

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 -

竹林 洋一[†] 永田仁史[†] 瀬戸重宣[†] 新地秀昭^{††} 橋本秀樹^{††}[†]株式会社 東芝 研究開発センター 情報通信システム研究所^{††}東芝ソフトウェアエンジニアリング株式会社

本報告では、システムからの音声応答を遮ってユーザから入力可能な音声自由対話システム TOSBURG II について述べる。先に試作した音声対話システム TOSBURG では、キーワードを基本に連続音声を理解し、音声、テキスト、人物像等のグラフィック情報からなるマルチモーダル応答を生成出力することにより、話し言葉による自由な対話を実現した。TOSBURG の評価を行ったところ、ユーザはシステムからの視覚的な応答からその内容を理解できるため、音声応答の終了前に発話する傾向が観察されたが、音声応答中は音声入力を受けつけられないという問題があった。TOSBURG II では、騒音の能動制御技術を用いてマイクロフォンからの入力信号から音声応答成分をキャンセルすることにより、ユーザからの割り込み入力を可能とし、より自由な計算機との対話を実現した。試作システムを用いて評価実験を行い、マルチメディア応答と音声応答キャンセルを利用した音声自由対話システムの有効性を確認した。

SPONTANEOUS SPEECH DIALOGUE SYSTEM TOSBURG II
USING MULTIMODAL RESPONSE AND SPEECH RESPONSE
CANCELLATIONYoichi Takebayashi[†] Yoshifumi Nagata[†] Shigenobu Seto[†]Hideaki Shinchi^{††} Hideki Hashimoto^{††}[†]Toshiba Corporation, Research and Development Center,
Communication and Information Systems Laboratory^{††}Toshiba Software Engineering Ltd.

1 Komukai-Toshiba-cho, Saiwai-ku, Kawasaki 210, Japan

We have developed the TOSBURG II speech dialogue system which enables spontaneous human-computer interaction for unspecified users. TOSBURG II is a development of an earlier speech dialogue system, TOSBURG, which understands spontaneous speech based on keywords and generates multimodal responses. Our evaluation of the TOSBURG system showed that the user often speaks as soon as the system outputs multimodal response. Conventional speech dialogue systems, however, are not designed to accept user's interruption during system response since the mixture of spoken inputs and system responses generally leads to a decrease in recognition accuracy. TOSBURG II enables spontaneous interaction, where the user can interrupt during system response by means of synthetic speech cancellation using active noise control technology. Experimental results have shown the effectiveness of the new system.

1. まえがき

情報処理の分野では、音声、テキスト、グラフィックス等の種々のメディアを利用した人目を引くデモンストラシステムに接する機会が増え、マルチメディアを用いたヒューマンインタフェース(HI)への期待が高まってきている。しかし、これらのマルチメディアシステムの大半は、知的機能を有するわけではなく、単に複数のメディアを入出力しているに過ぎない。

パーソナルコンピュータやワークステーションとの対話(Human-Computer Interaction)に関しては、従来の文字とキーボードから、ビットマップディスプレイ、マルチウィンドウ、マウス、アイコンを用いた対話へと移行し、さらに、音声入出力、3Dグラフィックス、データグローブ等の新しい入出力デバイスの利用[1][2]へと、表現能力と対話の自由度を増す方向で研究開発が進められている[3][4]。しかし、これらのマルチメディアを用いたHIは高性能である反面、逆にユーザにとって複雑さを増し、使い勝手を劣化させる場合が多く、ユーザフレンドリーなインターフェースの構築には至っていない。

一方、音声メディアに関しては、計算機へのテキスト入力の高精度化を目的として音声認識の研究が進められ、不特定話者大語彙連続音声に対しても、実験室レベルで高い認識性能が得られるようになってきた[5]。しかし、これらのHMM(Hidden Markov Model)をベースとしたシステムでは、限定した環境で高い認識精度が得られても、実際の音声対話の応用場面でのロバスト性は不十分であり、環境騒音や自由発話(Spontaneous Speech)への対処が問題となっている[6, 7, 8, 9]。

これに対して、筆者等は音声認識、音声合成を単なる要素技術としてではなく、実世界(real-world)のHIとして位置付け研究を行っており、先に、音声認識、音声言語理解、対話制御、音声応答を統合して、音声対話システム TOSBURG(Task-Oriented dialogue System Based on speech Understanding and Response Generation)を試作した[10]。音声の自然性を活かすために、TOSBURGは発話の制約が少ない自由発話理解を基本とし、日常の会話に頻繁にでてくる言い淀み、省略、語順の変動に対処することにした。また、試作システムを使用して、実環境の対話音声データの収集を行い、使い勝手を評価することができるように、実時間で音声理解とマルチメディアによる応答処理を実現した[11]。

以下、本文では、音声対話におけるマルチモーダル応答の利用に関する検討と、TOSBURGを拡張して試作した音声応答を遮って入力可能な音声自由対話システム TOSBURG II について述べる。

2. 音声自由対話へのアプローチ

2.1 メディアの理解と生成

音声や画像などのマルチメディアを用いた計算機との高度なインタラクションを実現するためには、図1(a)のような単なる入出力機能ではなく、メディアの理解と生成の機能が必要である[12][13]。図1(b)に示すように、メディアの理解は、情報量(information capacity)の多いパターン空間から情報量の少ない計算機の内部表現への写像としてとらえることができる。この情報圧縮の過程で認識エラーや曖昧性(Ambiguity)が生じるわけであり、対話システム構築のためには、エラーの削減に加えて曖昧性への対処が重要となる。また現在の音声対話システムでは単一の音声メディアのみを扱う場合が多いが、対話の効率と自然性を増すために、マルチメディア(マルチモーダル)の利用を考慮する必要がある。

一方、メディアの生成は、計算機の内部表現からパターン空間への一対多の ill-formed な写像と見なすことができる。すなわち、生成可能な出力パターンが多数存在し、ユーザに提示する応答パターンは複雑、多岐にわたるので、メディアの生成プロセスでは、理解プロセスとは別種の曖昧性が生じる。このため、ヒューマン・ファクタを考慮した複数のメディアを効果的に用いたシステム応答の提示、すなわちマルチモーダル応答の設計が重要である[14]。

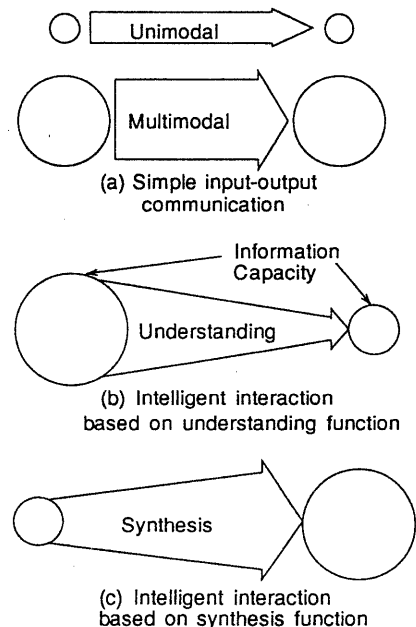


図1. メディアの理解と生成の概念図

2.2 音声自由対話システムの試作

不特定話者大語彙連続音声認識の研究成果を音声対話システムに適用する場合には、実世界の応用で重要なリアルタイム処理や耐雑音性が必要であり、発話の文型や形式に関する「制約のない自由発話」を理解することが望ましい[15][16]。自由発話(Spontaneous speech)を対象とするシステムでは、環境の雑音[17]、不要語、言い直し、省略、ポーズ、対象外の単語などを含む多様な話し言葉に対処する必要があるが[18]、このような現象を文法として完全に記述することは困難である。さらに、認識処理により生じる曖昧性や誤認識などに対しても、ユーザに不安感を与えずに、対話をスムーズに進行する必要があり、また、リアルタイム処理や音声以外のメディアの利用も必要である[12]。

上記の点を考慮し、筆者等は不特定のユーザに対して何ら制約を設けずに、通常の話し言葉で計算機と自然に対話できるシステムの構築を目指した。現状の技術レベルの制約下で自然な対話を実現するため、日常馴染みの深いハンバーガーショップでの注文をタスクとして選択した。この様なシステムの実現のためには、自由発話を対象とした音声理解の他に、対話用の音声合成、対話制御のための知識処理、視覚メディアの利用など、種々の知識やメディアの融合を試みた。また、HIの観点からリアルタイム処理が必須であるため、DSPを用いたアクセラレータも開発し[19]、音声対話システムTOSBURGを試作した。さらに、試作システムを用いてHIの評価と使い勝手の向上を検討し、図2に示すユーザ主体の音声自由対話システムTOSBURG IIを新しく開発した。

3. 音声自由対話システム TOSBURG II

3.1 TOSBURG II のシステム構成

TOSBURG II は、既開発のTOSBURGに音声応答

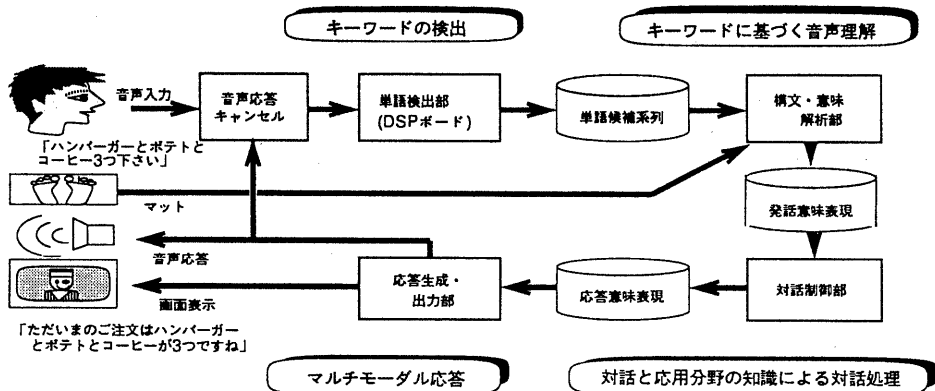


図2. 音声自由対話システム TOSBURG II のシステム構成

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音声入力可能な実時間音声対話システムである。システムは音声応答キャンセル部、キーワード検出部、音声理解部、対話制御部、応答生成部の5つのサブシステムより構成されている。

単語検出部は、不特定のユーザが発声した連続音声の中から、発話の内容理解に必要な49語のキーワードを検出する。キーワード検出は雑音や不要語に強い雑音免疫ワードスポッティング法[17]により、アクセラレータを利用して実時間で処理している。

構文・意味解析部は、検出された時間離散的なキーワードからユーザの発話の内容や意図を理解する。キーワードに基づくため詳細な音声認識は行わないが、対話中の不要語にも対応でき、不明な点や曖昧な点はユーザとの対話の中で明らかにすることとした。キーワード系列の構文意味解析は始終端非固定の拡張LRパーザにより行い、解析結果を発話意味表現として出力する。

対話制御部は、発話意味表現と対話の状況をもとに、対話の履歴を考慮して、省略表現などに対処し、応答意味表現を出力する。従来の計算機主導型の「穴埋め形式」ではなく、発話の制約を極力少なくし、不特定ユーザが自由に発話できるユーザ主導型の形式で対話を進行する。

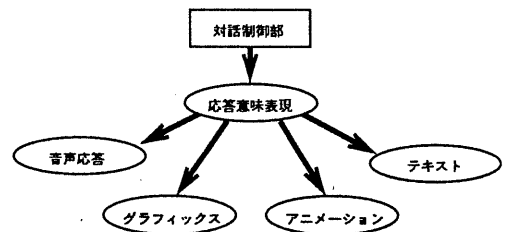


図3. 意味表現からの応答生成

応答生成部は、応答意味表現とシステムの内部状態をもとに、図3に示すように音声メディアや視覚的メディアを利用したマルチメディア応答を生成する。音声応答では、韻律を制御して強調した表現を行うことにより、対話の中で不明確な点を確認する。

TOSBURG II で追加した音声応答キャンセル部は、スピーカ-マイクロフォン間の伝達関数を逐次推定し、適応フィルタを用いてマイクロフォン信号中から音声応答成分を常時引き去り、単語検出部に音声応答のキャンセルされた信号を送る。これにより、システムからの音声応答中でも入力が可能となる。

以上述べたように、本システムでは、自由発話を一字一句認識するのではなく、キーワードに基づいて発話の内容や意図を理解し、対話の状況と対話の履歴情報を用い、また、音声応答キャンセルによりユーザー主導型で対話を進行する。次にマルチメディア応答、応答キャンセル、音声自由対話について詳しく述べる。

3.2 マルチモーダル応答生成

システムとの対話を円滑に進めるためには、計算機の内部状態や対話の状況を把握しやすいようにユーザーに提示する必要があり、システムからの応答としてマルチメディアをいかに利用するかが重要である。聴覚的メディアは情報を送るタイミングや対話進行のキーとなる一方、視覚的メディアは複数の情報をパラレルに送り多くの情報を一度に伝えることができるという特徴を持つ。

TOSBURGの応答生成部は、合成音声を用いた聴覚的メディアと、応答文テキスト、アニメーション、アイコンなどの視覚的メディアを用い、対話制御部から送られる応答意味表現からマルチモーダル応答を生成する。音声応答は、ターミナルアナログ型ホルマント合成器を用いた規則合成により生成する。応答意味表現により表現された応答内容に合った応答文の標準形を決め、状況に応じて省略表現を行い、生成された応答文に対して音韻処理などを行い合成音声を生成する(図4)。合成音声の韻律は、藤崎モデル[20]を用いて

与え、強調すべき情報がある場合は、対応する位置のイントネーションの変化を大きくする。応答文テキストは、音声応答と同じ応答文のテキストを視覚的に表示する。音声メディアには一過性という欠点があるが、応答文テキストはこの欠点を補うことができる。

注文品目とその個数および店員の姿を表わすアニメーションを提示し、単純明快でわかりやすい応答をユーザーに出力する(図5参照)。また、店員の表情は過

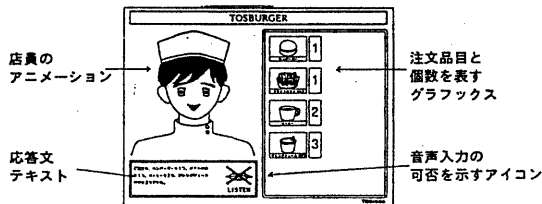


図5. 視覚メディアによる応答

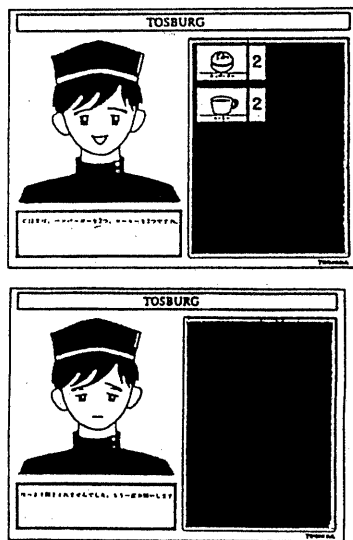


図6. 表情のあるインタフェース

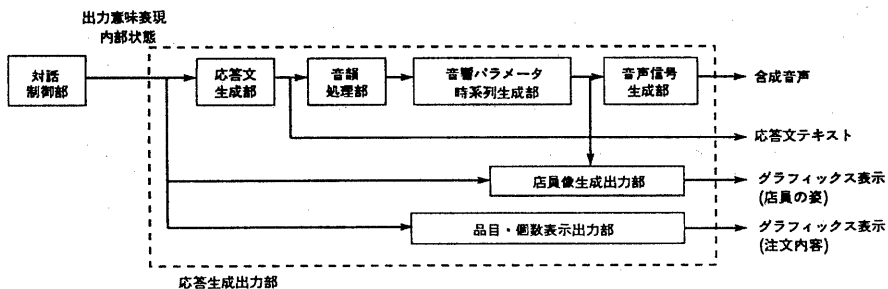


図4. 応答生成出力部の構成

度にリアルにせず設計した。注文品目とその個数を表わすグラフィックスは、システムの内部状態を視覚的に示す。店員の姿のアニメーションでは、音声応答にあわせて口を動かすことにより、動いている実感のある目標(attention getter)を具体的に示し、ユーザが自然に計算機へ音声入力できる雰囲気を作るとともに、店員というコミュニケーション対象を明示することにより、特に意識することなく自然な大きさの声で入力できる。さらに、店員の表情は、対話の状況に合わせて、微笑んだ顔、申し訳なさそうな顔などにかえ、システムからの応答の理解を助けている(図6参照)。

3.3 音声応答キャンセル

TOSBURGは自由な話し言葉による音声対話を可能としたが、音声応答が入力音声に混入し誤認識をもたらすことがあるため、システムの音声応答中は図5に示すような唇のアイコンを用い、ユーザの音声入力を受けつけないよう設計した。しかし、日常の自然な対話では、相手の応答を遮って発声し、テンポの良いインタラクションを実現している。筆者等は、ユーザ主導型でより制約の少ない自由対話(Spontaneous

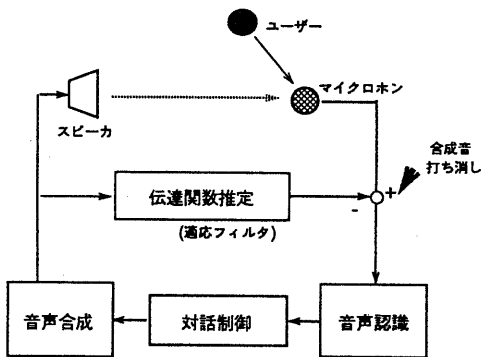


図7. 音声応答キャンセル機能の構成

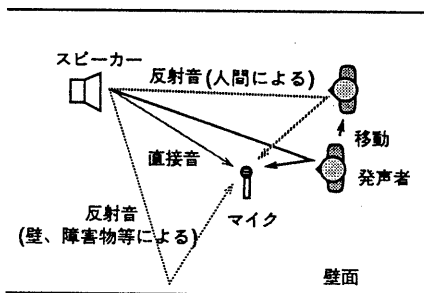


図8. 人の移動による反射音の変化

Interaction)を実現するためには、ユーザからの入力を常に受けつけられるようにすることが望ましいと考え、TOSBURG IIを構築した。

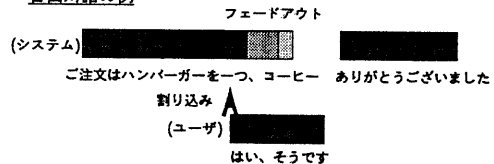
音声応答のキャンセルは、騒音能動制御技術を用い、図7に示すようにマイクロホンに入力された音声入力と音声応答が混合した信号中から音声応答の成分を常に引き去ることにより行う。音声応答のキャンセルでは、合成音声という非定常な帯域の広い信号を扱い、しかも図8のように、スピーカからマイクロホンに届く成分のほかに周囲の物体に反射して届く成分があり、これらを取り除く必要がある。とりわけユーザの身体を反射する成分は身体の動きの影響を受けるため、スピーカからマイクロホンへの伝達特性の推定を、対話中にリアルタイムに行う必要がある。このため、LMSアルゴリズムを音声対話システム用に改良し、ユーザの動きにも追従し騒音下でも安定に動作する方式を開発し[21][22]、既開発のDSPアクセラレータにより実時間処理を実現した。

3.4 音声自由対話の実現

上述した音声応答キャンセル機能を組み込んだ音声自由対話システムTOSBURG IIは、図5に示したような唇のアイコンが不要となり、ユーザは表示画面を意識せずに、常時音声応答を遮って入力できるようになった。

図9に自由対話と従来のシステムにおけるタイミングの比較を示す。マルチモーダルインタラクションでは、視覚的なメディアを利用して、音声メディアによるシステムの応答出力が終わるのを待たずに相手の伝えたい情報の内容が理解できるため、その時点で次の入力を行えば、よりテンポの良いインタラクションが実現できる。また、TOSBURG IIでは、ユーザの音声

自由対話の例



従来システムの対話例

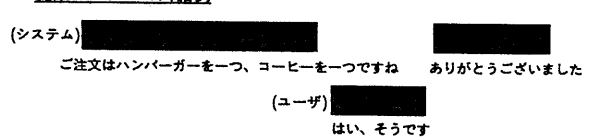


図9. 自由対話と従来システムにおける発声のタイミングの比較

入力に比較的長い間があると、システムは発話の終了と判断して応答を返すが、これは双方向のテンポの良いインタラクションを実現するためのシステムからの割り込みととらえることができる。このように、音声応答のキャンセルにより、システム、ユーザともに相手からの情報を受けている間の割り込み発話の手段を得たことになり、より制約のない自由対話を実現することができる。

4. 評価実験

4.1 マルチモーダル応答の評価実験

上述したように、TOSBURGの各応答メディアは、システムからのメッセージやシステムの内部状態を伝えるなど、それぞれの役割を持っている。応答メディアを制限した場合のインタラクションへの影響を調べマルチモーダル応答の効果を確認するため、以下のような評価実験を行った。

音声応答、テキスト、店員のアニメーション、注文品目とその個数を表わすグラフィクス(以下、トレイと呼ぶ)の各応答メディアのうち、すべてを出力(条件1)、トレイを出力しない(条件2)、トレイとテキストを出力しない(条件3)、音声応答を出力しない(条件4)、の4つの場合について対話実験

を行い、1回の対話に要した時間とユーザの発声回数、対話の間のユーザの反応を観察した。被験者は、本システムを使用した経験のある男性5名で、欠落するメディアを知らせている。また、あらかじめ決めておいた5品目の注文をタスクとした。

図10に実験結果を示す。システムの内部状態を視覚的に示すトレイが欠ける条件2や条件3では、全てを出力する条件1に比べて、対話に要する時間、発声回数とも明らかに増加している。また、条件2の対話中、複数の箇所を訂正する必要がある場合に、被験者は一度に訂正するよりも、インタラクションの回数が増えてでも少しずつ確実に訂正する戦略を選ぶケースが観察された。さらに、条件1や条件2では、システムの音声応答が出力し終わっていないのにユーザが入力を始めようとするケースがよく観察されたが、条件3の対話中では、システムの応答出力が終わるのを待つ傾向があった。これらの結果は、トレイの表示によりシステムの内部状態を視覚的に示すことが、対話の円滑な進行に貢献していることを意味している。特に条件3では、システムの内部状態を知る手がかりは音声応答以外はなく、視覚的なメディアを使った応答がユーザ

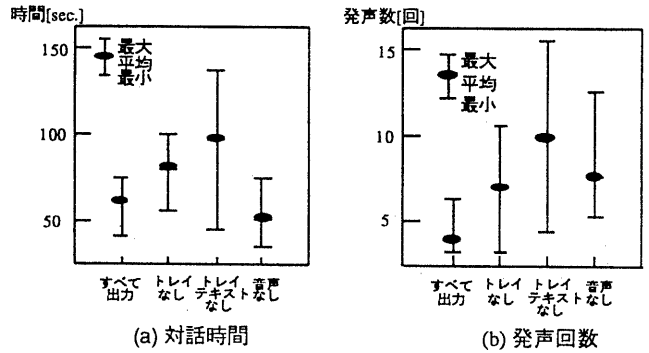


図10. マルチモーダル応答の実験結果

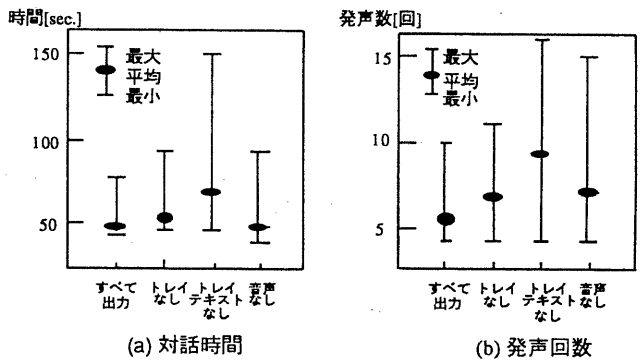


図11. 音声応答キャンセルの実験結果

への負担の少ないインタラクションの実現に重要であることがわかる。

音声応答のない条件4の対話では、システムが入力待の状態になっているにもかかわらず、ユーザがすぐに音声入力をしない場合がみられたが、実際、実験後に入力のタイミングがつかみづらいとの感想を述べた被験者がいた。条件4は、音声応答の終了を待つ必要がないが、条件1に比べ対話時間がさほど短くならない。対話のスピーディな進行のためには、ユーザの次の入力を促す応答の存在が重要であることを示している。

4.2 音声応答キャンセルの評価実験

TOSBURG IIを用いて音声応答キャンセルの効果を評価するため実験を行った。被験者は、本システムを使用した経験のある男性4名で、あらかじめ決めておいた前節と同じ5品目の注文をタスクとし、応答メディアを変えて、対話時間および発声回数を測定した。

実験結果を図11に示す。TOSBURG IIでは単語検出

用辞書および構文意味解析用文法などの改良も行ったため、単純にTOSBURGの結果と比較できないが、対話時間は大幅に短縮され、キャンセルの効果が現れていることがわかる。また、応答メディアの変化に伴う対話時間および発声回数の変化は、TOSBURGの評価実験と同様の傾向を示した。発声回数はTOSBURGの評価実験から減少傾向は見られない。実験後の感想では、音声応答の終了を待つ必要がなく気楽に入力できるとか違和感がないという意見があった。

これらの実験から、音声応答キャンセル機能を組み込むことにより、テンポの良いインタラクションが可能となり、また、ユーザの精神的な負担が軽減されることが明らかとなり、その有効性を確認できた。

5. マルチメディア対話技術の課題

これまでの各章節では計算機との自然な対話実現のため試作した音声自由対話システムTOSBURG IIとその評価実験について論じた。以下、音声などのマルチメディアの理解と生成機能を用いた対話技術に関する研究課題を列挙する。

(1) メディア理解における曖昧性

実際のメディア理解の応用場面では、まず低レベルのパターン認識精度の向上が大切である。筆者等が雑音下の音声認識のために開発した雑音免疫学習法等により、低次の認識性能が向上可能である[17]。低次レベルの処理が貧弱な場合、どんな巧妙な高次の知識処理を行ってもメディア理解の高性能化は困難である。しかし、パターン認識エラーは不可避であり、メディア理解の曖昧性の問題に対処するためには両者のバランス良い統合が必要である。

(2) メディア生成における曖昧性

マルチメディアの生成は、自動生成機能あるいはマルチメディア編集ツールを利用することにより行える。しかし、メディアの自動生成機能により生成可能なパターンは多数存在し、その曖昧性は避けられない。ユーザの目的や好みに合致したメディア生成を実現するためには、自動生成機能により、複数のパターン候補を提示し、ユーザが最終的に選択する提案型インタフェースが今後有望である。さらに、ユーザの好みの学習機能の研究が今後重要視されると考えられる[12]。

(3) 意図と感情

自然な計算機との対話を実現するためには、意図と感情の伝達が重要である[23,24]。人と人とのコミュニケーションは、表情や非言語音声、ジェスチャー等、様々なチャンネルを利用している。計算機との対話の場合も、新しいメディアを利用し、意図、感情に関する認知モデルを導入することにより、自然性、使

い勝手が向上できる。

(4) メディア・フュージョン

各メディアはそれぞれ長所と短所がある場合が多い。のが普通である。ロバストでフレンドリなマルチメディア・インタフェースを実現するためには、オーケストラの指揮者のように各人工メディアの性質を理解し、各メディアを融合することが必要である。さらに、知識を一種のメディアと見なし、Cyc[25]のような大規模知識ベースや電子化辞書[26]を用いた知的対話システムの構築が期待できる。さらに、Minskyによれば人間の心も数百種類の異なるアーキテクチャのコンピュータから構成されており、種々のメディア生成、理解の他に、種々の知的機能との統合に関する研究が自然な対話実現のために必要である[27]。

(5) システム評価

計算機との人工メディアによる対話システムを評価するためには十分なメディア生成、理解の精度とリアルタイム処理が必要である。上記の条件を満足しなければ、対話システムを開発しても単なるデモ用のtoyに終わってしまう。また、リアルタイムシステム上で、各メディア理解や生成機能の評価はHI研究者が行うべきであり、正確で客観的評価が可能となる。

(6) コラボレーション

伝統的に音声認識等のパターン認識とAI研究者間の交流は少ない。パターン認識の分野では、雑音や変形が多いパターンを処理対象とし、表層的なパターンから記号へのメディア変換を研究対象とし、統計的手法やニューラルネットに基づく低レベル(low-level)の認識精度向上を研究している。一方、AIの分野では、厳密性を重視し、限定された(well-defined)世界の中で、記号ベースにhigh-level(上位)の研究を行っているが、toyモデルの研究が多い。ニューラルネットワークの研究分野も、現状では、パターンレベルの研究が主流であるが、AIや認知科学研究者も参入し、上位レベルの研究も盛んになりつつある。マルチメディアによるコンピュータとの自然な対話の実現にむけて、今後、各メディア、AI、パターン認識、認知科学、ロボティクス、HI、基本ソフトウェア、計算機アーキテクチャー等の研究コミュニティー間の交流を深めることが必須である。

6. むすび

本文では、マルチメディアによる計算機との対話につき検討し、音声メディアの特徴を活かした音声自由対話システムTOSBURG IIの構成とその評価実験について述べた。実時間システムを用いた実験から、ユーザはマルチモーダル応答によって応答内容を迅速に理解することができるため、音声応答の終了を待たずに

次の入力始める傾向があることがわかった。この問題に対処し制約の少ない自由対話を実現するため、応答出力中でもユーザからの音声入力を受け付ける音声応答キャンセル機能を利用した新しい音声自由対話システムTOSBURG IIを開発し、その有効性を明らかにした。今後は、音声応答キャンセル機能を組み込んだシステムを使い、HIの観点から評価を行い、人間にとって使い勝手の良いシステムのあり方について検討していく。

謝辞

本システムの開発に際しご指導頂いた都立航空高専武井欣二教授と、システム試作で協力頂いた坪井宏之、金沢博史、貞本洋一、山下泰樹、新居孝章の諸氏に感謝致します。また、データ収集等に協力頂いた鈴木八重子、今井直子、松浦佐知子、茂木みどり他、東芝研究開発センターの諸氏に感謝致します。

参考文献

- [1] D.Weimer and S.Ganapathy : "Interaction Techniques Using Hand Tracking and Speech Recognition" in "Multimedia Interface Design", Blattner ed., Chap.7, ACM Press Book (1992).
- [2] A.Rudnický and A.Hauptmann : "Multimodal Interaction in Speech Systems" in "Multimedia Interface Design", Chap.10, ACM Press Book (1992).
- [3] M.Tatham : "Intelligent Speech Synthesis as part of an Integrated Speech Synthesis/Automatic Recognition System", in "The Structure of Multimodal Dialogue", M.Taylor et al. eds, North-Holland, pp.301-312 (1989).
- [4] M.Blattner and R.Dannenber : "CHI'90 Workshop on Multimedia and Multimodal Interface Design", SIGCHI Bulletin Vol.22, No.2, pp.54-58 (1990.10).
- [5] K.Lee : "Automatic Speech Recognition", Kluwer Academic Publishers (1989)
- [6] 古井 : "連続音声認識の今後の研究課題", 電子情報通信学会「マルコフモデル・ニューラルネットワークを包含する新しい音声認識手法」時限研究専門委員会資料, SPREC-91-2, p.41 (Feb. 1992).
- [7] 樽松 : "連続音声認識技術の将来的課題", 電子情報通信学会時限研究専門委員会資料, SPREC-91-2, p. 52 (Feb. 1992).
- [8] 小林, 白井 : "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, SP85-15 (1985.6).
- [9] 飯田, 有田 : "4階層プラン認識モデルを使った対話の理解", 情処学論, 31, 6, pp.810-821 (1990.6).
- [10] 竹林, 坪井, 金澤, 貞本, 山下, 永田, 瀬戸, 新居, 橋本, 新地 : "不特定話者音声対話システムTOSBURGの開発", 音講論, 1-P-16, (1992.3).
- [11] 山下, 瀬戸, 坪井, 竹林 : "自由発話による実時間音声対話システムTOSBURGの応答生成", 人工知能学会研究会資料, SIG-HICG-9201-2 (1992.5).
- [12] 竹林 : "マルチメディアによる計算機との対話", 信学技報, SP92-37, (1992.7).
- [13] Y. Takebayashi : "Integration of Understanding and Synthesis Functions for Multimedia Interfaces" in "Multimedia Interface Design", Chap.14, ACM Press Book (1992).
- [14] 竹林, E.Large, 相馬, 土井 : "図形理解に基づく知的プレゼンングラフシステムの構成", 第4回HIシンポジウム, pp.185-190 (1988.11).
- [15] P.Heisterkamp, S.McGlashan and N.Youd : "Dialogue Semantics for an Oral Dialogue System", ICSLP'92, Vol.2, Th.FPM.2.3, pp.643-646 (1992.10).
- [16] J.Allen : "Recognizing Intention from Natural Language Utterances", in "Computational Model of Discourse", M.Brady et al. eds, MIT Press, pp.107-166 (1983).
- [17] 竹林, 金澤 : "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", 信学誌 (D-II), J74-D-II, No.2, pp.121-129 (1991.2).
- [18] 坪井, 竹林, 橋本 : "連続音声理解のためのキーワードラティスの解析", 音講論, 1-5-11 (1991.10).
- [19] 坪井, 金沢, 竹林 : "音声認識研究用リアルタイム処理システムの開発", 信学技報, SP90-37 (1990.8).
- [20] 藤崎, 須藤 : "日本語単語アクセントの基本周波数ボタンとその生成機構のモデル", 音響誌, Vol.27, No.9, pp.445-453 (1971.9).
- [21] D.Gleaves and Y.Nagata : "The Cancellation of Synthetic Speech for a Man-Machine Dialogue System", 音講論, 2-5-5, (1992.12).
- [22] 永田, D.Gleaves, 竹林 : "音声対話システムにおける応答音声のキャンセルについての検討", 音講論, 1-P-21, (1992.3).
- [23] K.Maeda, Y.Yamashita and Y.Takebayashi : "Enhancement of Human-Computer Interaction through the Synthesis of Nonverbal Expressions", Proc. ICSLP'90, 19.15, pp.821-824 (1990.11).
- [24] 竹林, K.Maeda, 金沢 : "計算機との対話のための非言語音声の認識", 第7回HIシンポジウム, pp.123-128 (1991.10).
- [25] D.Lenat and R.Guha : "Building Large Knowledge-Based Systems", Addison-Wesley (1989)
- [26] 横井俊夫 : "自然言語処理の頑健さが知識処理の脆弱さを救う", 日経AI春号, pp.158-171, (1991).
- [27] M.Minsky : "The Society of Mind", Simon and Schuster (1985).