車駅，降車駅，座席の種類などの情報．

グラウンディング情報 ユーザ意図の推定結果に含まれる情報のうち，どの情報は確認済み（とシステムが思っている）かの情報．例えば図1のU6でユーザは「人数＝1」であることを確認している．

データベースの検索結果 ユーザの意図の推定結果に基づいてデータベースを検索した結果．ユーザの要求にマッチする列車，座席の空き状況など．

談話オブジェクト 今までの対話に出てきた，「もの」や「こと」．ユーザが，「あれ」，「それ」などの指示詞や省略表現などで指し示すことができる．

AI系対話モデルの研究では，さらに，談話（発話）プラン，ドメインプラン，談話スタックなどの情報を用いる．

対話状態に含まれる情報が多ければ多いほど，より適切にシステム応答を生成することができ，複雑なアプリケーションをつくることができるはずである．しかし，音声認識誤りが避けられないことから，対話状態の内容が信頼できるとは限らず，あまり細かい情報を保持することには意味がない．また対話状態が複雑になれば，計算時間もかかり，また，プログラムの保守も大変になる．したがって，アプリケーションと音声認識精度に応じてできるだけ単純なデータにする必要がある．AI系対話モデルの研究では対話状態を表現するのに論理式を用いる場合もあるが，音声対話システムでは図3のようなフレーム構造を用いる場合が多い [Chu-Carroll 99a, Goddeau 96, Lamel 00]．この場合，発話理解は，発話の内容に基づいて，フレームの属性を追加，変更，消去する処理として実現される．発話生成は，フレームを入力として，発話内容を生成し，かつフレームの属性を変更する．

また，古くから用いられてきた対話状態の表現法として，ネットワークを用いる方法がある．これは，図4のように，対話の進行をネットワークで表し，ユーザの発話内容に応じてネットワーク上を遷移する．そして，各ノードに割り当てられた発話を生成する．この方法は，比較的単純なドメインで，後述のシステム主導の対話システムを構築するのに便利な方法であるが，ユーザの自由な発話を許すシステムではアークが多くなりすぎる．なお，ユーザ発話の内容をユーザに確認してグラウンディングするまで遷移しないようにする方法が提案されている [Heeman 98]．

```
[乗車駅: 東京
 降車駅: 京都
 出発時間帯:[時間:9 時
             時間付加情報:すぎ]
 到着時間帯: null
 列車種別: のぞみ
 日にち:[曜日: 水曜日]
 直前の確認要求: {日にち}
 確認済情報: {乗車駅, 降車駅}]
```

図3 フレームによる対話状態の表現

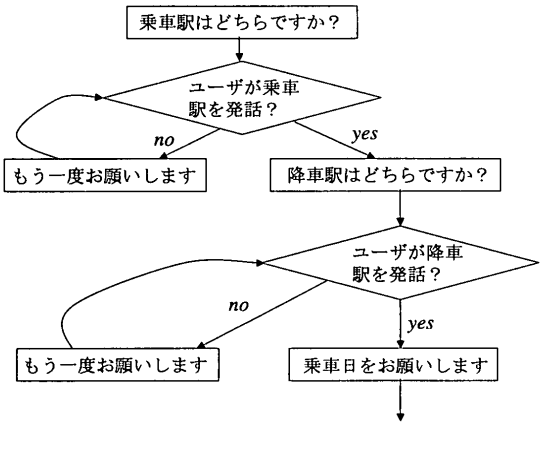

図4 ネットワークモデルによる対話制御

4. ユーザ発話の理解

4·1 言語理解

音声対話システムのユーザの発話は，書き言葉の文に比べると短いので，文の構造を解析するのは比較的簡単に見える．しかし，誤認識や，言い直し，言いよどみを含む発話，断片的な発話などに対処するため，別のタイプのロバストネスが必要である．これに対し，主に，ルールベースの理解と統計的言語理解の二つの方法が試みられている．

ルールベースの理解では，構文規則を用いて文の構造を解析し，対話行為を生成する，通常の言語理解の方法をベースにする．音声対話システムでは，意味的な制約も含む，ドメインに依存した規則が用いられる場合が多い．ただし，規則だけでは誤認識や言い直しなどに対応することができないため，いろいろな拡張が行われている．

部分解析（Partial Parsing）と呼ばれる方法は [Seneff 92]，発話全体を一つの構文木にまとめ上げられないときに，発話の中の解析可能な一部分を使って理解し，あとの部分は無視する方法である．例えば，「今日の，いや，明日のひかり号に乗りたいんですけど」という発話の中の，「明日のひかり号に乗りたいんですけど」という部分が解析可能なら，これだけを用いて理解する．また言い直しや言いよどみを理解するには，これらを検出するためのパターンを用いて，構文解析の前処理で文法的な発話に直す方法などが用いられている [Bear 92]．

これに対し，意味格フレームアプローチ [Hayes 86] は，認識結果の表層表現からキーワードやキーフレーズを取り出して，直接フレーム構造をつくる．例えば，「京都までの自由席特急券はいくらですか」から，「京都まで」，「自由席特急券」，「いくらですか」などの単語やフレーズを抽出して，[対話行為タイプ：値段の質問，目的地:京都，券：自由席特急券] という対話行為を得る．しかし，この方法だと，キーワードしか見ないため，対話行為タイプの推定に難がある．そこで，実際のユーザ発話のコーパスから統計的にパターンを学習する方法が用いられている．統計的機械翻訳の手法 [Epstein 96]，決定木学習 [Magerman 95]，単語ベクトル [Chu-Carroll 99b] を用いる方法などが提案されている．

ルールベースの理解と統計的手法のどちらにも一長一

短がある．規則をつくるにはエキスパートの労力が必要であるが，コーパスを集めるのも多くの人手と労力がかかる．どちらがよいかは，予算と人員に依存する．

言語理解部の仕事の一つに，認識結果の複数の候補の中から，構文意味的に正しいものを一つ選ぶ，または，構文意味制約を用いて順序付けをすることがある．例えば，NBest音声認識結果の中で，単語列全体が構文意味解析可能なものを優先して選択する方法がある．そのほかにも構文的意味的制約を用いて，順序付けをする方法などが提案されている [Zechner 98]．

4·2 文脈理解

文脈理解部の仕事は，言語理解の結果を用いて，対話状態を変更することである．これは逆に言えば，対話状態に応じて言語理解の結果を解釈することである．例えば，図1 U2の「京都です」は，降車駅を意味しているが，これは，その前の発話が，降車駅の確認要求であることから判断できる．図1 U7の「もう一つ早いのは」は，「何より早い」のかが省略されているので，文脈から補う必要がある．このように省略を補ったり，指示詞の指すものを同定する問題は，照応処理と呼ばれている．

これらの問題に関しては，人間どうしの対話やキーボード対話システムに関して多くのモデルが提案されているが [Jurafsky 00]，音声対話システムの場合，扱えるドメインが比較的単純であることもあり，簡単なルールのみを用いている場合が多い．

言語理解が複数の対話行為の候補を出せば，文脈情報を用いて最適な対話行為を選ぶことができ，文脈理解と同時に音声認識，言語理解の曖昧性を解消することができる [Abdou 01]．しかし，この方法はシステム構成や計算が複雑になることから，実現した例は多くない．

5. 対話制御

ユーザの要求を知り，適切に情報を提供するにはどのように対話を進行させればよいだろうか．図1では，まずユーザに要求内容を話してもらい（S1-U1），ユーザの言った内容を確認したり足りない情報を質問したりしながら，データベース検索に必要な情報を得る（S2-U6）．そしてデータベースの検索結果をユーザに提供し，ユーザによいものを選んでもらう（S7-U8）．さらにいくつかの付加情報を得る（S9-S11）．このような対話の進行を司るのが対話制御部の仕事である．データベースの検索に至るまでどのようにユーザの要求を聞き出すか，そしてデータベース検索の結果をどのようにユーザに提示すればより使いやすい対話システムになるか，といったことが対話制御の研究課題である．

一般に，対話制御部は，対話制御規則を用いて発話内容を決定する．対話制御規則は，対話状態を入力として対話行為を出力するようなプロダクションルールである．例えば，図1 S3は，対話状態がフレームで表されているなら，「出発地属性が空なら出発地を聞く」という規則から生成される．ネットワークモデルの場合は，一つ一つのアークが対話制御規則に相当する．

上記の例で，「出発地属性がわかっていないといけない」というのは，アプリケーションに依存した制約である．アプリケーションごとにドメイン依存の規則を用意するのは手間がかかるため，対話制御規則が，アプリケーション依存の知識と，一般的な規則から自動的につくられればよい．アプリケーション依存の知識として，フレームに対する制約が用いられる．例えば，「出発地属性が未指定ではいけない」や「出発時刻は到着時刻より前でなくてはならない」といった制約をアプリケーションごとに記述する．これと，「未指定ではいけない属性をユーザに聞く」などの一般的な対話規則からアプリケーション依存の規則がつくられる [Chu-Carroll 99a]．そのほかアプリケーションに依存しない対話制御の原理を指向した研究として，オブジェクト指向モデリングの方法を用いたもの [Abella 99] や記号論理に基づく方法 [Smith 94] などがある．

5·1 対話戦略

対話規則の内容および適用の仕方は対話戦略と呼ばれる．対話戦略には，いくつかの一般的な分類がある．

一つは対話の主導権をシステムとユーザのどちらにもたせるかによる分類である．主導権がシステムにあるときは，以下の例のように，タスクの遂行に必要な情報をシステムが質問し，ユーザはそれに答える．

S1 乗車駅はどちらですか？
U1 東京です．
S2 降車駅はどちらですか？
U2 京都です．

S1，S2はシステムが主導権をもっている発話である．システム主導では，ユーザの発話が簡潔になり，音声認識の誤りが少なくなるが，慣れたユーザには冗長に感じられる．

逆に主導権がユーザにあるとき，ユーザは自由に話すことができる．

S1 どのようなご用件でしょうか？
U1 東京から京都まで行きたいんですけど．

S1は主導権をユーザに譲渡する発話である．タスクを達成するためには，不足している情報をユーザに尋ねる必要が出てくるため，主導権が最初から最後までユーザにあるシステムは考えにくい．システムが必要なときに主導権を取る必要がある．このような対話システムは，主導権混合型（mixed-initiative）と呼ばれる．

主導権混合型システムは必要に応じて主導権を取ったり譲渡したりする必要がある．例えば，以下のように，

システムの発話が主導権を保持しようとするものであっても，ユーザが主導権を取ろうとする場合もあり，そのような場合を想定して対話を進行させる必要がある．

S1 乗車駅はどちらですか？
U1 東京からなんですが，京都までの自由席特急券はいくらですか？

対話戦略はユーザの発話内容の確認でも分類される．例えば，「東京まで」というユーザ発話に対し，「東京までですか？」のように陽に確認する方法は明示的確認要求（explicit confirmation request）と呼ばれ，「どちらから東京までですか？」のように，次の発話に含めてしまう方法は，暗示的確認要求（implicit confirmation request）と呼ばれている．

また確認の順序にも違いがある．不足している情報をすべて尋ね，データベースに問い合わせる直前に一括して確認する方法や，フレームの属性値が埋まるたびに確認する方法などが考えられる．

5·2 適応的な対話制御

以上述べたように，対話戦略は数多く考えられるため，タスクドメイン，システムのパフォーマンス（音声認識率など），ユーザによって最適な対話戦略を自動的に決定する方法が研究されている．

ユーザ発話の音声認識結果の信頼性尺度や，音声認識精度に応じて，不要な確認を減らし，最短の対話でタスクを達成できるように，最適な確認手順を選択する対話制御法がある [Niimi 96, Yasuda 01].

また，対話をマルコフ決定過程（Markov Decision Process:MDP）と見なし，強化学習（Reinforcement Learning）を用いて，最適な対話戦略を学習する方法がある [Levin 00, Singh 02, Walker 00]．マルコフ決定過程では，システムが，状態 S_i で行動 a を選択すると，状態が確率的に S_j に遷移し，報酬 r を得る．報酬の値も確率的に決まる．また，報酬の値は，そのときすぐに決まるのではなく，先の状態に依存して決まる．例えば，ロボットが目的地までの最短経路を試行錯誤を繰り返しながら決める場合を考えよう．ロボットは適当な道を選んで進むのだが，障害物などがあって，1ステップ後どこにたどり着くかは，確率的にしかわからないとする．選んだ道がどのくらい正しかったか（すなわち報酬の値）は，このステップでどのくらい最終目的地に近づいたかではなく，先のステップでどのくらい近づいていくかに依存して決まる．ここで，最も報酬の期待値が高くなるように，各状態における最適な行動 $a'(S)$ を，実際にシステムが行動を行った結果に基づいて学習するのが強化学習である．状態を対話状態，行動をシステム発話と読み替え，報酬の値を対話のパフォーマンス（例えばタスクが達成したか，対話の短さ，ユーザの満足度）で決めると，システムにとってユーザとの対話はMDPで表せることがわかる．したがって，人間との対話のデータを使い，強化学習によって，最適な対話戦略を計算することができる．

MDPは，ある時点でのシステムの状態は一意に決まることを仮定しており，音声の誤認識によって，システムの状態が一意に決まらないことを直接的にモデル化できない．この制限を外し，センサによる認識結果に応じてシステムの状態が確率的に決定されることをモデル化したものがPOMDP（Partially Observable MDP）である．POMDPに基づいた音声対話システムの対話制御法についても研究が進められている [Roy 00].

また，主導権の交代を適応的に行う対話システムもつくられている [Chu-Carroll 00]．ユーザ発話の中の，主導権の授受を行う鍵となる句（cue phrase）の存在などの情報を用い，主導権がシステムとユーザのどちらにあるかを推定し，それに基づいて対話制御を行う．常にシステムが主導権をもつシステムに比べて，よいパフォーマンスが得られることが実験的に示されている．

また列車予約のようにデータベースが刻々と変わる場合には，データベースの内容に応じて確認手順を変えることにより，効率的な対話が行える [堂坂 02]．例えば，ユーザが「5月3日の9時ごろに東京を出る新幹線を大阪まで予約したい」と言ったとき，もし，5月3日の東京から大阪までの新幹線がすべて満席であったとしたら，「5月3日」，「東京」，「大阪」だけを確認すれば，「9時ごろ」を確認する必要はなく，「すべて満席です」と言えばよい．

5·3 発話交代

現状の音声対話システムの多くは，ユーザとシステムの発話交代が整然と行われることを仮定している．すなわち，ユーザが一定の長さのポーズを置くとシステムが話し始め，システムが話し終わるまでユーザ発話の音声認識を開始しないシステムである．しかし，システムが話している間にユーザが話し始めることがある．これをバージイン（barge-in）と呼ぶ．バージインを正しく処理するためには，バージインを正しく検出し，その内容に応じてシステム発話をとめたり内容を変更したりすることが必要となる [Dohsaka 97]．例えばユーザのバージインが相槌ならそのまま話し続ける必要があるし，問い返しなら繰り返す必要がある．

また，ユーザのポーズが必ずしも発話交代のタイミングとは限らない．システムは，ユーザの発話の途中でも，いつ発話を開始すべきかを決定する必要がある．そのためには，ユーザが発話している途中でもシステムは発話理解を行う必要がある [Allen 01, Nakano 99]．また，（特に日本語の場合）ユーザが話している途中で相槌をうつことによって，ユーザの発話を聞いたり理解していることを示すことが効果的である．そのため，ユーザ発話の韻律，文法情報などを用いて相槌のタイミングを決

定することが試みられている [Noguchi 00, Ward 00].

6. 意図理解と協調的応答生成

ユーザの要求がわかると，システムはデータベースを参照して，ユーザに情報を提供する．この処理を一般に応答生成と呼ぶ．現状の音声対話システムでは，ドメインが比較的単純なため，あまり応答生成には注意が払われていない．例えば単純に新幹線の予約を行うようなアプリケーションでは，データベースをサーチして，ユーザの希望に最も近い候補の列車をあげれば，とりあえずタスクは達成できる．

しかし，例えば，「5月3日のひかり3号を新横浜から京都まで予約したい」というユーザの要求に対し，ただ「その列車は満席です」とだけ答えるより，「一つ早い列車なら空いています」と付け加えるほうが親切である．このような応答は協調的応答と呼ばれ，AI系対話モデルの枠組みで研究されている．

協調的な応答を行うためには，ユーザがこの対話によって本当は何をしたいのか，という意図理解が必要である．単純なタスクだけを行うシステムの場合，「要求されている列車が満席なら前後の空いている列車を提案する」といった規則を用意しておけば，単純な処理ですむ．しかし，複数の区間を乗り継ぐような予約が行える対話システムの場合，ユーザが最終的にどこに何時までに行きたいのかという意図を把握していなければ，乗継ぎが不可能な代替案を出してしまい，協調的でなくなる．

意図理解を行うための道具として，プラン認識が用いられている [Allen 80]．プランとは，目標を達成するために，どのような行為をどのような順序で行えばよいかを表現するものである．ユーザの発話からシステムはユーザのプランを推定し，それをもとに協調的応答を生成する方法が提案されている [Chu-Carroll 98]．また，複数のユーザ発話が一つのプランの一部をなすことを仮定すると，ユーザ発話の解釈に曖昧性があるとき，うまくプランが構成できるような解釈を選ぶことにより，曖昧性の解消ができることも示されている [Carberry 90].

ユーザのプランを推定するだけでなく，システム発話の生成にもプランニングが有効である．例えば，ユーザに経路や装置の使い方などを説明するシステムでは，システムの応答が長くなるため，一度に全部話してもうまく伝わらない場合がある．このため，それまでの対話の履歴からのユーザの知識の推定結果などをもとに，何をどういう順序で発話すべきかというプランを生成する必要がある [Walker 01]．また，説明中のユーザの反応に応じてプランをつくり直しながら対話を進めていくという方法が用いられる [Cawsey 92].

AI系対話モデルの詳細については最近の教科書 [石崎 01, Jurafsky 00, 田中 99] を参考にしていただきたい．

近年，これらのAI系対話モデルと音声認識・合成を組み合わせた対話システムがつくられている [Allen 01, Rich 01]．これらのシステムは，音声認識がうまくいかないときのロバストネスにあまり注意を払っていないため，一般ユーザが使用したときのパフォーマンスはあまりよくないと思われる．しかし，音声認識との統合を考慮してモデルが再構築されれば，現状の音声対話システムより複雑なタスクを行うシステムを構築することができるだろう．

7. 複数のモジュールに関わる技術

最近では，各モジュールの高性能化だけではなく，複数のモジュールの処理に関係するような技術が研究されている．そのようなもののうち，今後の発展が期待されるものをまとめる．

訂正発話検出 システムからの明示的または暗示的確認要求発話に対する，ユーザの訂正要求（図1U2など）を検出する．このとき，音声認識結果だけを手がかりにすると，訂正要求自体が誤認識される場合にうまくいかず，対話がなかなか前に進まない．したがって，対話履歴，韻律，発話長などの情報も合わせ，コーパスから学習された決定木を用いて検出する方法が提案されている [Litman 01].

未知語理解 システムが想定していない単語（Out-of-Vocabulary Word: OOV）をユーザが話すことがある．音声処理でOOVの検出・認識について研究が進んでいるが，言語・対話処理でも，OOVを含む発話を理解し，適切に対話を進行させたり，またOOVの情報を新しい知識として取り込んだりする必要がある [Meng 01].

韻律を利用した発話の理解 発話の意図推定，特に対話行為タイプの決定を，音声認識結果だけではなく，韻律も利用して行う方法が検討されているが [Shriberg 98]，音声対話システムで有効に活用された例はまだ少ない．

システム発話の韻律生成 システム発話の音声合成を行うとき，特定の単語や句を強調するほうが意図がうまく伝わり，自然に聞こえることがある．例えば図1S4では，「新横浜」より，確認を求める「京都」を強調するほうがよい．このために，言語・対話処理部は音声合成器に適切に情報を伝える必要がある [Nakatani 00].

言語モデルの動的変更 音声認識の言語モデルを，対話状態に応じて変更することにより，認識率を向上させる技術 [Esteve 01].

対話システムの総合評価 対話システムの個々のモジュールを評価するには，そのモジュールだけで評価するのでは不十分で，対話システム全体のパフォーマンスの向上にどのくらい貢献したかを測定

する必要がある．そのために，対話システム全体のパフォーマンスを測定する方法が研究されている．例えば，タスクの達成度と，所要時間などのユーザにかかった負担からパフォーマンスを計算する手法が提案されている [Walker 98].

ポータビリティ　音声対話システムは多くのモジュールからなり，おのおののモジュールがアプリケーション依存の規則や知識を必要とする．したがって，新しいアプリケーションをつくる際に知識や規則を書くためのエキスパートの労力が必要である．また，規則や知識間の整合性を保つためのメンテナンスにも労力がかかる．この労力を削減し，かつ，音声言語処理の研究者以外でも対話システムを開発できるようにするためのツールが構築されている [Glass 01, Kogure 00].

そのほか，本稿で取り上げなかった話題については，音声言語処理に関する最新の教科書やサーベイ [Huang 01, Jurafsky 00, 田中 99, Zue 00] を参考にされたい．

8. おわりに

以上，音声対話システムの言語・対話処理の技術を概観した．音声対話システム研究は，音声認識の性能向上を背景に，プロトタイプシステムと一般ユーザとの対話のデータの収集が可能になったことから，新たな展開を見せている．今後はこれらのデータをもとにしたアプリケーション依存の知識源の自動学習技術の研究がますます盛んになるだろう．また，現在のシステムがユーザに課しているさまざまな制限（認識語彙，発話の内容，発話のタイミングなど）を取り除いた対話システムが研究されていくと思われる．

謝　辞

NTT コミュニケーション科学基礎研究所マルチモーダル対話グループ各位および MIT 音声言語システムグループ各位をはじめとする，日頃議論させていただく方々に感謝します．

◇ 参考文献 ◇

[Abdou 01] Abdou, S. and Scordilis, M.: Integrating Multiple Knowledge Sources for Improved Speech Understanding, *Proc. 7th Eurospeech,* pp. 1783-1786 (2001)

[Abella 99] Abella, A. and Gorin, A. L.: Construct Algebra: Analytical Dialogue Management, *Proc. 37th ACL* (1999)

[Allen 80] Allen, J. F. and Perrault, C. R.: Analyzing Intention in Utterances, *Artificial Intelligence,* Vol. 15, pp. 143-178 (1980)

[Allen 01] Allen, J., Ferguson, G. and Stent, A.: An Architecture for More Realistic Conversational Systems, *Proc. IUI' 01,* pp. 1-8 (2001)

[Axelrod 00] Axelrod, S.: Natural Language Generation in the IBM Flight Information System, *Proc. NAACL-ANLP 2000 Workshop on Conversational Systems* (2000)

[Bear 92] Bear, J., Dowding, J. and Shriberg, E.: Integrating Multiple Knowledge Sources for the Detection and Correction of Repairs in Human-Computer Dialog, *Proc. 30th ACL,* pp. 56-63 (1992)

[Carberry 90] Carberry, S.: *Plan Recognition in Natural Language Dialogue,* MIT Press, Cambridge, Mass. (1990)

[Cawsey 92] Cawsey, A.: Explanation and Interaction: *The Computer Generation of Explanatory Dialogues,* MIT Press, Cambridge, Mass. (1992)

[Chu-Carroll 98] Chu-Carroll, J. and Carberry, S.: Collaborative Response Generation in Planning Dialogues, *Computational Linguistics,* Vol. 24, No. 3, pp. 355-400 (1998)

[Chu-Carroll 99a] Chu-Carroll, J.: Form-Based Reasoning for Mixed-Initiative Dialogue Management in Information-Query Systems, *Proc. 6th Eurospeech,* pp. 1519-1522 (1999)

[Chu-Carroll 99b] Chu-Carroll, J. and Carpenter, B.: Vector-Based Natural Language Call Routing, *Computational Linguistics,* Vol. 25, No. 3, pp. 361-388 (1999)

[Chu-Carroll 00] Chu-Carroll, J.: MIMIC: An Adaptive Mixed Initiative Spoken Dialogue System for Information Queries, *Proc. 6th Applied NLP,* pp. 97-104 (2000)

[Dohsaka 97] Dohsaka, K. and Shimazu, A.: System Architecture for Spoken Utterance Production in Collaborative Dialogue, *Working Notes of IJCAI 1997 Workshop on Collaboration, Cooperation and Conflict in Dialogue Systems* (1997)

[堂坂 02] 堂坂浩二, 安田宜仁, 相川清明: システム知識制限下での効率的音声対話制御法, 自然言語処理, Vol. 9, No. 1, pp. 43-63 (2002)

[Doran 01] Doran, C., Aberdeen, J., Damianos, L. and Hirschman, L.: Comparing Several Aspects of Human-Computer and Human-Human Dialogues, *Proc. Second SIGdial Workshop,* pp. 48-57 (2001)

[Epstein 96] Epstein, M., Papineni, K., Roukos, S., Ward, T. and Pietra, S. D.: Statistical natural language understanding using hidden clumpings, *Proc. ICASSP-96,* pp. 176-179 (1996)

[Esteve 01] Esteve, Y., Bechet, F., Nasr, A. and Mori, R. D.: Stochastic Finite State Automata Language Model Triggered by Dialogue States, *Proc. 7th Eurospeech,* pp. 725-728 (2001)

[Glass 01] Glass, J. and Weinstein, E.: SpeechBuilder: Facilitating Spoken Dialogue System Development, *Proc. 7th Eurospeech,* pp. 1335-1338 (2001)

[Goddeau 96] Goddeau, D., Meng, H., Polifroni, J., Seneff, S. and Busayapongchai, S.: A form-based dialogue manager for spoken language applications, *Proc. 4th ICSLP* (1996)

[Hayes 86] Hayes, P. J., Hauptmann, A. G., Carbonell, J. G. and Tomita, M.: Parsing Spoken Language: A Semantic Caseframe Approach, *Proc. 11th COLING,* pp. 587-592 (1986)

[Heeman 98] Heeman, P. A., Johnston, M., Denney, J. and Kaiser, E.: Beyond Structured Dialogues: Factoring Out Grounding, *Proc. 5th ICSLP,* pp. 863-866 (1998)

[Huang 01] Huang, X., Acero, A. and Hon, H. -W.: *Spoken Language Processing,* Prentice-Hall (2001)

[石崎 01] 石崎雅人, 伝 康晴: 談話と対話, 東京大学出版会 (2001)

[Jurafsky 00] Jurafsky, D. and Martin, J. H.: *Speech and Language Processing,* Prentice-Hall (2000)

[Kogure 00] Kogure, S. and Nakagawa, S.: A Portable Development Tool for Spoken Dialogue Systems, *Proc. 6th ICSLP* (2000)

[Lamel 00] Lamel, L., Rosset, S., Gauvain, J., Bennacef, S., Garnier-Rizet, M. and Prouts, B.: The LIMSI ARISE system, *Speech Communication,* Vol. 31, pp. 339-353 (2000)

[Levin 00] Levin, E., Pieraccini, R. and Eckert, W.: A Stochastic Model of Human-Machine Interaction for Learning Dialogue Strategies, *IEEE Transaction on Speech and Audio Processing,* Vol. 8, No. 1, pp. 11-23 (2000)

[Litman 01] Litman, D., Hirschberg, J. and Swerts, M.: Predicting User Reactions to System Error, *Proc. 39th ACL,* pp. 370-377 (2001)

[Magerman 95] Magerman, D. M.: Statistical Decision-Tree Models for Parsing, *Proc. 33th ACL,* pp. 276-283 (1995)

[Meng 01] Meng, H., Chan, S. F., Wong, Y. F., Chan, C. C., Wong, Y. W., Fung, T. Y., Tsui, W. C., Chen, K., Wang, L., Wu, T. Y., Li, X., Lee, T., Choi, W. N., Ching, P. C. and Chi, H.: ISIS: A Learning System with Combined Interaction and Delegation Dialogs, *Proc. 7th Eurospeech* (2001)

[Nakano 99] Nakano, M., Miyazaki, N., Hirasawa, J., Dohsaka, K. and Kawabata, T.: Understanding Unsegmented User Utterances in Real-Time Spoken Dialogue Systems, *Proc. 37th ACL,* pp. 200-207 (1999)

[Nakatani 00] Nakatani, C. H. and Chu-Carroll, J.: Coupling Dialogue and Prosody Computation in Spoken Dialogue Generation, *Proc. 6th ICSLP* (2000)

[Niimi 96] Niimi, Y. and Kobayashi, Y.: Dialog control strategy based on the reliability of speech recognition, *Proc. 4th ICSLP* (1996)

[Noguchi 00] Noguchi, H., Katagiri, Y. and Den, Y.: An Experimental Verification of the Prosodic Lexical Effects on the Occurrence of Backchannels, *Proc. 6th ICSLP* (2000)

[Oh 00] Oh, A. and Rudnicky, A.: Stochastic language generation for spoken dialogue systems, *Proc. NAACL-ANLP 2000 Workshop on Conversational Systems,* pp. 27-32 (2000)

[Rich 01] Rich, C., Sidner, C. and Lesh, N.: COLLAGEN: Applying Collaborative Discourse Theory, *AI Magazine,* Vol. 22, No. 4, pp. 15-25 (2001)

[Roy 00] Roy, N., Pineau, J. and Thrun, S.: Spoken Dialogue Management Using Probabilistic Reasoning, *Proc. 38th ACL,* pp. 93-100 (2000)

[Rudnicky 00] Rudnicky, A. I., Bennett, C., Black, A., Chotomongcol, A., Lenzo, K., Oh, A. and Singh, R.: Task and Domain Specific Modelling in the Carnegie Mellon Communicator System, *Proc. 6th ICSLP* (2000)

[Seneff 92] Seneff, S.: Robust Parsing for Spoken Language Systems, *Proc. ICASSP-92,* pp. 23-26 (1992)

[Seneff to appear] Seneff, S.: Response Planning and Generation in the MERCURY Flight Reservation System, *Computer Speech and Language* (to appear)

[Shriberg 98] Shriberg, E., Bates, R., Stolcke, A., Taylor, P., Jurafsky, D., Ries, K., Coccaro, N., Martin, R., Meteer, M. and Ess-Dykema, C. V.: Can Prosody Aid the Automatic Classification of Dialog Acts in Conversational Speech?, *Language and Speech,* Vol. 41, No. 3-4, pp. 439-487 (1998)

[Singh 02] Singh, S., Litman, D., Kearns, M. and Walker, M.: Optimizing Dialogue Management with Reinforcement Learning: Experiments with the NJFun System, *Journal of Artificial Intelligence Research,* Vol. 16, pp. 105-133 (2002)

[Smith 94] Smith, R. W. and Hipp, D. R.: *Spoken Natural Language Dialoue Systems,* Oxford University Press (1994)

[Sturm 99] Sturm, J., den Os E. and Boves, L.: Dialogue Management in the Dutch ARISE Train Timetable Information System, *Proc. 6th Eurospeech,* pp. 1419-1422 (1999)

[田中 99] 田中穂積 編: 自然言語処理—基礎と応用—, 電子情報通信学会 (1999)

[Walker 98] Walker, M. A., Litman, D. J., Kamm, C. A. and Abella, A.: Evaluating Spoken Dialogue Agents with PARADISE: Two Case Studies, *Computer Speech and Language,* Vol. 12, No. 3, pp. 317-347 (1998)

[Walker 00] Walker, M. A.: An Application of Reinforcement Learning to Dialogue Strategy Selection in a Spoken Dialogue System for Email, *Journal of Artificial Intelligence Research,* Vol. 12, pp. 387-416 (2000)

[Walker 01] Walker, M. A., Rambow, O. and Rogati, M.: SPoT: A Trainable Sentence Planner, *Proc. 2nd NAACL,* pp. 17-24 (2001)

[Ward 00] Ward, N. and Tsukahara, W.: Prosodic Features which Cue Back-channel Responses in English and Japanese, *Journal of Pragmatics,* Vol. 23, pp. 1177-1207 (2000)

[Yasuda 01] Yasuda, N., Dohsaka, K. and Aikawa, K.: Spoken Dialogue Control Based on a Turn-minimization Criterion Depending on the Speech Recognition Accuracy, *Proc. Second SIGdial Workshop,* pp. 210-213 (2001)

[Zechner 98] Zechner, K. and Waibel, A.: Using Chunk Based Partial Parsing of Spontaneous Speech in Unrestricted Domains for Reducing Word Error Rate in Speech Recognition, *Proc. 17th COLING and 36th ACL,* pp. 1453-1459 (1998)

[Zue 00] Zue, V. W. and Glass, J. R.: Conversational Interfaces: Advances and Challenges, *Proc. IEEE,* Vol. 88, No. 8, pp. 1166-1180 (2000)

2002年3月26日 受理

著者紹介

中野 幹生（正会員）

1988年東京大学教養学部基礎科学科第一卒業．1990年同大学院理学系研究科相関理化学専攻修士課程修了．同年日本電信電話（株）入社．2000年よりMIT計算機科学研究所客員研究員．音声対話システム，音声言語理解の研究に従事．博士（理学）．ACL，ISCA，ACM，情報処理学会，言語処理学会各会員．

堂坂 浩二（正会員）

1984年大阪大学基礎工学部情報工学科卒業．1986年同大学院博士前期課程修了．同年，日本電信電話（株）入社．現在，NTTコミュニケーション科学基礎研究所勤務．音声対話システム，言語生成，文脈理解の研究に従事．情報処理学会平成9年度論文賞受賞．言語処理学会，ACL，情報処理学会，電子情報通信学会各会員．