

## 解説

## 音声言語情報処理の現状と研究課題



## 5. GUIからマルチモーダルUI(MUI)に向けて†

新田 恒雄††

## 1. はじめに

「人間にとって音声は最も自然な対話手段である」という言葉を、音声研究者はこれまで繰り返し使用してきた。しかし現実に応用システムを構築すると、ユーザへのプロンプトの与え方に始まり誤認識を生じた場合の対処方法まで、音声チャンネルだけではうまく解決できない数多くの問題に突き当たる。こうした体験を経て、現在、音声研究者は“人間とコンピュータの対話ではどのような形態が自然なのか”という問題設定に立ち返って、音声以外の対話チャンネルを含むインタラクティブな対話システムに取り組み始めている。他方、コンピュータのユーザインタフェース(UI)を研究する立場からも、グラフィカルUI(GUI)の限界を乗り越える新しい枠組みが求められている。特にマルチメディアを対象とする次世代の多様なサービスでは、新しい対話スタイルと共に、コンピュータに慣れない人も利用できるUI環境が強く望まれている。

マルチモーダルUIは、音声言語やノンバーバルインタフェースを介してマルチメディア情報と直接対話することを目指しており、次世代UIの候補として期待される。本稿では、音声入力とコンピュータUIの現状を概観した後、将来のマルチメディア・サービスに求められているUI上の課題を述べる。続いて、マルチモーダルUIの概要と研究課題を解説する。

## 2. 音声入力とユーザインタフェース

## 2.1 音声認識に今何ができるか

音声認識の実用化は1970年代に始まる。当初

は、利用者を限定した数10単語を認識する装置が、物流仕分けや工場の検査といった産業応用に使われた。1980年代に入って装置の小形化が進み、音声ダイヤル電話や家電機器制御などの民生応用も試みられた。また、電話サービス向け音声認識応答システムとして、不特定多数の人の音声(16単語程度)を認識する装置も導入された。しかし、これらの装置は性能面での頑健性不足とUI設計の経験不足から、一部の応用を除いて広く普及するには至らなかった<sup>1)</sup>。

では1990年代に入って音声認識に何ができるようになったのか<sup>2),3)</sup>。大きな進歩は、音声を登録せずに(キーボードから単語を登録したり、アプリケーションごとに使用する単語リストを読み取るだけで)任意の単語を認識できるようになったことと、自由発話(例: えー、都庁はどこです?)の中から単語(“都庁”)を識別できるようになったことであろう。また語彙数が制限されるものの、ソフトウェアのみで音声入力を実現できるようになったこともあげられる。アプリケーション・プログラムインタフェース(API)の整備が進めば、システム開発者が専門家の助けなしに音声入力を様々なシステムに組み込むことができるようになるだろう。

一方、音声認識にはまだまだ課題も多い。音声パターンは、文脈の違い、話者の違い、環境騒音の違いによって大きく変動する。また、認識対象が単語から文、自由発話へと拡大するにつれ、発話タイミングのずれや息継ぎ・息漏れ・舌打ちなど話者自身から出る非言語音の問題、助詞落ち・倒置など話し言葉のゆらぎの問題、さらに言い淀みなど困難も拡大する。表-1は音声認識における1990年代後半の利用形態と制約を展望したものである。音声入力の導入にあたっては頑健性を確認し、またユーザの習熟度を配慮して目的に合

† From GUI towards Multimodal UI (MUI) by Tsuneo NITTA (Multimedia Engineering Laboratory, Toshiba Corporation).

†† (株)東芝マルチメディア技術研究所

表-1 1990年代後半の音声認識と制約の展望

発話モード	語彙数の目安	その他の制約	アプリケーション例
孤立単語	10,000	前後に余計な発声が入らないこと	コマンド入力 データ入力
連続音声	10,000 (右の例の A, X, … で各数 100 程度)	構文を固定した短文 例: A さんに X を …してください。	コマンド+データの 一括入力
自由	1,000	発話中の数単語を抽出 例: えーと、郵便局は？	情報 Kiosk など
<p>〈備考〉</p> <ul style="list-style-type: none"> <li>・音声登録について: 原則, 不特定話者かつ任意語彙を対象に利用可, すなわち, 個々の単語を音声登録する必要はなく, キーボードから登録したり, リストを読み取るだけでよい。</li> <li>・共通の制約: 単語内で言い淀まないこと, 騒音が 70 ホン程度の環境で使用すること, 現在の認識方式では S/N が 6~10 dB 程度より悪化すると, ノイズキャンセラなどの特別な騒音対策が必要になる。</li> </ul>			

った利用形態を選択することが大切である。さらにアプリケーション組込みの際は, 多様な入出力手段を用いて効率的かつ頑健な UI, すなわち後述するマルチモーダル UI を提供することが重要になる。

## 2.2 コンピュータ UI の現在

コンピュータ UI の世界では, キーボードを操作する“テキストコマンドベースの対話”の時代が長く続いた(第一世代の UI)。この対話方法は, ユーザがコンピュータの機能を知っていることが前提になるが, 多くの場合専門家が会計/資料管理といった定型業務を対象に使用していたため, 現在から見ると UI が問題になることは少なかった。

1980年代の初頭に始まったダウンサイジングの流れの中で, ワークステーション (WS) やパーソナルコンピュータ (PC) へとコンピュータの主役が代わり, 用途も WP/電子メール/表計算/DB 利用など不特定の個人向け業務が中心になった。キーボードとマウスを操作する“GUI ベースの対話”は, 画面からある程度機能を読み取れるため, 専門家でなくとも使用できる(第二世代の UI)。GUI はアイコン, プルダウンメニュー, さらに drag&drop などの操作を提供して格段に使い勝手の良い UI 環境をもたらした。

さて 1990 年代も後半に向かい, 現在, マルチメディア処理機能が情報機器に標準装備されつつある。様々なマルチメディア・サービスが提案されるのと並行して, UI についても新しいサービ

スへの対応が要請されている。情報機器に関わる道具の歩みを見ると, 紙&ペンそして印刷メディアの長い歴史の後, 音声メディア (電話) と映像メディア (TV) が生活の中に定着し, 多くの人々に受け入れられてきた。コンピュータが過去の道具と異なる点は, 操作対象がデジタルメディアであること, それゆえに外見は 1 つの道具でも様々な機能を持ち, 多彩な道具を演じられることであろう。しかし個々の機能には, それに見合った“使いこなす技術”が必要になる。また, ユーザはビット・ストリームとしてのメディア操作だけでなく, 情報 (コンテンツ) を直接操作することも望んでいる。現在は, 画面上のボタンオブジェクトや, それをクリックすると現れる様々なオブジェクト (表示・音…) の指示に従い, 対象にたどり着く方法が主流である。しかし, 選択肢が  $7 \pm 2$  の範囲を越えると操作効率が低下することから<sup>4)</sup>, 広い範囲の検索が可能な言葉によるアクセスを組み込んだ次世代の UI が要請されている。複数のメディア情報を直接操作する“マルチモーダル UI ベースの対話”は, この第三世代 UI の有力な候補として期待される。

## 2.3 UI に今求められていること

ここでは, 近い将来におけるマルチメディア・サービスの例を通して“使いこなす技術”の課題を概観する。まず, 通信は今後ますます重要なサービスになると考えられる。通信では人と人の対話を中心になる。このため, 多様なメディア (文字, 音, 映像) 情報を活用した新しい会話スタイ

ルを、どのように支援し定着させていくかが大きな課題になる。また携帯型の情報機器では操作性が問題になる。

次に、複合文書の作成では専門家を対象にした現在のツール(Hyper media authoring toolなど)を、いかにして一般の人が使える道具にするかが課題である。マルチメディアデータの構造化とデータ間のリンク作成は専門知識を必要とし、かつこれらは個々人で視点が異なる。このため、一般の人にも使える“賢い”支援ツールの開発は容易ではない。また、作成された複合文書を利用する際のUIは、マルチメディアを対象にした新しい操作メソッドを必要としている。

さらに、街中で様々な情報を提供する端末、いわゆる情報Kioskなどのマルチメディア情報利用では、サービス内容が多岐にわたるため、応用分野(観光案内、地理案内、金融サービスなど)とタスク(情報アクセス、スケジュールリングなど)に応じた対話様式を提供することが望まれる。このためには、タスクのシナリオと対話履歴からユーザの意図を理解するなど、能動的で深いレベルのインタラクションを可能にする“対話管理能力”が必要になる。

一方、コンピュータのユーザからは以下のような要望もある。

- (a) マニュアルを読まずに使える
- (b) 同じアプリケーションは、異機種でもすぐ使える
- (c) ストレスなく使える
- (d) 使うほどにリテラシーも上がる

厚いマニュアルを前にすれば、普通の人なら使う気も萎えるだろう。やりたいことは同じなのに、買い替えたらずしく習わないと使えない機器では身近な道具とは呼べない。また、現代のようにストレスの多い社会では、さらにストレスを増やすようなものは受け入れられない。逆に、使って楽しいといった遊びの要素が大切になる。最後に、難しい操作をとともなう場合には、スキルが向上する枠組みも必要である。

以上に述べたUIへの様々な要請に応えるには、多様でインタラクティブな対話を可能にするマルチモーダルUIの枠組みとその応用技術の研究が必要になる。

### 3. マルチモーダルUI

マルチモーダルはモダリティとモードという2つの概念を受け継いでいる。モダリティ(modality; 知覚的様相)はもともと心理学で使われている用語で、情報が授受されるチャンネル(たとえば、音声・ジェスチャ・)の種類を指し、また情報が表現・知覚される様相までを含んでいる。たとえば音声チャンネルは、言語のほか感情や個人性といった異なるモダリティを含む。したがって、音声単一チャンネルを使用しても、言葉と感情を授受するならモダリティは複数(マルチ)になる。一方、コンピュータUIの分野ではモード(mode)という用語が使われてきた。キーボード入力を例にとると、操作効率を高めるためにコマンドを1打鍵で表現するが、その際多くのコマンドを利用するために、モード指定キーを使用し同じ入力操作でもモードによって異なる意味を付与している。モードは情報を解釈して意味を抽出したり、意味を担わせるためのチャンネルの組合せを指して使用される。以上から分かるように、マルチモーダル・インタフェースあるいはマルチモーダルUI(本稿ではこの言葉を使用し、以下ではMUIと呼ぶ)と言うときには、マルチメディア情報の送り手(ユーザ/コンピュータ)と受け手(コンピュータ/ユーザ)が、複数のモダリティを利用して相手の意図を解釈しながら対話を進めることが期待されている。したがって、コンピュータ側にはユーザの意図を理解し、その結果を反映した新しい対話戦略に基づく応答をユーザ側に提示する仕掛け、すなわちインタラクティブな対話管理機能が必要とされる。

近年、音声/タッチ/ジェスチャ/顔画像など様々な入出力チャンネルを備えたマルチモーダルUIの試作が行われている<sup>5)~7)</sup>。以下では、MUIを多様性、相補性、頑健性といった面から眺めた後、MUIフレームワーク、円滑な対話を実現する際の留意点、およびMUI開発環境について説明する。

#### 3.1 マルチモーダルUIの諸相

MUIには様々な対話形態がある。表-2は対話に使われるメディアとメッセージの組合せの例を、入力モダリティと出力モダリティに分けて示したものである。人間はここに見られる多様な形

表-2 マルチモーダル UI の多様性の例

形態	入力モダリティ (理解)	出力モダリティ (生成)
単一メディア 複数メッセージ	<ul style="list-style-type: none"> <li>ペンでジェスチャと文字を入力</li> <li>音声で言葉と個人性を入力</li> </ul>	<ul style="list-style-type: none"> <li>複数 Window を利用しての表示</li> <li>応答音に言葉と送り手の感情を表現</li> </ul>
マルチメディア 単一メッセージ	<ul style="list-style-type: none"> <li>「このグラフ」と音声で入力し、同時にペンで囲み直指</li> </ul>	<ul style="list-style-type: none"> <li>Text と Graphics での説明</li> <li>画面に Text を表示し、同時に内容を読み上げ</li> </ul>
マルチメディア 複数メッセージ	<ul style="list-style-type: none"> <li>Text (ASCII) に絵 (Digital ink) を付して入力</li> <li>音声とペンおよび Eye tracker による入力作業</li> </ul>	<ul style="list-style-type: none"> <li>旅行案内ビデオを写しながら地図と時刻表を表示</li> <li>予定表を表示しながら、会議通知メールを読み上げ</li> </ul>

表-3 マルチモーダル UI の相補性—ペンと音声の場合

	ペン入力	音声入力
ユーザの拘束	<ul style="list-style-type: none"> <li>パッドの上に構えて操作しなければならず、目と手を拘束</li> </ul>	<ul style="list-style-type: none"> <li>動作は自由 (目と手を拘束せず)</li> <li>会議中や高騒音下では使えない</li> </ul>
入力速度	<ul style="list-style-type: none"> <li>遅い</li> </ul>	<ul style="list-style-type: none"> <li>速い (だれでも)</li> </ul>
記録/編集	<ul style="list-style-type: none"> <li>記録に残り、簡単に編集できる (考えながらの入力に適している)</li> </ul>	<ul style="list-style-type: none"> <li>記録/編集には不向き (画像との組合せでは効果的)</li> </ul>
入力対象	<ul style="list-style-type: none"> <li>少項目への確実な直指に向く</li> <li>文字・図形・ジェスチャと多彩な機能を持つ</li> </ul>	<ul style="list-style-type: none"> <li>多項目への直指が可能</li> <li>感情・個人性を表現しやすい</li> </ul>

態から、相補的な入出力のモダリティと対話モードを柔軟に選択しながら円滑に対話を進めている。MUI の中での相補性利用には以下の 3 つの型がある。

- 状況に応じて個々のモダリティを選択使用する
- 複数のモダリティを組み合わせて冗長性を持たせ、頑健な UI を構成する
- 複数のモダリティを組み合わせて意味を確定・強化する

MUI が頑健性に優れている例としては、騒音環境の中で音声だけで対話する場合を、絵や図を参照しながら話す場合と比べると容易に理解できよう。表-3 は相補性を示す例として、ペン入力と音声入力の比較を示したものである。ここでは音声入力を利用した多項目への直指について説明を加える。視覚的メタファとペンによるポインティング操作の組合せは、対象が少項目の場合に効果的なためよく利用される。一方、ペンによるポインティングはメニュー選択と同じく目を使うため、選択肢が  $7 \pm 2$  の範囲を越えると操作効率が低下する。音声入力 (による検索) と組合せたポインティングは視覚を使わないため、直接、多項

表-4 マルチモーダル UI の 4 つのカテゴリ<sup>9)</sup>

順序的		並行的
結合的	交互的 (Alternate)	共働的 (Synergistic)
独立的	排他的 (Exclusive)	同時的 (Concurrent)

目をアクセスすることができる。電話をかける場合を例にとろう。ユーザが机の上の電話メタファをペンで押しながら「A さんにつないでくれ」と言うと、A さんを含むディレクトリが表示されポインタは A さんを示す。ユーザはこれを確認してマウスボタンを押すだけでよい。これをディレクトリから目視だけで探す場合と比べると、言葉と組み合わせた指示の優位さが分かるであろう。

一方、MUI の多様性をモダリティ間の同時並行性と融合度から分類することが提案されている<sup>9)</sup>。人間は、表-4 の 4 つの対話形態を状況に応じて柔軟に選択しながら対話を進めている。当面は、タスクに合わせて 4 つのカテゴリを使い分けることになろうが、将来は同じタスクの中でも、ユーザが自由に対話形態を選べることが望ましい。

以上に述べた対話構造を柔軟に選択する際の戦

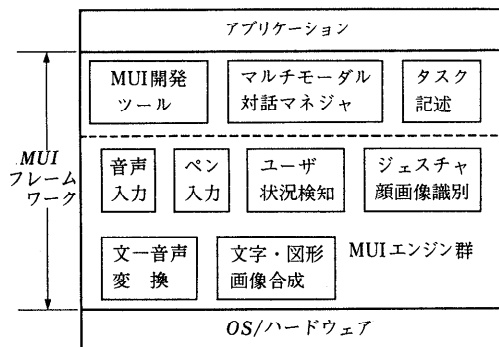


図-1 MUIフレームワーク

略は、今後の大きな課題である。当面は、人間がシステムの対話機能を代行しながら模擬対話データを収集・解析し、結果を対話システムへ組み込み、実環境での評価を含めて評価-再設計を繰り返しながら最適解を求めることになる。こうした手法は Wizard of Oz (WOZ) 方式と呼ばれている。

### 3.2 マルチモーダル UI フレームワーク

MUI フレームワークは、多様な入出力モダリティに対応するエンジン群、対話マネージャ、タスク記述、および MUI 開発ツールを含む (図-1 参照)。MUI ベースの対話では応答の遅れが円滑な対話を阻害し、しばしば対話の破綻につながるため、個々のエンジンと対話マネージャには実時間処理・並行処理・時間同期処理が要求される。また、3.1 に述べた MUI の多様性を実現するには、複数のモダリティを統合し抽象化してメッセージを抽出する機能 (理解)、およびメッセージを複数のモダリティに分配して伝達する機能 (生成) の研究が重要になる。現在、クライアント/サーバモデルを構成する形で、個々のサーバエンジンに対応する API が整備されつつある。MUI では、ここにモダリティを統合・抽象化するレイヤなどのメカニズムを組み込むことが重要になる<sup>9)</sup>。

対話マネージャの役割はタスク記述に沿った対話の運用である (ここでは chat のような対話は除いて考える)。すなわち対話マネージャは、ユーザの入力操作、タスクのシナリオ、および対話履歴からユーザの目的を予測して対話戦略を立て、ゴールに向けて対話を誘導する。対話戦略では、従来のシステム主導をシステム/ユーザ混在主導に

変える試みが始まっているが、音声対話などでの認識結果に対する信頼性の向上、熟練度の異なるユーザへの対応方法、主導権の移行の制御など課題も多い。次に、多くの知識源を参照しながらマルチモーダル対話を進める際のモデルとしては、一般的な階層的モデルの他、マルチエージェントモデル<sup>10),11)</sup>や、ネットワークモデル<sup>12)</sup>などが提案されている。なお、実際の局面では、次節に一部述べる MUI ガイドラインを参照して、円滑な対話を実現することが重要になる。

対話マネージャ、タスク記述、およびアプリケーション・ドメインは、できるだけ独立に設計を進められることが望ましい。このためには、MUI の設計者と応用システムの設計者が共同して、対象システムの基本的概念と語彙の体系、すなわちオントロジー<sup>13)</sup>を抽出、整理する作業が必要になる。「航空機の乗り継ぎ案内」を例にとると、ドメインから航空機予約案内サービスにおける語彙やその意味など (これらは対話にも利用される)、タスクからはスケジューリングにおけるゴールとサブゴールの関係など、さらにマルチモーダル対話からはモダリティの定義、関係などを抽出することになる。タスクとマルチモーダル対話に関するオントロジーは対話構造を決定する。対話マネージャとアプリケーションその他とのプロトコルに関するオントロジーの定義も必要になる。その他、対話文理解/生成 (語彙・構文・意味・語用論) や一般常識まで含めると、これらの広範な作業はオープンな環境で進める必要があろう。

### 3.3 円滑な対話を実現するために

GUI ではアイコンにより操作上の情報が与えられ、そこではアフォードンス<sup>14)</sup>が重要な要素になっていた。コンピュータ UI におけるアフォードンスは、“ユーザに対してコンピュータが提供する行為の可能性についての情報”と言い換えることができよう。GUI では、この仕掛けとして操作環境に現実世界のメタファを取り入れる手法がしばしば用いられている。机上を模したデスクトップ・メタファを例にとると、ユーザはそこに電話を目にする。その電話は「電話をかける」ことをアフォードする。電話をクリックすると受話器が動き次にプッシュボタンが表示され...と、ユーザは操作にともなって変化する画面から次の行為の可能性を受けとり、アフォードンスに沿っ

て目的を達成することができる。

人間同士の対話では、音声と視覚的メディアがうまく使い分けられており、対話の流れの中に各モダリティが自然に埋め込まれている。コンピュータとの対話においても、メディアの違いを活かすと同時に、モダリティの違いや切り換えを意識させず、自然な流れを構成することが大切である。MUIでは、情報の送り手（ユーザ/コンピュータ）が意図したモダリティと、受け手（コンピュータ/ユーザ）の期待しているモダリティにズレを生じることがある。複数のモダリティを使うなど情報の担わせ方に冗長性を持たせることも大切であるが、受け手の予期するメディアとメッセージを考慮して、上述のアフォーダンスをうまく利用することがGUI以上に必要になろう。音や擬人化エージェントの表情といったノンバーバルインタフェースも、円滑な対話の進行に大きな役割を果たすようになると考えられる<sup>15)</sup>。

対話の際に各メディアを入力/提示する時間順序も重要になることがある。視覚メディアと聴覚メディアを例にとると、文の読み上げや図の説明などでは前者を先行させ、注意を喚起する際には後者を先行させると対話が円滑になることが多い。また、対話に使用するメディアの組合せには、音声に対する聴覚応答（“呼べば答える”）やジェスチャに対する可視表示などの相性がある。MUI特有の設計指針は、今後、“MUIのガイドライン”として整備することが望まれる。

システム内部の複雑な処理をユーザから隠し、対話をシンプルに見せる仕掛けとして、顔画像やキャラクタを使用した擬人化エージェントがよく利用される。この代表はApple ComputerのKnowledge Navigatorで登場した蝶ネクタイのエージェント“Phil”であろう。エージェントベースの対話の理想は、“サービスに応じて現れるエージェントが、対話を通してユーザの意図を実現すること”であろうが、この実現には真に知的なエージェントの登場を待たねばならない。しかし、その前にはコンピュータが知的になることの定義に始まり、多くの課題が横たわっている。一方、対話の目的を“ユーザの抱えるタスクをいかにして解決に導くか”に置くなら、単純なタスクではエージェントを脇に退かせ、見慣れたボタンやパネルの世界を提示すれば済むであろう。また

複雑なシステムでは、今後、擬人化エージェントの親しみやすさの面を強調するだけでなく、システムの能力、状態あるいは意図をエージェントが演ずる出力のモダリティ（音、表情、しぐさ、…）に乗せて伝える研究が重要になってくると考えられる。

### 3.4 マルチモーダル UI の開発環境

MUIの設計は、GUIの設計に比べて複雑なことから多くの時間を要する。様々なタスクにわたるマルチモーダル対話システムを効率よく開発するには、先に述べたWOZ法の利用と共に、今後は実環境において組織的に対話データを収集できるラピッドプロトタイプング技術が大変重要になる。

図-2は著者等によって開発されたMUI開発支援ツールを利用した作業フローを示している<sup>16)</sup>。MUI設計エディタ(Muse: Multimodal UI-design Support Editor)は、マウス操作によりUI-object(ボタン/音声入力/文字入力/ユーザ状況検知センサなどの入力オブジェクト、イメージ部品/テキスト部品/録音編集音/文一音声変換出力などの出力オブジェクト)を画面に配置した後、UI-object相互、および画面間にリンクを張って対話シナリオを作成できるようになっている。また同時に、UI-objectの様々なプロパティ(文一音声変換出力の例では、応答文/speaker type/発声速度/音量/声の高さなど)もダイアログボックスを表示して設定することができる。こ

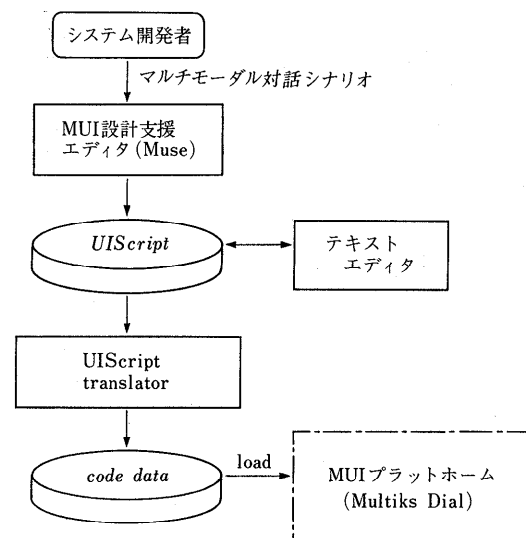


図-2 MUI開発支援ツールと作業フロー

これらの結果はUIScriptに変換され、テキストエディタにより追加・修正を加えることができる。UIScriptのリストは、MUIプラットホーム(MultiksDial<sup>17)</sup>)上で動作するcode dataに変換されてプロトタイプが完成する。開発環境の整備により、設計—プロトタイプ開発—評価のサイクルを短期間にまわすことが可能になった。

MUIの開発支援ツールの機能としては、この他、ユーザの状況や使用環境に応じてモダリティを切り換えるメカニズム<sup>18)</sup>、モダリティ間の時間同期や同時/並行処理などを記述する文法<sup>19)</sup>なども重要なテーマである。

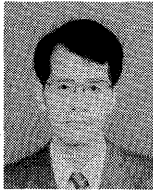
#### 4. おわりに

音声対話を中心とするマルチモーダルUIの周辺と今後の研究課題を概括した。この分野の研究はまだ始められたばかりで多くの課題を抱えており、かつ対象も広い学問分野にわたっている。今後とも、隣接学会、研究会との間で共同討議を行っていくことが大切である。本稿には述べなかったが、マルチモーダルUIと密接な関係にある、インタラクティブ対話システムの評価方法や、論文中の報告事項(システム・実験条件・結果)の標準的枠組みに関する討議も始められており<sup>20),21)</sup>、これらについては別の機会に紹介したい。最後に、本稿に述べた本格的なマルチモーダルUI環境を利用できる日の近いことを期待する。

#### 参考文献

- 1) 音声の知的処理に関する調査研究報告書, 日本情報処理開発協会刊 (Mar. 1992, Mar. 1993).
- 2) Nitta, T.: Speech Recognition Applications in Japan, Proc. Int. Conf. Spoken Language Proc. (ICSLP 94), pp. 671-674 (1994).
- 3) Rudnicky, A. I., Hauptmann, A. G. and Lee, K.: Survey of Current Speech Technology, Commun. ACM, Vol. 37, No. 3, pp. 52-57 (1994).
- 4) Schneiderman, B.: Designing the User Interface 2nd Edit., Addison-Wesley Publishing Co. Inc., 1992 (日本語訳: 東, 井関監訳「ユーザインタフェースの設計—第2版」, pp. 49, 日経BP社).
- 5) Bolt, R. A.: The Integrated Multi-modal Interface, 信学論D, Vol. 70, No. 11, pp. 2017-2025 (1987).
- 6) 新田: 音声マルチモーダル対話の動向, 人工知能学会研究会資料, SIG-SLUD-9202-2 (1992).
- 7) 竹林: 音声自由対話システム TOSBURG II—ユーザ中心のマルチモーダルインタフェースの実現に向けて, 信学論D-II, Vol. J77-D-II, No. 8, pp. 1417-1428 (1994).
- 8) Nigay, L. and Coutaz, J.: A Design Space for Multimodal Systems; Concurrent Processing and Data Fusion, Proc. INTERCHI'93, pp. 172-178 (1993).
- 9) Blattner, M. M., Glinert, J. A. and Ormsby, G. R.: Metawidgets: Towards a Theory of Multimodal Interface Design, Proc. COMP-SAC'92, pp. 115-120 (1992).
- 10) Ueda, H., Hidai, Y. and Uyama, M.: Agency Model—An Architecture for Dynamic Mapping of Appearances and Functions, Proc. of the 7th Symp. on Human Interface SICE, pp. 531-538 (1991).
- 11) Cohen, P. R., Cheyer, A., Wang, M. and Baeg, S. C.: An Open Agent Architecture, Proc. AAAI'94-SA (Stanford), pp. 1-8 (1994).
- 12) 岡, 木山, 伊藤: 認識単位の粒度自由・並列アーキテクチャとその実現のための Reference Interval-free 連続 DP, 情報処理学会音声言語情報処理研究会資料, 7-23, pp. 145-150 (1995).
- 13) Mizoguchi, R.: Knowledge Acquisition and Ontology, Proc. KB & KS 93, pp. 121-128 (1993).
- 14) Gibson, J. J.: The Ecological Approach to Visual Perception, Houghton Mifflin Company, 1979 (日本語訳: 古崎ほか「生態学的視覚論」サイエンス社).
- 15) 黒川隆夫: ノンバーバルインタフェース, オーム社 (1994).
- 16) 新田, 神尾, 雨宮, 松浦, 内山, 田村: マルチモーダルUIとラビッドプロトタイピング, 情報処理学会音声言語情報処理研究会資料, 7-5, pp. 29-34 (1995).
- 17) 神尾, 松浦, 正井, 新田: マルチモーダル対話システム MultiksDial, 信学論D-II, Vol. J77-D-II, No. 8, pp. 1429-1437 (1994).
- 18) 川合, 相田, 齊藤: An Interface Toolkit with Dynamic Selectable Modality, 情報処理学会ヒューマンインタフェース研究会資料, 61-6, pp. 41-48 (1995).
- 19) 島津, 高島: マルチモーダル Definite Clause Grammar(MM-DCG), 信学論D-II, Vol. J77-D-II, No. 8, pp. 1438-1446 (1994).
- 20) 新田, 浜田, 中川, 古井, 平井: パネルディスカッション: インタラクティブなシステムの評価をどのように考えていくか, 情報処理学会音声言語情報処理研究会資料, 7-19-1~5, pp. 119-128 (1995).
- 21) EAGLES: Spoken Language Handbook <http://www.ilc.pi.cnr.it/EAGLES/home.html> (1995).

(平成7年8月29日受付)



**新田 恒雄 (正会員)**

1946年生。1969年東北大学工学部電気工学科卒業。工学博士。(株)東芝マルチメディア技術研究所勤務。現在、音声認識・合成、文字認識、およびこれらの技術をベースとしたマルチモーダル対話システムの研究に従事。1989年より工学院大学電子工学科兼任講師を兼務。1989年電子情報通信学会論文賞受賞。電子情報通信学会、日本音響学会、IEEE各会員。

