

## 複数ユーザによる音声とジェスチャのマルチモーダル インタフェースシステム：Real-time GSI の一評価実験

伊藤慶明†\* 古川清† 中沢正幸† 木山次郎†\*\* 張建新† 岡隆一†

†新情報処理開発機構つくば研究センター ‡メディアドライブ株式会社

\* 現在、川崎製鉄(株) 情報システム部  
〒100 千代田区内幸町 2-2-3 日比谷国際ビル 26F  
E-mail: itoh@system.kawasaki-steel.co.jp

\*\* 現在、シャープ(株) 応用システム研究所

あらまし - 我々は、マルチモーダルインタフェースシステム Real-time GSI (Gesture and Speech Interface) を既に開発している。本システムは、音声と動画像理解の統合、数人の自然な発話の同時認識/理解、及びリアルタイムの視覚的漸次的フィードバックを、人間同士の会話進行とともに実現した点に特徴がある。本稿では、まずマルチモーダルインタフェース・システムの評価方法について検討する。次に、本システムの利用可能性について評価するため、Wizard of OZ 法を用いた簡単な心理実験を行った。その実験方法と、評価結果についての報告を行う。

### An Evaluation Experiment of Real-time GSI : Multi-modal Interface System for Gesture and Speech by Multiple Users

Yoshiaki ITOH †\* Kiyoshi FURUKAWA † Masayuki NAKAZAWA †

Jiro KIYAMA †\*\* JianXin ZHANG † and Ryuichi OKA †

†TRC RWCP, ‡Media Drive Corporation

\* Now, Informations System Dept., Kawasaki Steel Corporation \*\* Now, Sharp Corporation

**Abstract** - We have proposed a new type of multi-modal interface system, which allows input of speech and gesture by multiple users. This system has features such as integration for understanding for speech and gesture, simultaneous recognition of conversational speech by multiple speakers and real-time and gradual visual feedback of system understanding state. In the paper, we first study how to evaluate a multi-modal interface. Next, we report some results of our psychological experiment, using Wizard of OZ method to evaluate the feasibility of the system.

#### 1. はじめに

音声認識システムは100%の認識率<sup>1</sup>ではないと使われないのだろうか。答えは否である、と音声認識に関わる研究者は信じている。では、何%の認識率になったら使われるようになるのかという、非常に単純かつ重要な質問に明確に答えられる研究者も殆んどいない。もちろん、アプリケーションによって、要求される認識率は異なるだろう。100%の認識率を要求するアプリケーションも十分考えられる。しかし、あなたのアプリケーションシステムは認識率が何%になったら実用できるのか、という質問に明確に答えられる研究者もあまりいないようである。

一方、実験室で行われる評価実験の結果からは、多くのシステムでかなり高い(90%以上)認識率が得られている。しかし、このような性能は、ユーザ全体

の平均の認識率に対して利用可能性を評価するもので、平均の認識率に到達しないユーザへの対処方法については、十分な検討が行われていないのが現状である。一方、実際のフィールドテストでは、雑音、発声の変動等が加わる。これらの要因に対するロバスト性が不足しているため、実験室レベルの性能にはとうてい到達しない。ともすると、10~30%以上の認識性能の低下が見られる。こうしたことを前提とする現状の音声対話システムは少ない。

また、従来の音声認識インタフェースシステム(I/F)では、音声は「あったら便利」程度のシステムが多く、「音声ならでは」というシステムは少ない。電話を介するシステムも多いが、最近のマルチメディア時代では、必ずしも音声だけしか使えないという状況も考えにくくなるのではないだろうか。マルチメディア環境においては、もし、人間の明確な意図をコンピュータに伝えるだけであれば、タッチパネル等の

<sup>1</sup>ここではタスク達成率と考えてもよい

もっと確実な手段があり、手軽さゆえに音声が使われるということは、性能が Real World においても相当高くない限り、少ないのではないだろうか。

音声対話システムを実用化する上では、上記の 1. 認識性能の要件設定、2. 実フィールドでの性能低下、3. 音声を必須入力手段とするアプリケーション、という 3 つの問題への対応方法を検討しておく必要があると考える。そこで、これらの問題に対する我々の考え方とアプローチ方法を以下に示す。また、今回、マルチモーダルシステムの評価方法の検討を含めて、ここで述べたことを検証するため、被験者による簡単な心理実験を行うことにした。

我々は既に、複数ユーザによる音声とジェスチャのマルチモーダルリアルタイム I/F システム：Real-time GSI (Gesture and Speech Interface) を構築した。本システムに主には以下の 3 つの特徴がある。

- (1) 音声・ジェスチャの相補的マルチモダリティ
- (2) (複数)ユーザが任意の時刻に自由に発話可能
- (3) ユーザ - 認識/理解 - 結果反応の One Loop の One frame real-time 完結処理

文献 [2] において、我々は上記の特徴を有するシステムでは、ユーザに対し臨場感や計算機との一体感を与え、ユーザの思考の促進手段を提供するのではないかと主張した。このようなコンピュータとの I/F 環境では、たとえ認識率が低くても、アミューズメント性が高く、利用され得るシステムを提供できるのではないかと考える。また、我々が今回提案しているシステムでは、ユーザがシステムを使用するに際して、例えば自分たちの希望する家を設計したいという漠然とした目的はあっても、明確なイメージを持っているとは仮定しない。従って、発声される音声、ジェスチャは自分の思考している様子、それ自体を表すものであり、従来のように、音声コマンドをシステムに与えるようなものではない。音声やジェスチャはまさにリアルタイムで反応することに意味がある情報媒体であり、キーボードやタッチパネルに代替しがたい音声及びジェスチャによる入力が必要となる世界と考える。以上が第 2、3 の問題に対する我々の考え方である。

音声認識システムは、現在 I/F システムにならんとしている。ここでは、単なる認識率や、タスク達成率の評価のみではなく、インタラクティブシステムあるいは I/F としての評価、即ち、ユーザビリティ等のユーザ主観を交えて評価する必要がある。そこで今回は、第 1 の問題である性能がどの程度ならば利用され得るかを評価するため、被験者を用いた簡単な心理実験を行った。3 章でも述べるが、今回は被験者の条件を揃えるため、Wizard of OZ (WOZ) 方式を用いた。

以下、第 2 章で、本システムの特徴と対象としたタスク、システム構成について簡単に述べる。第 3 章

ではマルチモーダルシステムの性能評価方法を検討し、第 4 章で、今回行った実験方法と結果を紹介する。最後に今後の課題とまとめを述べる。

## 2. システムの概要

### 2.1 システムの特徴

本システムは、人間同士のコミュニケーション同様、音声だけではなく手ぶり等のジェスチャも、コンピュータを意識することなく、ヒューマン・マシンインタフェースとして実現しようというものである。本システムでは、以下の機能および特徴を有する。

- (1) Multi-Modal 機能：音声の認識と動画像の認識技術を統合し、自然な発話とジェスチャの同時かつ相補的な理解を可能にした
- (2) 入力のための制限排除：ユーザの思考を妨げないため、任意のタイミングで入力できる。また、ユーザに厳しい文法的制約を与えない
- (3) 複数ユーザを対象：知的作業や発想支援を行うため、複数ユーザの討論形態の入力を実現した
- (4) 画像によるリアルタイム漸次的フィードバック：システムの理解内容のリアルタイム漸次的視覚化は、ユーザに、臨場感や一体感を提供するだけでなく、ユーザの思考の画面への反映となり、思考の支援となり得る

上述した機能を備えた I/F は、人間と計算機の新しい対話形態、新しい I/F と考えることもできる。我々は、frame-wise and realtime spotting 技術を適用し、上述の機能を人間同士の会話進行とともに実現する新しい I/F システムを構築した。

### 2.2 タスク

本システムのタスクは、一般ユーザが自分(達)の家の大雑把な配置を決定していくという、「家の配置設計」である。ユーザは、通常、「こうしたい」という漠然としたイメージを抱いているが、予め仕様が確定しているわけではない。ユーザ間の会話や一時的な変更要求に対する視覚的フィードバックを通して、試行錯誤的、漸時的に自分の要求を満足する仕様にしていく。このため、視覚的な漸時的フィードバックは対話を進める上で重要な要素となる。また、配置問題では例えば、「こっち」という指示のようにジェスチャが音声言語と相補的に大きな役割を担うことが多い。このように、会話音声による思考の促進、視覚的な漸時的フィードバックの役割、音声とジェスチャの相補的入力、さらにユーザにとってのアミューズメント性を考慮すると、「家の配置設計」はそれに適合するものだと考えられる。

ユーザには図 2 のように、システムの現在理解状況の画像が提示され、希望するイメージをユーザが発

声、あるいは会話、ジェスチャで指示すると、即座に画像に反映されていくことになる。人間へのフィードバックは、グラフィクスを用いて、過去の理解状況とともに現在の理解状況をリアルタイムに表示する。

## 2.3 システム構成

システムの構成例を図1に示す。本システムでは、各ユーザにマイクロフォンおよびカメラが向けられ、それぞれ音声認識部 (Speech Processing Unit : SPU), 動画像認識部 (Image Processing Unit : IPU), に送られる。音声認識部では、フレームと同期したワードスポッティングによって、リアルタイムに結果を連続オートマトンへ出力する。動画像認識部では、各ユーザ毎に特徴分析、ジェスチャスポッティングを行い、同様に連続オートマトンへ出力する。

音声認識部と動画像認識部から得られるスポッティング結果の統合方式としては、連続オートマトン [3]-[5] を用いた。連続オートマトンによってスポッティングされた概念は、ユーザ側に即座にシステムの理解状況として、画像によってフィードバックされる。これにより、ユーザはある概念を発話、あるいはジェスチャによって示した瞬間に、その概念の実現イメージが画面に現れる。

コンピュータとのユーザフレンドリな対話を実現するために、画面上に女性を模したエージェントを表示し、ユーザからの入力がしばらくない場合には、エージェントが適度に口と顔を動かしながら、最新の認識結果を音声合成で答え、ユーザの注意を喚起し会話の維持に努めている。

## 3. 評価方法の検討

### 3.1 音声 I/F システムの評価の難しさ

インタラクティブシステムの評価は難しい [12]。音声認識システムでは、認識率等の絶対的評価尺度で、方法論等の優劣を簡単に比較することができた。しかし、これを I/F システムあるいはインタラクティブ

システムと捉えたと、認識率に代わる絶対的評価尺度を設定することはできず、現状では、客観的にシステムを概観するために、個々システムの特徴や条件などを列記するに留まっている。I/F システムでは認識率だけではなく、ユーザビリティに起因するユーザの主観的判断がそのシステムを評価する上で重要な要素、あるいは最も信頼できる評価指標になってくる。

マルチモーダルシステムの評価はさらに難しいと考えられる。マルチモーダルシステムにおいても、絶対的評価尺度はなく、これらの特徴や条件が、個々のシステムで様々であるため、性能の比較を困難にしている。1つのシステムにおいても、あるモードが加わったことによる効果や、各モードの性能 (機能) 向上による定性的な効果は、ある程度求められるものの、マルチモーダルシステムとしての定量的な効果は、測定困難である。例えば、現在エージェントを画面に表示し、コンピュータからの出力はエージェントを介して音声で会話した場合の有効性が報告されている [10] (本システムでも有効性が確認できている) が、エージェントがどの程度有効か、認識性能をどの程度上げるのと等価かといった定量評価は難しい。

我々は、このような状況の中で、今回、我々のシステムの利用を考えた場合、要求される性能およびリアルタイムフィードバックの重要性を確認するため、認識率とシステムの性能 (ユーザの主観的評価) について評価したいと考えた。

### 3.2 認識系 I/F システムの評価方法の検討

構築した I/F システムを、被験者に利用して評価してもらうことによって、現システムの利用可能性については評価できるだろう。しかし、今回の目的である、認識性能と利用可能性の評価および要求される性能に関する知見は、被験者毎の認識性能が異なるため、このような実験からは得ることが困難である。

そこで、この目的のためには、認識性能が被験者全員で同一の条件となるように、Wizard of OZ (WOZ) 方式を適用するのが適切ではないかと考える。即ち、設定した認識性能毎のユーザの評価を得ることによって、認識性能と利用可能性の関係、及び要求される性能が、ユーザの満足度を基準に得られると考えた。しかし、WOZ 方式を用いる上でも、次節で述べるようないくつかの問題点がある。

### 3.3 WOZ 方式による性能シミュレーションと評価方法の課題

WOZ 方式を用いて実験を行おうという場合、既に行われているような 100% の認識性能を実現 [10][8] することは容易である。しかし、ある認識率をシミュレーションするとすると、その認識システムの誤りをシミュレートする必要があり、例えば以下のような点を考慮する必要がある。

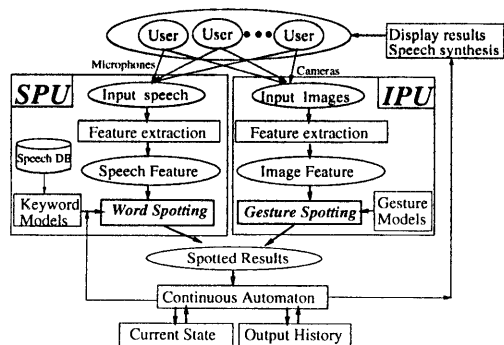


図1: 音声動画像認識対話システムの構成

- (1) エラーの種類：置換，挿入，脱落エラーのうちの誤りを発生させるか。
- (2) エラーの発生箇所：エラーをいつ発生させるか（特に挿入エラー）。
- (3) エラーの内容：置換エラーの場合，どの単語をどの単語への誤認識とするか。挿入エラーの場合，どういった認識／理解内容を挿入するか。
- (4) 発声とエラーの一貫性：発声した言葉とエラーの対応の一貫性を考慮するか。

本来なら，上記の問題をクリアするためには，システムの誤り方を分析して，それをシミュレーションするような枠組が必要である。

被験者を使った心理実験では，複数の条件で比較実験を行うことになる。例えば N 人の被験者で 4 つの条件を比較したい場合，以下のような方法が考えられる。但し，括弧内は 1 つの条件を受ける人数を示す。

- (1) 各被験者：一条件での実験 (N/4 人)
- (2) 各被験者：一定順の 4 条件での実験 (N 人)
- (3) X 通りの条件の順序を用意，各被験者：1 つの順序，4 条件での実験 (N/X：X<sub>Max</sub> = 24)

被験者が受ける条件の順序によって，印象は大きく影響されるため（予備実験でも確認），本来なら条件の順序の影響を客観的に除去した (1)，あるいは様々な順序で行った (3) で比較評価することが望ましい。

#### 4. 本システムの一評価実験

##### 4.1 実験システム

現在，リアルタイムの制約から，音声認識部のキーワードは 50 単語，動画像認識部は「前へ」，「後ろへ」，「不要」を表す 3 種類となっている。音声認識と動画像認識は異なるフレーム周期で処理されているが，音声側の短い周期 (8msec) で連続オートマトンの更新及び出力を行っている。2 人のユーザがシステムと対話をしている様子を図 2 に示す。

##### 4.2 ユーザへの Instruction と対話サポート

システムのユーザへの説明は，付録 1 に示した内容のものを読んでもらった後，システムが受理でき



図 2: 2 人のユーザによる本システムの利用例

0: 扉	1: ベランダ	2: 池	3: 緑の扉	4: 赤の扉	5: 左の窓	6: 右の窓	7: 表の扉
1: 緑の扉	2: 池	3: 花壇	4: 2 階左の窓	5: 2 階右の窓	6: 2 階右の窓		
7: 窓	8: 駐車場	9: 屋根	10: 門	11: 1 階左の窓	12: 1 階左の窓	13: 1 階右の窓	
14: 窓	15: 消火器	16: 階段	17: 階段	18: 階段	19: 階段	20: 階段	21: 階段
22: 階段	23: 階段	24: 階段	25: 階段	26: 階段	27: 階段	28: 階段	29: 階段
30: 階段	31: 階段	32: 階段	33: 階段	34: 階段	35: 階段	36: 階段	37: 階段
38: 階段	39: 階段	40: 階段	41: 階段	42: 階段	43: 階段	44: 階段	45: 階段
46: 階段	47: 階段	48: 階段	49: 階段	50: 階段	51: 階段	52: 階段	53: 階段
54: 階段	55: 階段	56: 階段	57: 階段	58: 階段	59: 階段	60: 階段	61: 階段
62: 階段	63: 階段	64: 階段	65: 階段	66: 階段	67: 階段	68: 階段	69: 階段
70: 階段	71: 階段	72: 階段	73: 階段	74: 階段	75: 階段	76: 階段	77: 階段
78: 階段	79: 階段	80: 階段	81: 階段	82: 階段	83: 階段	84: 階段	85: 階段
86: 階段	87: 階段	88: 階段	89: 階段	90: 階段	91: 階段	92: 階段	93: 階段
94: 階段	95: 階段	96: 階段	97: 階段	98: 階段	99: 階段	100: 階段	101: 階段
102: 階段	103: 階段	104: 階段	105: 階段	106: 階段	107: 階段	108: 階段	109: 階段
110: 階段	111: 階段	112: 階段	113: 階段	114: 階段	115: 階段	116: 階段	117: 階段
118: 階段	119: 階段	120: 階段	121: 階段	122: 階段	123: 階段	124: 階段	125: 階段
126: 階段	127: 階段	128: 階段	129: 階段	130: 階段	131: 階段	132: 階段	133: 階段
134: 階段	135: 階段	136: 階段	137: 階段	138: 階段	139: 階段	140: 階段	141: 階段
142: 階段	143: 階段	144: 階段	145: 階段	146: 階段	147: 階段	148: 階段	149: 階段
150: 階段	151: 階段	152: 階段	153: 階段	154: 階段	155: 階段	156: 階段	157: 階段
158: 階段	159: 階段	160: 階段	161: 階段	162: 階段	163: 階段	164: 階段	165: 階段
166: 階段	167: 階段	168: 階段	169: 階段	170: 階段	171: 階段	172: 階段	173: 階段
174: 階段	175: 階段	176: 階段	177: 階段	178: 階段	179: 階段	180: 階段	181: 階段
182: 階段	183: 階段	184: 階段	185: 階段	186: 階段	187: 階段	188: 階段	189: 階段
190: 階段	191: 階段	192: 階段	193: 階段	194: 階段	195: 階段	196: 階段	197: 階段
198: 階段	199: 階段	200: 階段	201: 階段	202: 階段	203: 階段	204: 階段	205: 階段
206: 階段	207: 階段	208: 階段	209: 階段	210: 階段	211: 階段	212: 階段	213: 階段
214: 階段	215: 階段	216: 階段	217: 階段	218: 階段	219: 階段	220: 階段	221: 階段
222: 階段	223: 階段	224: 階段	225: 階段	226: 階段	227: 階段	228: 階段	229: 階段
230: 階段	231: 階段	232: 階段	233: 階段	234: 階段	235: 階段	236: 階段	237: 階段
238: 階段	239: 階段	240: 階段	241: 階段	242: 階段	243: 階段	244: 階段	245: 階段
246: 階段	247: 階段	248: 階段	249: 階段	250: 階段	251: 階段	252: 階段	253: 階段
254: 階段	255: 階段	256: 階段	257: 階段	258: 階段	259: 階段	260: 階段	261: 階段
262: 階段	263: 階段	264: 階段	265: 階段	266: 階段	267: 階段	268: 階段	269: 階段
270: 階段	271: 階段	272: 階段	273: 階段	274: 階段	275: 階段	276: 階段	277: 階段
278: 階段	279: 階段	280: 階段	281: 階段	282: 階段	283: 階段	284: 階段	285: 階段
286: 階段	287: 階段	288: 階段	289: 階段	290: 階段	291: 階段	292: 階段	293: 階段
294: 階段	295: 階段	296: 階段	297: 階段	298: 階段	299: 階段	300: 階段	301: 階段
302: 階段	303: 階段	304: 階段	305: 階段	306: 階段	307: 階段	308: 階段	309: 階段
310: 階段	311: 階段	312: 階段	313: 階段	314: 階段	315: 階段	316: 階段	317: 階段
318: 階段	319: 階段	320: 階段	321: 階段	322: 階段	323: 階段	324: 階段	325: 階段
326: 階段	327: 階段	328: 階段	329: 階段	330: 階段	331: 階段	332: 階段	333: 階段
334: 階段	335: 階段	336: 階段	337: 階段	338: 階段	339: 階段	340: 階段	341: 階段
342: 階段	343: 階段	344: 階段	345: 階段	346: 階段	347: 階段	348: 階段	349: 階段
350: 階段	351: 階段	352: 階段	353: 階段	354: 階段	355: 階段	356: 階段	357: 階段
358: 階段	359: 階段	360: 階段	361: 階段	362: 階段	363: 階段	364: 階段	365: 階段
366: 階段	367: 階段	368: 階段	369: 階段	370: 階段	371: 階段	372: 階段	373: 階段
374: 階段	375: 階段	376: 階段	377: 階段	378: 階段	379: 階段	380: 階段	381: 階段
382: 階段	383: 階段	384: 階段	385: 階段	386: 階段	387: 階段	388: 階段	389: 階段
390: 階段	391: 階段	392: 階段	393: 階段	394: 階段	395: 階段	396: 階段	397: 階段
398: 階段	399: 階段	400: 階段	401: 階段	402: 階段	403: 階段	404: 階段	405: 階段
406: 階段	407: 階段	408: 階段	409: 階段	410: 階段	411: 階段	412: 階段	413: 階段
414: 階段	415: 階段	416: 階段	417: 階段	418: 階段	419: 階段	420: 階段	421: 階段
422: 階段	423: 階段	424: 階段	425: 階段	426: 階段	427: 階段	428: 階段	429: 階段
430: 階段	431: 階段	432: 階段	433: 階段	434: 階段	435: 階段	436: 階段	437: 階段
438: 階段	439: 階段	440: 階段	441: 階段	442: 階段	443: 階段	444: 階段	445: 階段
446: 階段	447: 階段	448: 階段	449: 階段	450: 階段	451: 階段	452: 階段	453: 階段
454: 階段	455: 階段	456: 階段	457: 階段	458: 階段	459: 階段	460: 階段	461: 階段
462: 階段	463: 階段	464: 階段	465: 階段	466: 階段	467: 階段	468: 階段	469: 階段
470: 階段	471: 階段	472: 階段	473: 階段	474: 階段	475: 階段	476: 階段	477: 階段
478: 階段	479: 階段	480: 階段	481: 階段	482: 階段	483: 階段	484: 階段	485: 階段
486: 階段	487: 階段	488: 階段	489: 階段	490: 階段	491: 階段	492: 階段	493: 階段
494: 階段	495: 階段	496: 階段	497: 階段	498: 階段	499: 階段	500: 階段	501: 階段
502: 階段	503: 階段	504: 階段	505: 階段	506: 階段	507: 階段	508: 階段	509: 階段
510: 階段	511: 階段	512: 階段	513: 階段	514: 階段	515: 階段	516: 階段	517: 階段
518: 階段	519: 階段	520: 階段	521: 階段	522: 階段	523: 階段	524: 階段	525: 階段
526: 階段	527: 階段	528: 階段	529: 階段	530: 階段	531: 階段	532: 階段	533: 階段
534: 階段	535: 階段	536: 階段	537: 階段	538: 階段	539: 階段	540: 階段	541: 階段
542: 階段	543: 階段	544: 階段	545: 階段	546: 階段	547: 階段	548: 階段	549: 階段
550: 階段	551: 階段	552: 階段	553: 階段	554: 階段	555: 階段	556: 階段	557: 階段
558: 階段	559: 階段	560: 階段	561: 階段	562: 階段	563: 階段	564: 階段	565: 階段
566: 階段	567: 階段	568: 階段	569: 階段	570: 階段	571: 階段	572: 階段	573: 階段
574: 階段	575: 階段	576: 階段	577: 階段	578: 階段	579: 階段	580: 階段	581: 階段
582: 階段	583: 階段	584: 階段	585: 階段	586: 階段	587: 階段	588: 階段	589: 階段
590: 階段	591: 階段	592: 階段	593: 階段	594: 階段	595: 階段	596: 階段	597: 階段
598: 階段	599: 階段	600: 階段	601: 階段	602: 階段	603: 階段	604: 階段	605: 階段
606: 階段	607: 階段	608: 階段	609: 階段	610: 階段	611: 階段	612: 階段	613: 階段
614: 階段	615: 階段	616: 階段	617: 階段	618: 階段	619: 階段	620: 階段	621: 階段
622: 階段	623: 階段	624: 階段	625: 階段	626: 階段	627: 階段	628: 階段	629: 階段
630: 階段	631: 階段	632: 階段	633: 階段	634: 階段	635: 階段	636: 階段	637: 階段
638: 階段	639: 階段	640: 階段	641: 階段	642: 階段	643: 階段	644: 階段	645: 階段
646: 階段	647: 階段	648: 階段	649: 階段	650: 階段	651: 階段	652: 階段	653: 階段
654: 階段	655: 階段	656: 階段	657: 階段	658: 階段	659: 階段	660: 階段	661: 階段
662: 階段	663: 階段	664: 階段	665: 階段	666: 階段	667: 階段	668: 階段	669: 階段
670: 階段	671: 階段	672: 階段	673: 階段	674: 階段	675: 階段	676: 階段	677: 階段
678: 階段	679: 階段	680: 階段	681: 階段	682: 階段	683: 階段	684: 階段	685: 階段
686: 階段	687: 階段	688: 階段	689: 階段	690: 階段	691: 階段	692: 階段	693: 階段
694: 階段	695: 階段	696: 階段	697: 階段	698: 階段	699: 階段	700: 階段	701: 階段
702: 階段	703: 階段	704: 階段	705: 階段	706: 階段	707: 階段	708: 階段	709: 階段
710: 階段	711: 階段	712: 階段	713: 階段	714: 階段	715: 階段	716: 階段	717: 階段
718: 階段	719: 階段	720: 階段	721: 階段	722: 階段	723: 階段	724: 階段	725: 階段
726: 階段	727: 階段	728: 階段	729: 階段	730: 階段	731: 階段	732: 階段	733: 階段
734: 階段	735: 階段	736: 階段	737: 階段	738: 階段	739: 階段	740: 階段	741: 階段
742: 階段	743: 階段	744: 階段	745: 階段	746: 階段	747: 階段	748: 階段	749: 階段
750: 階段	751: 階段	752: 階段	753: 階段	754: 階段	755: 階段	756: 階段	757: 階段
758: 階段	759: 階段	760: 階段	761: 階段	762: 階段	763: 階段	764: 階段	765: 階段
766: 階段	767: 階段	768: 階段	769: 階段	770: 階段	771: 階段	772: 階段	773: 階段
774: 階段	775: 階段	776: 階段	777: 階段	778: 階段	779: 階段	780: 階段	781: 階段
782: 階段	783: 階段	784: 階段	785: 階段	786: 階段	787: 階段	788: 階段	789: 階段
790: 階段	791: 階段	792: 階段	793: 階段	794: 階段	795: 階段	796: 階段	797: 階段
798: 階段	799: 階段	800: 階段	801: 階段	802: 階段	803: 階段	804: 階段	805: 階段
806: 階段	807: 階段	808: 階段	809: 階段	810: 階段	811: 階段	812: 階段	813: 階段
814: 階段	815: 階段	816: 階段	817: 階段	818: 階段	819: 階段	820: 階段	821: 階段
822: 階段	823: 階段	824: 階段	825: 階段	826: 階段	827: 階段	828: 階段	829: 階段
830: 階段	831: 階段	832: 階段	833: 階段	834: 階段	835: 階段	836: 階段	837: 階段
838: 階段	839: 階段	840: 階段	841: 階段	842: 階段	843: 階段	844: 階段	845: 階段
846: 階段	847: 階段	848: 階段	849: 階段	850: 階段	851: 階段	852: 階段	853: 階段
854: 階段	855: 階段	856: 階段	857: 階段	858: 階段	859: 階段	860: 階段	861: 階段
862: 階段	863: 階段	864: 階段	865: 階段	866: 階段	867: 階段	868: 階段	869: 階段
870: 階段	871: 階段	872: 階段	873: 階段				

#### 4.4 ユーザへのアンケート

実験の対象としたユーザは、主に筑波大学の学生20名である。アンケートの結果より、実際に音声認識システムを利用した経験者は、PCやWS付随のシステムを使った2人だけだった。また、アンケートおよび口頭での質問から、音声認識システムを理解している人は対象となっていないと考える。アンケート内容の抜粋は付録2に示す。

#### 4.5 結果と検討

本稿では、1章で述べたように、利用可能性と性能に関する議論を中心に、実験結果の検討を行う。ユーザに前節の4つの条件でシステムを(1)～(4)の順で使ってもらい、アンケートにより、システムの比較、及びそれぞれのシステムの印象について調査した。その結果を表1に示す。

表1. System Impression Comparison

Conditions	(1)	(2)	(3)	(4)
T.A.R.	90%	70%	50%	70%
Speed	R/T	5 sec.	R/T	R/T
Impression Grade	2.2	3.3	3.1	1.3
Usability Judge	81%	23%	27%	95%

T.A.R.: Task Accomplishment Rate  
R/T:real-time, 5 sec.:5 second delay for response

ユーザには最後にシステムの順位をつけてもらった。その平均順位を表中の Impression Grade で示す。また、実際の利用に耐えられるか質問し、利用できるかと判断された割合を Usability Judge として表中に示した。(1)は(4)より設定性能で優れているが、ユーザは(4)の方が良いと判断している。これは初回の印象が薄いためと、習熟度合が後の方が高いためと考えられる。また、反応に遅れがある場合や、タスク達成率が50%程度になると、ユーザの印象は極端に悪くなる。この結果から、我々は、本システムを利用する場合、タスク達成率が90%以上にならなくとも、70%程度でも実用に耐えられるのではないかと認識を得た。

現在、システムに慣れたユーザの会話に対する簡単なテストでは、挿入エラーを含め7割近いタスク達成率を得ている。

#### 4.6 今後の課題

3章で述べた課題に対し、今回の実験では、必ずしも最適な方法で対処した訳ではなく、様々な角度からの実験が必要と考える。評価方法自体についても、各々の課題を整理して少ない被験者で、定量的に評価する方法を検討する必要があるだろう。また、例えば、今回の実験の結果の中でも、画面で受け答えを行うエージェントの必要性は、5段階評価で4.6で高い

と認められたが、マルチモーダル I/F の個々の要素技術の機能や性能と、Usability の関係も検討が必要ではないかと考える。

また、I/F の評価では、ユーザの主観的評価が重要である。同一条件を被験者に与えるには、WOZ 法が効果的だが、この場合のシステムのシミュレート方法、及びシステムの動作とユーザの主観との関係等は、今後の重要な研究課題と位置付けられるだろう。

一方、認識の世界では主に技術比較のため、平均の認識率で評価がなされている場合が多い。しかし、平均認識率95%と言っても、95%の人が満足する訳ではない。I/F を構築する上では、被験者各人の認識率における分散は非常に重要であり、例えば、「良い方からの累積で90%の人が含まれる最低認識率」というような新しい評価尺度が必要であろう。さらに、認識系の I/F においては、同様に、「90%の人が満足する認識率」というような、指標も必要ではないだろうか。このような適切な評価尺度の決め方、及びそれに付随するアプリケーションの目標の設定方法は、今後認識を応用する I/F で重要な課題になるのではないかと考える。

#### 5. おわりに

本稿では、既に開発した複数ユーザによる音声とジェスチャのマルチモーダル I/F システムをベースに、マルチモーダルシステムの評価方法及びその課題を検討した。また、Wizard of OZ 方式によってシステムの性能をシミュレーションし、性能と Usability について、被験者を用いての簡単な評価実験を行った。この結果から、One frame real-time 完結処理方式に基づく我々のシステムでは、タスク達成率が高くなくとも十分利用に耐えられるという認識を得た。

謝辞 本研究の機会を与えて下さった新情報処理開発機構 島田潤一所長に深く感謝致します。

#### 参考文献

- [1] 木山次郎, 伊藤慶明, 岡隆一: “リアルタイム発話視覚化システムの試作”, 情処研報, SLP2-3 (1994-7).
- [2] 木山次郎, 伊藤慶明, 岡隆一: “同時複数話者の会話音声およびジェスチャのリアルタイム統合理解による Novel Interface System”, 情処研報, SLP7-3(1995-7).
- [3] 岡隆一, 木山次郎, 伊藤慶明: “概念スボットイングのための画像オートマトン”, 音講論 3-4-12 (1995-3).
- [4] R. Oka, J. Kiyama, H. Kojima, Y. Itoh, S. Seki, and S. Nagaya, “Real-time Integration of Speech, Gesture, Graphics and Data-base” '95 RWC Symposium (1995-6).
- [5] 岡隆一, 木山次郎, 伊藤慶明: “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”, 情処研報, SLP7-23(1995-7).

- [6] K. Nagao and A. Takeuchi, "A New Modality for Natural Human-Computer Interaction: Integration of Speech Dialogue and Facial Animation", Proc. ISSD-93, (1993-11)
- [7] 竹林洋一: "音声自由対話システム TOSUBURG 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), J77-D-II, No.8 (1994-8).
- [8] 伊藤克巨, 秋葉友良, 長谷川修, 速水悟, 田中和世: "音声対話システム構築のための実対話データの収録実験", 情処研報, SLP2-6(1994-7).
- [9] 吉岡理, 南泰浩, 山田智一, 鹿野清宏, "電話番号案内を対象としたマルチモーダル対話システムの作成", 音講論, 1-8-19, (1993-10).
- [10] 安藤ハル, 畑岡信夫, "マルチモーダルなエージェント型インタフェースの評価と対話制御の検討", 情処研報, SLP7-15(1995-7).
- [11] 畑崎香一郎, 野口淳, F. Ehsani, 渡辺隆夫: "発話同時理解による音声対話インタフェースの検討", 音講論, 3-4-13 (1993-3).
- [12] 新田恒雄, 浜田 洋, 中川聖一, 古井 貞熙, 平井 誠, "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" 情処研報, SLP7-19-1 ~ 5(1995-7).
- [13] 高橋勝彦, 関進, 小島浩, 岡隆一: "ジェスチャー動画像のスポッティング認識," 信学論 (D-II), J77-D-II, 8, (1994).
- [14] 嵯峨山茂樹: "なぜ音声認識は使われないか・どうすれば使われるか?", 情処研報, SLP1-4(1994-5).

**[付録 1 : INSTRUCTION]**

私たちは、マルチメディア環境における新しいマン・マシンインタフェースを研究しています。今回ご利用頂くシステムでは、キーボードとマウスといった従来のコンピュータの使い方は異なり、音声とジェスチャを入力手段に利用します。では、どのようにシステムをご利用頂くか簡単にご説明します。

**A. 音声入力**

- (1) 画面に表示されているものの形や色を音声で変更することができます。但し、できないものもあります。変更する時は、対象物を指定して、変更後の形、色等を指定します。

例: 「窓を大きくして」, 「扉は生け垣の方がいいな」

[(2)は移動または削除, (3)削除の復活について]

**B. ジェスチャ入力**

- (1) 移動または削除する以下の3つの場合にジェスチャで入力することができます。

「前」, 「後ろ」, 「いない」

この場合も対象物を指定した後にどうするのかをジェスチャで入力します。

例: 「池はいない」, 「植木はいない」

↑            ↑            ↑            ↑  
 音声    ジェスチャ    音声    ジェスチャ

- C. 次に実際の会話例のビデオを見て下さい。  
 この例のようにシステムは間違っても認識したり、反応しない場合もあります。その場合、くり返し入力して下さい。
- D. 今回は、私どもの研究員と会話をしながら、自分の希望の家にしていくという設定でシステムを使って下さい。なお、比較実験のため、4つの異なるシステムで会話を行って頂いて、システム終了後、どちらのシステムの方がよいかについて、簡単なアンケートを取らせて頂きます。

**[付録 2 : アンケート]**

**2. 次の質問にお答え下さい。**

1. 前回のシステムと今回のシステムの違いがわかりましたか?  
 よく分かった |-----| 全く分からない
2. ジェスチャによる入力はどちらのシステムがうまくいったと思いますか?  
 前回のシステム |-----| 今回のシステム  
 前前回のシステム |-----| 今回のシステム
5. コンピュータの反応はどちらのシステムが速かったですか?  
 前回のシステム |-----| 今回のシステム  
 前前回のシステム |-----| 今回のシステム
6. 音声での入力はどちらのシステムがうまくいったと思いますか?  
 前回のシステム |-----| 今回のシステム  
 前前回のシステム |-----| 今回のシステム
9. システム全体としてどちらのシステムがよかったですか?  
 前回のシステム |-----| 今回のシステム  
 前前回のシステム |-----| 今回のシステム
10. 9の質問で「よくない方のシステム」を「よくする」には、どうすればよいと思いますか?

**4. 次の質問にお答え下さい。**

氏名 ( ) 出身地 大卒の場合の専攻

1. 音声認識システムをどの程度知っていますか?  
 知らない |-----| 知っている  
 1 2 3 4 5
2. 音声認識システムを使ったことがありますか?  
 Yes [どんな? : ]  
 No
3. 画面に出てくる人の声は聞きとりやすかったですか?  
 聞きにくい |-----| 聞きやすい  
 1 2 3 4 5
4. 画面に出てくる人は自然な感じがしましたか?  
 不自然 |-----| 自然  
 1 2 3 4 5
5. 画面に出てくる人は必要だと思いますか?  
 不必要 |-----| 必要  
 1 2 3 4 5
6. このようなシステムで家を設計してみたいと思いますか?  
 思わない |-----| 思う  
 1 2 3 4 5
7. 6の質問で「このシステムで家を設計してみたいと思わない」と感じた方、システムがどうすれば「このシステムで設計してみたい」と思うようになりますか?
8. もう一度このシステムを(奥様や彼女と)使ってみたいと思いますか?  
 使ってみたいと思わない |-----| 使ってみたい
9. システムの順位をつけるとすると、どうなりますか?  
 違いを感じなかった場合は括弧に入れて下さい。  
 また、どれ位までなら実際に利用可能だと思えますか?  
 1位            2位            3位            4位
10. 最も印象の悪かったシステムでも使えると思いますか?  
 使えない |-----| 使える
11. このシステムは1人で使った方が良いですか? 今回のように2人で使った方が良いですか? その他?
12. 改善すべき点があれば御指摘下さい。
13. このシステムに対する御意見、気付いたことがあれば、ご自由にご記入下さい。

ご協力、ありがとうございます。