Vol. 34 No. 10　　情　報　処　理　　Oct. 1993

解　説

IPS

自然言語処理技術の応用

7. 音声理解と対話における自然言語処理†

竹　林　洋　一††

1. はじめに

音声は人間の最も自然なコミュニケーション手段であり，文字の発明前から人間は音声言語により日常の会話を行っている．音声メディアの有する自然性などの性質は，人間同士の会話に限らず，計算機との対話手段としても望ましいものである．このため，音声合成は計算機からユーザへの情報提示手段として，音声認識・理解は計算機への情報入力手段として期待され，各国の研究機関で巨費を投じて多くの研究がなされてきた．

音声入力は連続音声認識が可能な場合，計算機のインタフェースとして他のメディアよりも入力速度が速く，手や足を他の用途に用いていても並行入力可能であり，キーボードに比べて訓練が不要であるなどの利点がある．しかし，音声認識・理解システムを実際に構築する場合，シンボル・レベルの自然言語の曖昧性のほかに，音声パターンの認識エラーや曖昧性が避けられないため，現状のシステムでは音声による計算機との対話を円滑に進行するには至っていない．音声符合化や音声合成の応用と同様に，音声認識の応用を拡大するためには，認識性能の向上と曖昧性への対処が必須であり，自然言語処理はそのためのキー技術として重要視されている[1),2)]．

音声処理の分野では，音声から自然言語へのメディア変換を音声認識 (Speech Recognition) と呼ぶ．音声ワードプロセッサやディクテーションマシンが連続音声認識システムに相当し，HMM (Hidden Markov Model, 隠れマルコフモデル) に代表される確率統計モデルに基づく大語彙連続音声認識の研究が盛んに行われてきた[3)]．このような表層的なメディア変換を指す連続音声認識に対して，発話の意味を抽出する音声理解 (Speech Understanding) や，音声の言語の側面を重視した音声言語システム (Spoken Language System)，さらに，計算機とのインタフェースとして音声対話 (Speech Dialogue) の研究が盛んになり，音声処理と自然言語処理の融合への気運が急速に高まっている[4)~7)]．

一方，音声入力の普及を後押しする要因として，計算機の高速化とマルチメディア化の進展があげられる．汎用ワークステーションにオーディオ入出力が標準装備されるようになり，音声信号の取込みや音声分析に必要であった外付けのハードウェアが不要となり，音声処理のコストが低下し，ユーザの裾野が広がってきた[8)]．この傾向は今後も加速化するのは確実であり，シミュレーションや理論研究に留まらず，音声による計算機とのインタフェースの構築が活発になることが予想される[9)]．

以下，本稿では自然言語処理技術の応用という観点から，音声理解・対話処理について解説する．

2. 音声言語処理

2.1 音声認識・理解と自然言語処理

音声言語処理と自然言語処理は言語情報を扱うという点では共通であるが，音声信号は発話者の個人性や感情などの情報を含み，さらに環境騒音などの影響を受けるために文字記号列に比べて多様でバリエーションが格段に多い．また，連続音声の場合には前後の音素環境の影響を受ける調音結合と呼ばれる現象があり，音響分析の結果観察される特徴パラメータの音素境界が不明確となったり，音素に対応する特徴パラメータの性質が単独発声のものと異なってしまうため，連続音声を機械的に音素などの離散的な記号に変換できな

† Natural Language Processing in Speech Understanding and Dialogue by Yoichi TAKEBAYASHI (Research and Development Center, Toshiba Corporation).

†† 東芝研究開発センター情報・通信システム研究所

い．このため，音声処理はインタフェースとして自然で迅速であるという利点がある反面，分析などに多くの計算機パワーを要するためコストが高く，音声認識・理解の曖昧性やエラーにより信頼性が低下するという大きなハンディがある．

音声認識に利用できる言語情報としては，音素の設定や単語の音韻情報の表現(音韻論)，単語の並びとしての文に関する構文情報(統語論)，意味情報(意味論)，文脈情報・プラグマティクス（語用論）がある．すなわち，自然言語処理で開発された各種言語情報や処理機構が利用可能である．音響特徴パラメータに近い音素などに関する低次言語情報と，記号化された後の構文・意味・文脈情報などの高次言語情報の適切な活用により，音声認識・理解・対話処理の性能向上が可能である．

図-1 に典型的な音声理解の処理過程を示す．入力音声は音響分析され，音素認識とセグメンテーションが行われ，音素ラティスが出力される．次に，音素ラティスを解析して単語認識を行い，単語ラティスを構文解析し，意味解析して発話文の意味表現を求める．このような連続音声理解の研究は，1970年代の米国の ARPA の音声理解プロジェクトで本格的に始められた．応用タスクを明確に定義し，大規模な音声理解システムの構築を目指して，音声処理と自然言語処理の統合が試みられた．CMU の HARPY を除いて大部分のシステムは未完成に終わったが，野心的なプロジェクトを通じて構文や意味レベルの知識の有効性が明らかとなった．たとえば，Hearsay-II の開発を

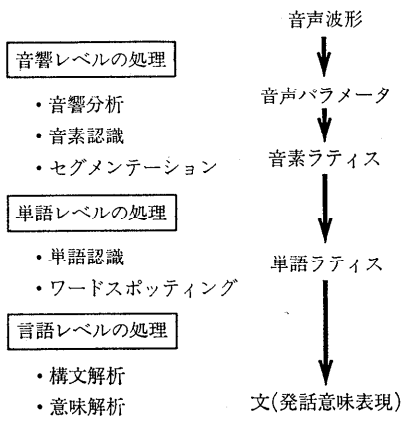


図-1　典型的な音声理解の処理過程

通じて各種知識を利用するブラックボード・モデルが提案され，人工知能や自然言語処理の進歩にも大きな貢献をした．また，音響分析や音韻・単語レベルのパターン認識などの低次処理の精度が不十分な場合，いかに高度な高次処理を行っても砂上の楼閣となるとの教訓を残した．その結果，過度な知識処理への期待を反省して，アルゴリズムとデータ中心の確率モデルに基づく音声処理に研究が移行することとなった．

1985年に始まった DARPA (Difence, Advanced Research Program Agency) の音声認識プロジェクトは，音声信号を自然言語へ変換する際の単語や文の認識性能の向上を目的とし，語彙の拡大と不特定話者化に注力した．その中で HMM は，音声データベースの整備と計算機パワーの増大を背景に，音声認識の中核的手法として用いられ，高精度の不特定語彙連続音声認識システムが開発された．CMU の SPHINX や BBN の BYBLOS などの代表的連続音声認識システムでは，確率統計的なアプローチを基本とし，伝統的な統語構造や意味構造の抽出を行わずに，音声を文字列へ変換する．現状では，言語的制約として，単語や音韻などの2連鎖，3連鎖の共起確率 (Bigram, Trigram) が用いられている場合が多い．すなわち，音声認識の曖昧性に対処するため，単語などの表層的な言語情報の確率・統計的な制約により，低レベルの音声認識誤りを修正し，文認識率を向上させている．このように強力な HMM の隆盛により，音声認識精度向上のための高次言語処理や知識処理の研究は最近まで下火となっていた．

確率統計モデルに基づく音声認識はモデルの詳細化と改良が進み，性能向上が飽和してきた1988年ごろから，音声認識と自然言語処理の統合の重要性が再認識され，音声言語システムの研究に重心が移ってきた．自然言語処理技術については，従来の書き言葉 (written language) を対象とするのではなく，話し言葉 (spoken language) を扱う新しい自然言語処理の重要性が指摘され，音声と自然言語処理の有機的な融合が研究課題として浮かび上がってきた．

2.2　音声認識における言語モデル

自然言語処理における言語モデル（文法）は，言語として妥当な解析的構造を得るために用いら

れるのに対して，音声認識における言語モデルは，言語として妥当でない単語系列（文字列）を排除して音声認識の探索空間を狭め，音声認識の不完全さを補う役割を担っている．

80年代の音声認識システムでは，確率モデルの利用により音声の曖昧性へ対処するというアプローチが主流となった．このようなシステムでは，音素レベルの処理にはHMMモデルを用い，単語・文レベルの処理は言語モデルにより単語系列に対して制約がかけられる．この制約は伝統的な自然言語処理の枠組とは異なる単語（品詞）の2連鎖，3連鎖の確率を表すBigram文法やTrigram文法，あるいは，単語ペア文法が用いられる場合が多い[10]．また，かな・漢字のTrigramを日本語のディクテーションシステムに用いた例もある[11]．これらの単語などの連鎖確率は大量のテキストを統計処理することにより推定する．また，正規文法や文脈自由文法を言語モデルとして採用するシステムも増えつつある．連続音声認識は，上述した各種言語モデルの下で，単語候補系列（単語ラティス）をパージングし，文全体の尤度を最大にする単語系列を求める探索問題として定式化できる．

音韻や単語などの低次認識過程に生じる誤認識や曖昧性に対処するため，音韻や単語などの候補系列からなるラティスのパージングを効果的に行う種々の方法が提案されている．パージング方式には，単語の認識検出と単語系列の評価とを別処理として行うラティス・パージングと，単語認識と単語系列の評価を同時進行する連鎖単語認識アルゴリズムの二種類がある．自然言語処理で提案された一般化LRパーザ，Earleyアルゴリズムなどが適用可能であり，リアルタイム処理のための高速解析方法が提案されている[12]．他にも，柔軟な制御が可能なチャートパーザや，C.Y.K.アルゴリズムも音声認識に用いられている[13),14]．

音声認識のパージングにおける曖昧性には，上述した低次音声認識処理の不完全さに起因する曖昧性のほかに，単語系列などの高次の解析における解釈上の曖昧性がある．音声認識の分野では，前者の曖昧性に対する処理に主眼が置かれてきたが，音声言語システム構築の際には，後者の曖昧性によりさらにパージングに要する処理量が増大するという問題がある．大語彙連続音声認識システムでは，Trigramなどの緩い言語的制約により連結音声中から文候補（単語系列）を探索し，文脈自由文法をテーブル化して，探索により得られたN-bestな単語系列を一般化LRパーザにより，高速にパージングする方法が提案されている[15]．また，上述した言語モデル下の探索の方法では最適経路を効率的に求めるため，ビーム探索を採用している場合が多く，動的計画法の利用やA*探索が検討されている[16]．

一種の確率文法であるBigram文法やTrigram文法の詳細化も検討されている．音声認識における言語モデルの役割は，発話文として存在しえない単語系列を排除し，探索空間を縮小して認識精度を上げることである．ところが，しばしば用いられる文脈自由文法などでも文法的制約が不十分であり，非文を多く生成してしまうという問題がある．そこで，文脈自由文法などの文法を確率化した確率文脈自由文法による音声言語のモデル化法が提案されている．確率文法の導入により，文脈自由な書換え規則にその規則が適用される確率が与えられるので，非文の生成を抑制できる．また，文脈に依存した確率を与えれば，さらに高精度化が可能であるが，このような文法の学習には大量の学習データを必要とするので，学習用データの自動作成方式も検討されている[17]．

2.3 音声理解と音声対話

音声言語と文字言語は共通な点も多いが，音声は対話的（interactive）なメディアであるため，話し言葉が多様であることに加えて，対話の際に話し手と聞き手が相互に入れ替わり相互に影響し合うダイナミックな性質を有する．このため，音声言語処理では，単文（一発話）の音声理解にとどまらず，人間と計算機の双方向のコミュニケーションを研究対象とする音声対話の研究が重要視されてきた．**図-2**に示すように，計算機との自然な対話を実現するためには，音声認識，自然言語処理，Human-Computer Interactionの研究者が互いに密接に連携し，各要素技術を融合した新しい音声対話システムを開発する必要がある．しかし現状では，音声，自然言語，HIの各分野は個別に研究され，共通部分を指向した研究は少ないが，コンピュータ側からの音声言語処理への期待は急速に高まってきている．

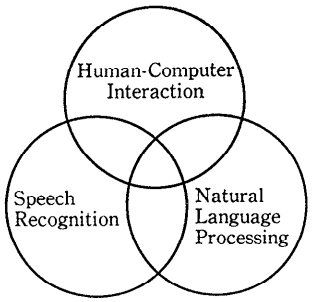


図-2 計算機との対話における音声認識と自然言語処理の関係

音声による計算機との自然な対話を実現するためには，ユーザが自然かつ自由に発声した話し言葉を聞きとる自由発話理解の高度化が必須である．自由発話音声 (Spontaneous Speech) には，言い淀み，ポーズ，繰返しなどが含まれ，日本語では助詞の脱落や倒置が頻発し，そのバリエーションは書き言葉よりもはるかに多い[18]．このため，自由発話の現象を文法として記述することは困難であり，従来の自然言語処理の枠組では対処できない．自由発話理解の方法として，ユーザの息つぎ，ポーズ，言い直し，環境雑音などに対して，音響レベル対処するガーベジ・モデルが提案されている．また，フレーズごとに文法を用意する柔軟なパージング方法やキーワードを基本とする方法が提案されている[19),35]．自由発話理解ではロバスト性が最重要であり，未登録語への対処方式も報告されている[20]．

人工知能や自然言語処理の分野でも談話理解や対話処理の研究がなされているが，人間と人間の対話の解析やモデル化が中心であり，文字言語を処理対象としているため，音声認識のエラーや曖昧性に関する検討はなされていない．人間が音声を聞きとる場合，発話の状況や文脈，常識や話し手との話題に関する知識を利用して，絶えず次の発話を予測し，推論を行っている．話し手の発話の予測が困難なときには，人間でも音声を正確に聞きとることは容易ではない．ロバストな音声理解・対話システムを実現するためには，上述した状況や文脈情報および話題に関する知識の利用が不可欠である．対話の文脈情報を用いた対話音声理解と，状況，対話の状況に応じた質問・確認による適切な曖昧性解消を，対話システムに組み込む必要がある．

音声言語には言語情報のほかにイントネーションやストレスなどの韻律情報が含まれており，句の境界や，強調点，疑い，感情などの情報を伝達している[21]．また，音韻情報を利用した連続音声中の句や文節の境界の検出も検討されている．音声対話システム構築のためには，積極的に韻律情報を利用して，強調点や意図・感情の認識に活用することが可能である．日常会話で頻繁に発話する「えー，えっ，ああ」などの声的感動詞（非言語音声）の音声認識とその利用についても検討されている[22]．このような韻律情報の利用は，音声認識率向上にも有効であるが，日常会話で頻繁に使われる情報であるため，迅速で自然な計算機とのコミュニケーションを実現するためのキー技術になると考えられる．

3. 音声理解・対話システム

3.1 研究動向

計算機性能の向上とマルチメディア化の進展により，オーディオインタフェースが標準装備となったワークステーションが普及し，コンピュータで手軽に音声が処理できる環境が整ってきた．HMM に基づく強力な音声認識も，高性能ワークステーション上で実時間動作が可能となり，音声による計算機との対話が現実味をおびてきた．音声関連の国際会議である ICASSP, ICSLP や DARPA のワークショップでも，音声言語システム，および，その実時間システムの開発に対する関心が高まってきた．従来は，ニューラルネットによる音韻認識率の向上や，HMM モデルの詳細化に研究者の労力が集中し過ぎていたきらいがあるが，実際の応用場面で役立つ音声言語処理システムに熱い視線が集まってきた[23)~25)]．

音声理解・対話システムの研究では，CMU, SRI, BBN, MIT などが研究遂行中の DARPA プロジェクトの規模が最も大きい．これらの研究機関では，定められたタスクに対して HMM をベースにひたすら認識率の向上に努め，音声認識処理後に複数の発話文候補を自然言語処理部へ送り，音声認識の曖昧性に対処していた[26),27)]．MIT では，音声認識と自然言語処理などを密接に統合した知的音声言語システム VOYAGER を開発した[28]．また，CMU では，マルチモーダル・インタフェースとして，SPHINX をベースにした応用

システムの検討も行っている[29]．一方，ヨーロッパでは，音声を含むマルチモーダル対話の研究がさかんであり[30),31]，また，EC諸国が参加した音声対話に関する SUNDIAL プロジェクトが行われた[7]．

国内では，ATR で自動翻訳電話プロジェクトが行われ，日，英，独の3カ国語を対象とした国際会議の問合せをタスクとした実験システム“ASURA”が開発された[32]．ASURA は，音声認識，機械翻訳，訳文テキストの伝送，訳文音声合成を行う．ATR 以外の研究機関でも音声対話の実システムが，電総研[6]，NEC[33]，KDD[34]，東芝[35),36] などで開発された．また，音声言語インタフェースの研究として，音声認識，自然言語理解，対話処理の研究が活発になり，新しい研究会が設立されるなどして，従来の音声研究の枠に収まらない研究が推進されている[37)~42]．さらに，音声理解・対話の研究のために重要な対話音声データベースの作成が，電子協を中心に進められている[6]．

3.2 DARPA の音声言語システム

1970年代の ARPA プロジェクトの終了後，1987年に DARPA の音声言語理解プロジェクトが再開された．計算機パワーの増大と強力な HMM の利用を背景に，対話音声データベースの整備を行い，自由な発話を認識対象として，CMU，BBN，SRI，MIT[42),43] などで，ユーザにとって制約の少ない音声理解システムの研究開発が進められている．また，ATIS (Air Travel Information System) を共通の応用タスクとしたプロジェクトについては，人とコンピュータとの対話を模擬して対話データベースを構築するとともに，システム開発が精力的に行われている．

BBN では，**図-3** の HARC (Hear And Respond to Continuous speech) 音声対話システムの研究を進めている[25]．HARC は，音声処理システムとして，HMM をベースとした BYBLOS 連続音声認識システムを用い，言語処理システムとして DELPHI システムを用いている．BYBLOS は，入力音声を単語 Bigram モデルに基づいて Forward-Backward サーチで N-best の候補を出力する．DELPHI は，N 個の文章の候補を再評価して，動詞のカテゴリ分けを主体とした Mapping Unit と呼ぶ定式化に基づいて解釈し，データベース検索言語 SQL に変換する．BBN では，言語処理では Mapping Unit の動詞から名詞や形容詞への拡張と，語順の利用，部分的パージングの導入なども検討している．

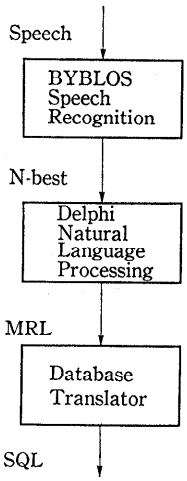


図-3 BBN の音声理解システム HARC の処理の流れ

CMU では，自由発話を対象とした PHOENIX 音声対話システムの研究を行っている[23]．ATIS をタスクとし，音声認識部では，CMU の SPHINX システムを用いて入力音声を意味フレームのスロットごとに認識し，情報検索言語 SQL に変換する．このシステムでは，雑音や環境雑音などに対するガーベジ HMM モデルを採用し，ユーザの自由発話に対処している．また，フレーズごとに文法を用意し，これらの文法を状態ネットワークで表現して解析を行っている．さらに，未知語への対処として音韻の確率的 Bigram モデルを利用している．

3.3 音声自由対話システム TOSBURG II

筆者らは，不特定のユーザに対して「なんら制約を設けない」というコンセプトの下で身近なタスクを選定し，音声対話システム TOSBURG (Task-Oriented dialogue System Based on speech Understanding and Response Generation) を試作し[35]，次に，より自由な音声自由対話システム TOSBURG II を開発した[36]．**図-4** に示すように TOSBURG II は，キーワード検出部，自由発話理解部，ユーザ主導型対話処理部，マルチモーダル応答生成部，音声応答キャンセル部からなり，実時間で動作する．

TOSBURG II は 49 語のキーワードによる自由

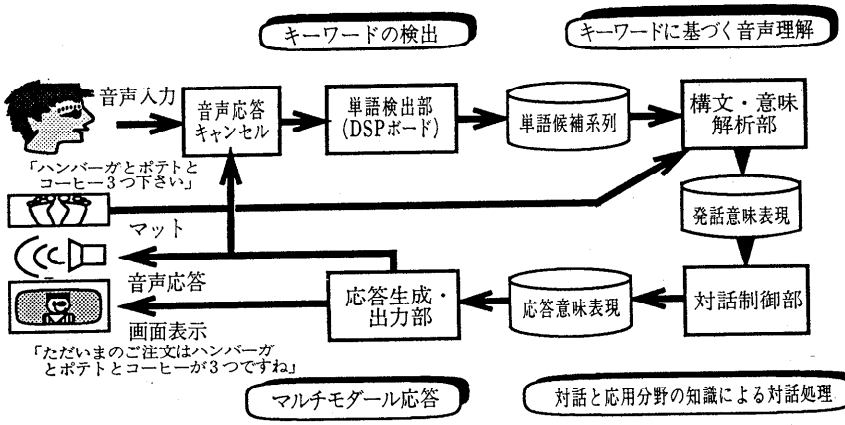


図-4　音声自由対話システム TOSBURG II のシステム構成

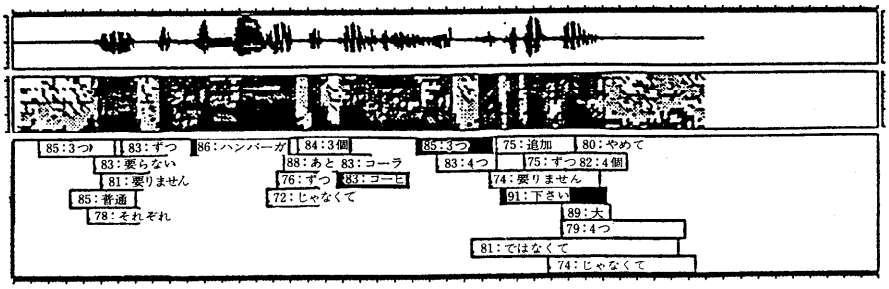


図-5　キーワードラティスの例

発話理解をベースにしている．連続音声中のキーワード検出は，雑音免疫学習[44]により雑音や不要語に対するロバスト性を高めた認識辞書と単語特徴ベクトルの間の始終端フリーな連続的なパターン照合により行う．**図-5** にキーワードラティス(候補系列)の例を示す．

構文意味解析部では，キーワードが検出されるごとにパーザが駆動され，キーワードラティスを始終端フリーに解析して意味表現を求め，複数の発話意味表現候補を対話処理部にわたす．構文意味解析は，文始端判定，文候補解析，文終端判定からなり，解析途中の文候補や部分文候補を保持しながら，拡張 LR 法を基本とし行われる．意味情報はフレーム形式で表現し，拡張項にフレームやスロットの生成などの意味解析手続きを記述し，解析と同時に意味表現を作成する．**図-6** は自由発話音声の解析例である．“ハンバーガ”，“下さい”などのキーワードから，“注文”などの行為を表すアクト，注文品，個数などの対話に必要な意味内容を得る．また，対話中の省略表現や不要語にも対応し，不明な点や曖昧な点は対話処理で補う．

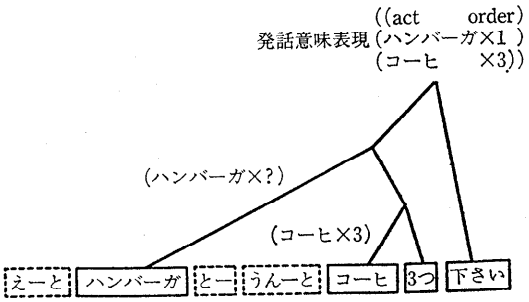


図-6　キーワード系列より得られた解析木と意味表現

対話システムの多くは，システム主導型の「穴埋め形式」の対話処理を採用しており，ユーザは型にはまった発話を強いられている．試作システムはユーザ主導型対話を目指し，対話中のさまざまな状況下で，システム応答に対してユーザの多様な発話を許し，対話の履歴や状況を考慮して省略表現された発話理解などを行う．また，状況に応じて適切な応答を生成し，ユーザが安心して対話できるように設計した．

図-7 に示すように本システムでは，対話処理の流れを，ユーザの発話を理解するユーザ状態と，タスクを管理して応答を生成するシステム状態と

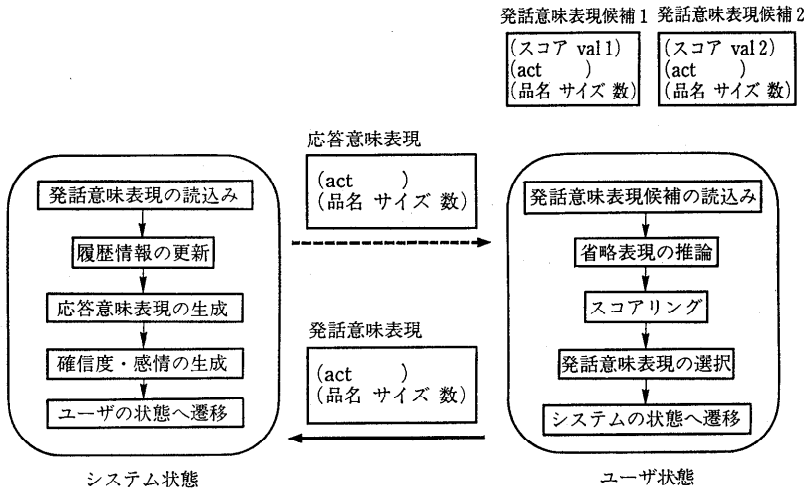


図-7 対話処理における音声理解と応答生成

に分け，モデル化し，ATN によりインプリメントした．このモデルの採用により，ユーザとシステムの状態変化をタスクの達成手順とは独立に，対話の進行状況に即した記述で表すことができ，ユーザ主導の多様な対話の流れに対応できた．図-7 に示すユーザ状態では，音声理解部からの複数の意味表現候補に対してユーザ状態に依存した解析を行い，直前のシステムの応答や対話の履歴情報との意味的な整合性を調べ，ユーザの発話の内容を理解する．たとえば，自由発話ではサイズなどの省略が頻繁に起こるので，直前のシステムの応答を参照して省略された品名，サイズ，数を補う．直前のシステムの応答に適切なものがない場合は，デフォルト値を設定する．図-7 の例のように，直前のシステムの応答内容と入力意味表現候補の内容が適合しない場合には，その候補のスコアを下げるなどの処理を行い，最も高スコアの入力意味表現候補を選択する．一方，システム状態では，対話音声の理解結果にしたがって対話の履歴情報を更新し，応答意味表現を生成する．入力意味表現は音声認識・理解の曖昧性を含むので，本システムでは，ユーザにシステムの理解が曖昧な点を確認するための応答を出力し対話を進める．また，この応答意味表現を，応答生成・出力部に出力するとともに，次からのユーザ発話の内容理解に用いる．さらに，本システムでは，発話理解の確信度や対話の進行状況などのシステムの内部状態を表す感情情報も応答生成・出力部に出力する．

応答生成部では，応答意味表現とシステムの内部状態から，合成音声とおよび **図-8** に示す応答文テキスト，店員の姿，注文品目と個数の視覚表示からなるマルチモーダル応答を出力する．このとき，店員の表情を対話状況に応じて変えるとともに，口の動きの開始タイミング・時間を合成音声に合わせている[36]．音声応答出力は，韻律（イントネーション）を制御して強調点を伝える．たとえば，ユーザが追加注文の際に省略した個数を，システムが推論して音声理解した場合，確認のために個数を強調するようにして合成音声を出力する．このようにマルチモーダル応答により，合成音声が対話状況に即したメッセージを聴覚的に，同時に応答文テキストを視覚的に提示することにより，合成音声の品質が不十分な場合でもユーザのシステム応答の理解を助け，また，音声メディアの一過性という欠点を補うことができる．

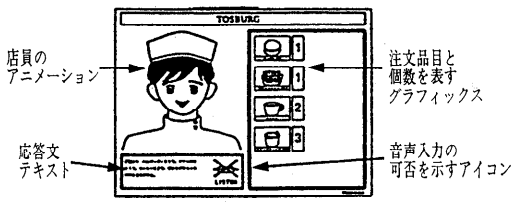


図-8 マルチモーダル応答の画面表示

TOSBURG I では図-8 に示すような唇のアイコンを表示し，音声応答中はユーザの音声入力を受けつけないように設計した．しかし，人間同士の対話ではしばしば相手の応答を遮って発声しており，より制約の少ない自由対話 (Spontaneous

Interaction) を実現するために，TOSBURG II では，図-4 に示すように音声応答キャンセル機能を組み込み，システムからの音声応答を遮って常に音声入力を可能とした[36].

4. 音声対話システム実現の課題

音声対話システム実現のための課題を以下に列挙し検討を加える.

● ロバストな音声認識の必要性　音声認識の曖昧性解消や音声理解に自然言語処理は必須であるが，低次の音声認識性能が水準以下の場合，高度な知的自然言語処理を駆使しても十分な性能は実現できない．耐雑音性を重視した低次音声認識ロバストな自然言語処理の融合が重要である[44),45)].

● 意図と感情　人間同士の対話では表情や非言語音声などを利用している．計算機との自然な対話実現のためには，人工的メディアの観点から音声言語を見直し，積極的に意図や感情のモデルを導入すべきであると考えられる[22)].

● ユーザ適応　音声認識では，認識エラー削減のために（話者）適応を行う場合が多い．ユーザが使い込むにつれて性能が向上するユーザ適応機能は望ましいが，未熟なシステムの学習はユーザが予測不能な状況を招くので好ましくない．ユーザのシステムへの適応とシステムのユーザ適応能力とのバランスを考慮して，インタフェースを設計する必要がある.

● メディア・フュージョン　音声および自然言語と，図形，画像，ジェスチャなどを併用したマルチモーダルインタフェースにより，対話性能の向上が可能である．また，知識メディアである EDR 辞書や Cyc のような大規模知識ベースは，知的対話システムの構築に有用である[46),47)]．種々のメディアおよび知的機能の統合と超並列音声言語処理の研究が，自然な対話実現の鍵となる[48),49)].

● 実システムによる性能評価とデータ収集　システムのヒューマンファクタを評価するためには十分な性能とリアルタイム処理が不可欠である．実システム上にデータ収集と評価支援ツールを開発し，性能評価は HI 研究者が行うべきである．また，各国で対話音声データが蓄積されつつあるが，実システムで収集した大量の音声データの利用により，Real-World システムの開発が可能となる.

● コラボレーション　音声などのアナログ信号やパターンを対象とするパターン認識と自然言語などの記号を対象とする AI コミュニティの間のギャップは大きい．前者は，統計的パターン認識などによる表層的メディア変換の高精度化の研究に注力し，後者は，厳密性を重視して限定された (well-defined) 世界で上位の研究を行っている．音声対話システムの広範な応用に向けて，音声と言語などの各メディア研究者と，AI, HI, OS などの研究者との一層の研究交流が必須である.

5. おわりに

計算機は種々の入出力センサやデバイスを取り込み，マルチメディア・ワークステーションへと進化している．Alan Kay は，次世代ワークステーションのインタフェースは，“ASK and TELL” であると予想している[9)]．その中で，知識や情報の表現メディアとしての自然言語と，計算機との対話メディアとしての音声が，情報処理技術の中心的役割を演じることは確実である[46)]．自然で快適なヒューマンインタフェースの実現に向けて，本稿で述べた自然言語処理と音声処理技術が融合し，言語理解，言語生成，文脈理解，対話処理の研究がさらに活性化することを望みたい.

なお，音声認識における言語処理に関する最近の解説論文のいくつかを参考文献としてあげてあるので参照されたい[1),2),50),51)].

最後に，本稿の執筆にあたり有益なご助言をいただいた(株)リコーの澤井秀文氏と，日頃から音声および自然言語処理について討論いただく坪井宏之，金沢博史をはじめとする東芝研究開発センターの諸氏に深謝いたします.

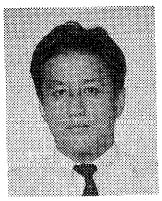
参考文献

1) 岡田美智男：音声言語システムの研究動向と今後の課題，音学誌，Vol. 48, No. 1, pp. 33-38 (1992).
2) 日本情報処理開発協会編：音声の知的処理に関する調査研究報告書，システム技術開発調査研究 3-R-2, 機械システム振興協会 (1992-3).
3) Lee, K. F.: Automatic Speech Recognition, Kluwer Academic Publishers (1989).
4) Goodine, D. et al.: Evaluating Interactive Spoken Language Systems, ICSLP '92, pp. 201-204 (1992).
5) 小林哲則他：ネットワークモデルによる会話音声理解における焦点の表現法，信学技報，SP 85-15 (1985).

6) 速水 悟他：音声対話システムの構築とそれを用いた会話音声収集，信学技報，SP 91-101 (1991).
7) Peckham, J.: Speech Understanding and Dialogue Over The Telephone: An Overview of Progress in the SUNDIAL Project, EUROSPEECH '91, pp. 1469-1472 (1991).
8) 竹林洋一，永田仁史，貞本洋一：ワークステーション上でのソフトウェアによる実時間音声認識機能の実現，HI シンポジウム，pp. 485-490 (1991).
9) Mclaughlin, F.: ICSE 11 Prompts Engineers to Reflect on Yesterday, Look to Tomorrow, IEEE COMPUTER, pp. 110-112 (1989-7).
10) Stern, R. M. et al.: Sentence Parsing with Weak Grammatical Constraints, ICASSP '87, pp. 380-383 (1987).
11) 山田智一他：かな・漢字連鎖統計モデルを用いた日本語 Dictation システム，信学技報，SP 92-32 (1992).
12) Tomita, M.: An Efficient Word Lattice Parsing Algorithm for Continuous Speech Recognition, ICASSP '86, pp. 1569-1572 (1986).
13) Kay, M.: Algorithm Schemata and Data Structures in Syntactic Processing, Tech. Rep. CSL-80-12, Xerox PARC (1980).
14) Ney, H.: Dynamic Programming Speech Recognition Using a Context-Free Grammar, ICASSP '87, pp. 69-72 (1987).
15) Schwartz, R. et al.: The N-best Algorithm: An Efficient and Exact Procedure for Finding the N Most Likely Sentence Hypotheses, ICASSP '90, pp. 81-84 (1990).
16) 松本真治他：語彙・構文・意味制約を統合した A* 探索による会話音声認識，信学技報，SP 91-93 (1991).
17) 北 研二他：音声認識における言語モデルについて，人工知能学会研究会資料，SIG-SLUD-9201-9 (1992).
18) 山本幹雄他：音声対話文における助詞落ち・倒置の分析とその解析手法，「自然言語処理の新しい応用」シンポジウム，pp. 86-93 (1992).
19) Ward, W.: Understanding Spontaneous Speech: The Phoenix System, ICASSP '91, pp. 365-367 (1991).
20) 伊藤克亘他：連続音声認識システム niNja への未知語処理の導入，音講論，pp. 115-116 (1992-3).
21) Grosz, B. and Hirschberg, J.: Some Intonational Characteristics of Discourse Structure, ICSLP '92, pp. 429-432 (1992).
22) 竹林洋一他：計算機との対話のための非言語音声の認識，HI シンポジウム，pp. 123-128 (1991).
23) Ward, W.: The CMU Air Travel Information Service: Understanding Spontaneous Speech, DARPA Speech and Natural Language Workshop, pp. 127-129 (1990).
24) Das, S. et al.: Influence of Background Noise and Microphone on the Performance of the IBM Tangora Speech Recognition System, ICASSP '93, pp. II-71-II-74 (1993).
25) Bates, M. et al.: The BBN/Harc Spoken Language Understanding System, ICASSP '93, pp. II-111-II-114 (1993).
26) Murveit, H. et al.: Speech Recognition in SRI's Resource Management and ATIS Systems, DARPA Speech and Natural Language Workshop, pp. 94-100 (1991).
27) Seneff, S. et al.: Development and Preliminary Evaluation of the MIT ATIS System, DARPA Speech and Natural Language Workshop, pp. 88-93 (1991).
28) Zue, V. et al.: The VOYAGER Speech Understanding System: Preliminary Development and Evaluation, ICASSP '90, pp. 73-76 (1990).
29) Rudnicky, A. I. et al.: Spoken Language Recognition in an office Management Domain, ICASSP '91, pp. 829-832 (1991).
30) Gaiffe, B., et al.: Reference in a Multimodal Dialogue: Towards a Unified Processing, EUROSPEECH '91, pp. 1481-1486 (1991).
31) Arndt, B.: Adoption of Verbal and Visual Dialogue Behaviour in Document Handling Systems, EUROSPEECH '91, pp. 1491-1494 (1991).
32) 嵯峨山茂樹他：自動翻訳電話実験システム ASURA の概要，音講論，pp. 83-84 (1993-3).
33) 渡辺隆夫他：自動通訳のための不特定話者連続音声認識システム，信学技報，SP 91-115 (1992).
34) 武田一哉他：連続音声認識に基づく内線番号案内システムの試作，音講論，pp. 79-80 (1993-3).
35) 竹林洋一，坪井宏之，貞本洋一，橋本秀樹，新地秀昭：不特定ユーザを対象とした音声対話システムの試作，人工知能学会研究会資料，SIG-SLUD-9201-4 (1992).
36) 竹林洋一，永田仁史，瀬戸重宣，新地秀昭，橋本秀樹：音声自由対話システム TOSBURG II—マルチモーダル応答と音声応答キャンセルの利用—，情報処理学会 HI 研資料，HI-45-13 (1992).
37) 飯田 仁他：4階層プラン認識モデルを使った対話の理解，情報処理学会論文誌，Vol. 31, No. 6, pp. 810-821 (June. 1990).
38) 中川聖一他：構文解析駆動型日本語連続音声認識システム SPOJUS-SYNO，信学論 (D-II), Vol. J 72-D-II, No. 8, pp. 1276-1283 (1989-8).
39) Kobayashi, Y. et al.: SUSKIT-II—A Speech Understanding System Based on Robust Phone Spotting, 信学論 (E), Vol. E 74, No. 7, pp. 1863-1869 (1991).
40) 南 泰浩他：自由発声音声認識における意味を考慮した2段 LR パーザの検討，音講論，pp. 69-70 (1993-3).
41) 岡 隆一：部分整合法の出力へのベクトル連続 DP適用による文スポッティング型連続音声認識，信学論 (D-II), Vol. J 76-D-II, No. 5, pp. 921-931 (1993).
42) Butzberger, J. et al.: Spontaneous Speech Effects in Large Vocabulary Speech Recognition Applications, DARPA Speech and Natural Language Workshop, pp. 339-343 (1992).
43) Pieraccini, B. et al.: Progress Report on the Chronus System: ATIS Benchmark Results,

DARPA Speech and Natural Language Workshop, pp. 67-71 (1992).
44) 竹林洋一，金沢博史：ワードスポッティングによる音声認識における雑音免疫学習，信学論 (D-II), Vol. J74-D-II, No. 2, pp. 121-129 (1991).
45) 松本裕治：頑健な自然言語処理へのアプローチ，情報処理，Vol. 29, No. 7, pp. 757-767 (1988).
46) 横井俊夫：日本語の情報化—その技術をめぐって—12章，共立出版 (1990).
47) Lenat, D. and Guha, R.: Building Large Knowledge-Based Systems—Representation and Interface in the Cyc Project—, Addison Wesley (1990).
48) Minsky, M.: The Society of Mind, Simon and Schuster (1985).
49) 北野宏明：超並列人工知能と音声認識システム，人工知能学会研究会資料，SIG-SLUD-9201-8 (1992).
50) 好田正紀：音声認識における言語処理，人工知能学会誌，Vol. 3, No. 4, pp. 424-430 (1988).
51) 坪井宏之，浮田輝彦：音声認識・理解研究の動向，情報処理，Vol. 29, No. 1, pp. 30-41 (1988).

(平成5年6月1日受付)



竹林 洋一（正会員）

1974年慶應義塾大学工学部卒業．1980年東北大学情報工学博士課程修了．工学博士．現在，(株)東芝研究開発センター情報・通信システム研究所主任研究員．1985～1987年MITメディア研究所客員研究員．1992年よりEDR第五研究室室長．1992年人工知能学会全国大会優秀論文賞，1993年音響学会技術開発賞受賞．音声認識，知的インタフェースの研究に従事．電子情報通信学会など会員．