

MITSUBISHI

三菱電機技報 Vol.76 No.8

特集「IT時代のヒューマンインタフェース技術」

2002 8

Human Interface



In a new generation of
Information Technology

目次

特集「IT時代のヒューマンインタフェース技術」

「ヒューマンインタフェース技術」に寄せて	1
岡田美智男	
IT時代のヒューマンインタフェース——現状と展望——	2
依田文夫・石川 泰	
音声認識技術とその応用	7
岩崎知弘・花沢利行	
テキスト音声合成技術	11
藤井洋一・石川 泰	
音声対話技術	15
石川 泰・渡邊圭輔・谷垣宏一	
人物監視技術	19
ポール バーズリー・レシエク チプリンスキー・ポール ビオラ・鷺見和彦	
接触認識技術“ダイヤモンドタッチ”	25
ポール ディーツ・ダレンリー	
携帯端末用“Z3D”グラフィックスエンジン	29
亀山正俊・藤本仁志・岩崎建樹・河合浩行・児玉幸夫	
携帯電話向け画像処理技術とその応用	33
的場成浩・田中正二・ミロスロー ボーバ	
マルチプロジェクト技術	38
ラメシュ ラスカル・ジェロエン ファンバー・芦崎能広	
ネットワーク上の文書テキスト検索・利用技術	42
鈴木克志・高山泰博	
情報化オフィスにおける文書インタフェース技術	46
岡田康裕・亀代泰三・平野 敬	
インタフェースデザイン評価技術	50
若松正晴・沢田久美子	
ヒューマンセンシング技術の開発と作業計測への応用	54
白松直樹・坂口貴司・平澤宏祐・大須賀美恵子	
産業用ヒューマンインタフェース評価システム	58
坂上聡子・寺下尚孝	
特許と新案	
「対話処理装置」「表を含む帳票処理装置」	63
「位置決め制御装置および位置決め制御方法」	64
スポットライト	
ライフエンドまでの省エネルギー性も考慮した	
“快測センサー霧ヶ峰” WXシリーズ	62
三菱FAXOCRシステム“MELFOS”	(表3)

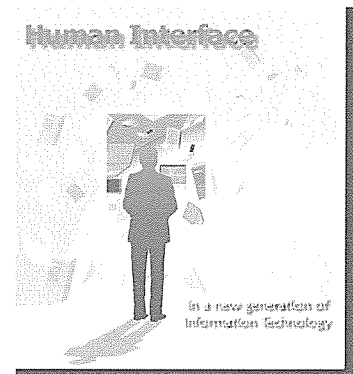
表紙

IT時代のヒューマンインタフェース技術

本格的なIT(情報技術)社会を迎え、我々は多くの情報機器に接し、また、それらの機器を通じて莫大(ばくだい)な情報と向かい合うようになってきた。我々がそのIT社会の恩恵を受けるために、ITシステムと人間とを結ぶヒューマンインタフェース技術の重要性はますます大きくなっている。

表紙は、携帯電話、パソコン、カーナビ、PDA(携帯情報端末)など多種多様な機器と対面し、また、それを通じて膨大な情報や情報通信サービスを受けるための扉を開く技術こそがヒューマンインタフェース技術であることを表している。

三菱電機は、これからもユーザーの立場に立った人間に優しいインタフェース技術の開発に注力していく予定である。



「ヒューマンインタフェース技術」に寄せて

ATRメディア情報科学研究所 主任研究員
京都大学大学院情報学研究科 客員助教授

岡田美智男



幼児の笑顔を眺めていると、なぜかホッとした気分になる。その母親の胸の中に身体を委(ゆだ)ねきった姿に畏敬(いけい)の念さえ感じてしまう。“もっと肩の力を抜いたら…”と、その視線は私たちに語りかけているようでもある。顔の下半分にある大きな目、丸い頬(ほほ)、ヨタヨタした動き、柔らかな体表など、動物の赤ちゃんが共通して備える“かわいらしさ”は、動物行動学のローレンツによって幼児図式と呼ばれている。それは親たちの関与を引き出すための生得的な解発機構であり、子育ての行動を駆り立てる原動力になる。“かわいらしさ”も幼児にとっては大切な機能の一部なのだろう。

いま、モノ作りの世界では“役に立つもの”から“かわいいもの”へと、その価値観の変容が著しい。モノそのものは、目的を達成するための道具や手段ではなく、むしろその存在自体に意味を持つ。その意味で利便性や効率性ではなく、私たちとモノとのかかわりの中から生み出される新たな意味や価値が求められつつある。“全く役に立たないけれど、いないと何だか寂しい”，そんな存在感なのである。

“これは何ですか、これは何に役に立つモノなの？”，これは20世紀のモノ作りにおいて繰り返されてきた問いである。私たち大人はモノに対して一方的に機能や役割を帰属させやすい。あるいは、その原理からモノの価値を把握しようとする。その一方で、子供たちのモノに対する振る舞いは面白い。とにかく撫(な)ぜたり、叩(たた)いたり、眺めたりと、その対象との身体的なかかわりを通して自分と一緒に何ができるものなのかを探ろうとする。そしていつの間にか、新しい遊びやそのモノとの間で新たな意味や価値を見付けだしている。

最近の発達研究の視点も、子供たちの個々の能力にではなく、むしろ母親など養育者とのかかわりの中で立ち現れ

る“能力”や子供たちを取り囲む養育者のかかわり方に向けられる。その意味で“この子供は〇〇ができる・できない”といった、能力や障害を子供たちに一方的に帰属させた議論は避けられつつある。例えば、一人では上手に遊べない子供たちも、養育者の豊かな意味づけやアシストに支えられて、一緒に遊びを達成できる。また、サポートフルな環境では、障害は障害でなくなることもある。そして、子供とかかわる養育者も、子供たちとの遊びに参加する過程で新たなスキルを見いだしていく。これは“発達における関係論的な視点”と呼ばれている。

これまで高齢者や障害者など情報弱者とコンピュータとのかかわりでも、その能力や障害を使用者・コンピュータのいずれかに帰属させて議論されることが多い。“このコンピュータは〇〇ができる・できない”“この使用者は〇〇が使える・使えない”。こうした大人の視点そのものが新たな情報弱者を生み出しているように思える。

私たちは大地の支えなくしては“歩行”という行為さえも作り出せない。相手の応答なくしては、自分の言葉の意味さえ、自分の中では完結できない。本質的に何らかの不定さ、不完結さを備えており、そのかかわりの中で新たな意味や価値を見いだしていく。これはコンピュータにあっても他人事ではない。自分の中に閉じていては自分の行為の意味さえ知りえない。いつも社会に身を委ねながら、一つのシステムを形作る必要がある。

IT社会に向けたヒューマンインタフェース技術は、ヒトとコンピュータをつなぐための“境界面”としてではなく、むしろヒトと一緒に新たな価値や意味を創りだしていく“関係性そのもの”を議論すべき時代を迎えつつある。効率や利便性だけでは測れない何か。すこし遠回りな作業だが“幼児や子供たちの存在に学び、その遊び心と一緒に競いあう”ことを真剣に考えてみたい。

IT時代のヒューマンインタフェース —現状と展望—



依田文夫*



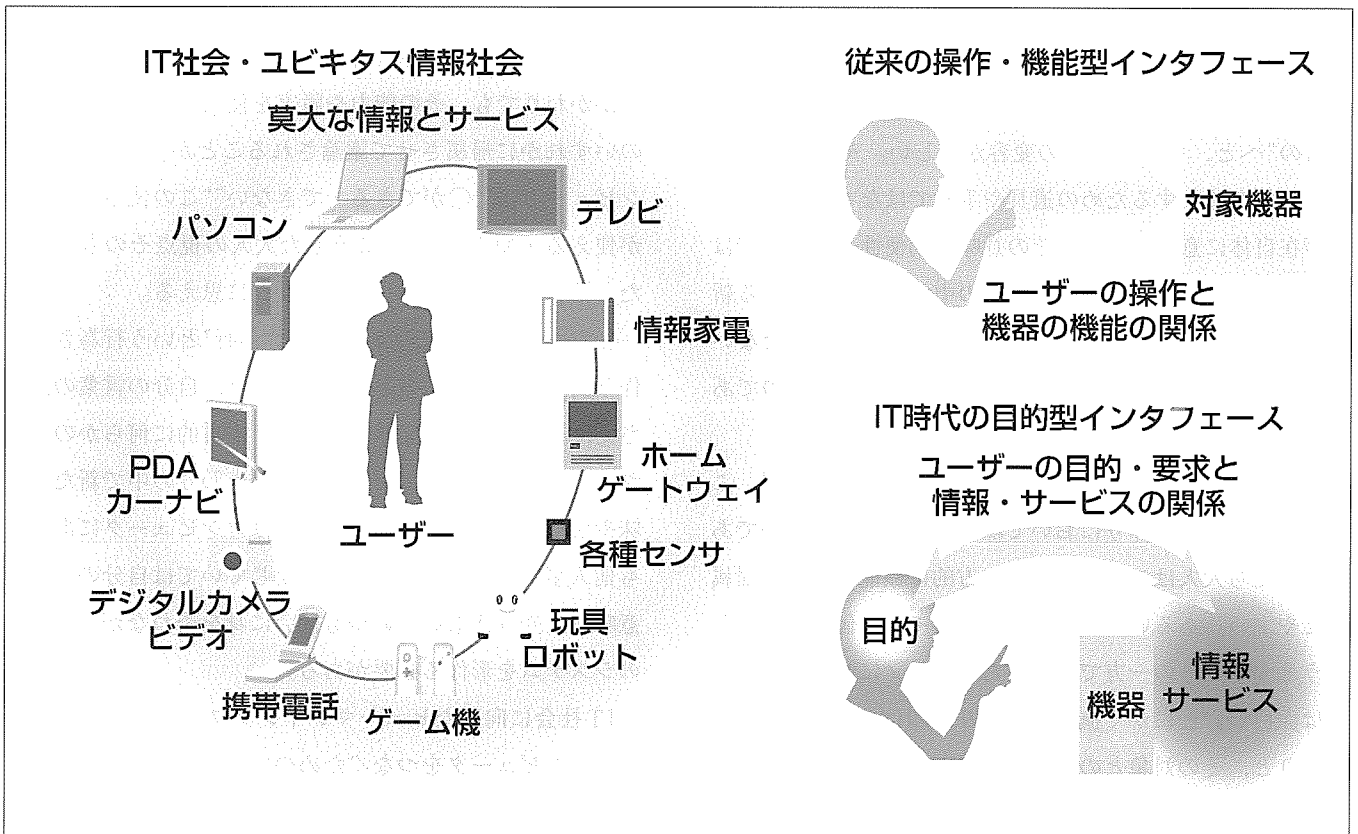
石川 泰**

要 旨

インターネット利用者数の増加，携帯電話の普及など，IT (Information Technology：情報技術) は，“情報社会”と呼ぶにふさわしい広まりを見せている。今後，この社会は，どこにでもコンピュータが存在し，だれでもが情報通信サービスを利用できるユビキタスコンピューティングの世界へと更に展開するだろう。このとき，このサービスがだれにでも真に利用できるものとするには，通信技術，ハードウェア技術，セキュリティ技術などの基盤技術に加え，人間との接点であるヒューマンインタフェース技術が重要となる。どのように有益な情報も，便利なサービスも，機器の使用法が難しければ活用されない。また，莫大(ばくだい)な情報やサービスの海の中に埋没してしまえば，利用者と出会うこともない。

人間が莫大な情報と向かい合うIT時代では，インタフェース技術は，人間の操作と機械の機能を関連付けた従来の操作性という観点では語る事ができず，人間の要求や行動目的と情報や支援という観点で考えなければならない技術である。この人間の要求や目的をいかに的確に把握しそれを支援するかという人間中心の視点に立ったインタフェース技術，それを我々は，IT時代から今後ユビキタス情報通信社会のヒューマンインタフェースと考えている。

この特集では，その基盤技術である音声処理，画像処理，情報処理，人間視点での評価技術を紹介し，それぞれの技術の今後の進展から，三菱電機を目指すヒューマンインタフェースを展望する。



人間と情報の関係とIT時代のヒューマンインタフェース

情報通信技術の進展により，ユーザーは，多種多様な機器と対面し，またそれを通じて，莫大な情報や情報通信サービスを受けられるようになる。このような環境では，機器の種類，場所，時間によらず，必要とする情報やサービスをだれでもが得られる必要がある。すなわち，ヒューマンインタフェース技術は，この人間の要求と情報やサービスとの接点という観点で，新たな時代を迎えていると言える。

1. ま え が き

ITという言葉もやや古臭さを感じるようになったが、国民のインターネットの使用率も半数近くになり、ITは、一部の人が特定の目的に使うための技術から、今まさにだれもが利用する技術となりつつある。さらに、今後は、ユビキタスコンピューティング、すなわち、どこにでもコンピュータが存在し、だれでもが、いつでも、そのコンピュータを活用する世界へと向かっていく。このような世界では、情報通信機器・情報サービスは、ガスや水道と同じように社会生活の基盤となるだろう。

今後の情報通信社会をより良いものとしていくためには、基盤となる通信技術、ハードウェア技術、セキュリティ技術に加え、ヒューマンインタフェース技術の役割がますます大きくなる。有益な情報もサービスも、利用者がそれを享受できるためには、人間と機械の接点又は人間と情報との接点であるインタフェースが、だれにでも使用できる利便性の高いものであることが不可欠となる。本稿では、まずIT時代のヒューマンインタフェースに求められることを従来のインタフェースとの比較で考察し、それを実現する技術を概観し、この特集に掲載した各論文について技術の領域や関係を述べ、最後に今後のヒューマンインタフェースを展望する。

2. IT時代のヒューマンインタフェース

2.1 ヒューマンインタフェース技術とは

ヒューマンインタフェース技術は、機械とその利用者のかかわりに関する全般的・総合的な技術である。狭い意味では、人間にとって使いやすいという意味でのヒューマンフレンドリなインタフェースを意味したり、機械が人間のように振る舞うことを目標とする意味でヒューマンインタフェースという言葉が用いられることもあるが、ここでは、利用者と機械のかかわりのすべてを、人間(ヒューマン)中心の視点で考える技術ととらえる。すなわち、単に入力手段・出力手段の技術又はその機械的な性能の技術ではなく、入出力手段によって利用者は何ができるようになるのか、また利用者にとり機械がどう振る舞うべきかという視点に立脚し、個々の要素技術から評価、デザインに至るまでの技術をヒューマンインタフェース技術と考える。

2.2 機械の進歩とヒューマンインタフェース技術の機能レベル

ヒューマンインタフェース技術は、対象の機械に大きく依存したものとなる。そこで、対象の機械の進展と求められるインタフェースについて説明する(図1)。

(1) 操作型インタフェース

ごく簡単な動作をする機械を考えると、人間と機械の関

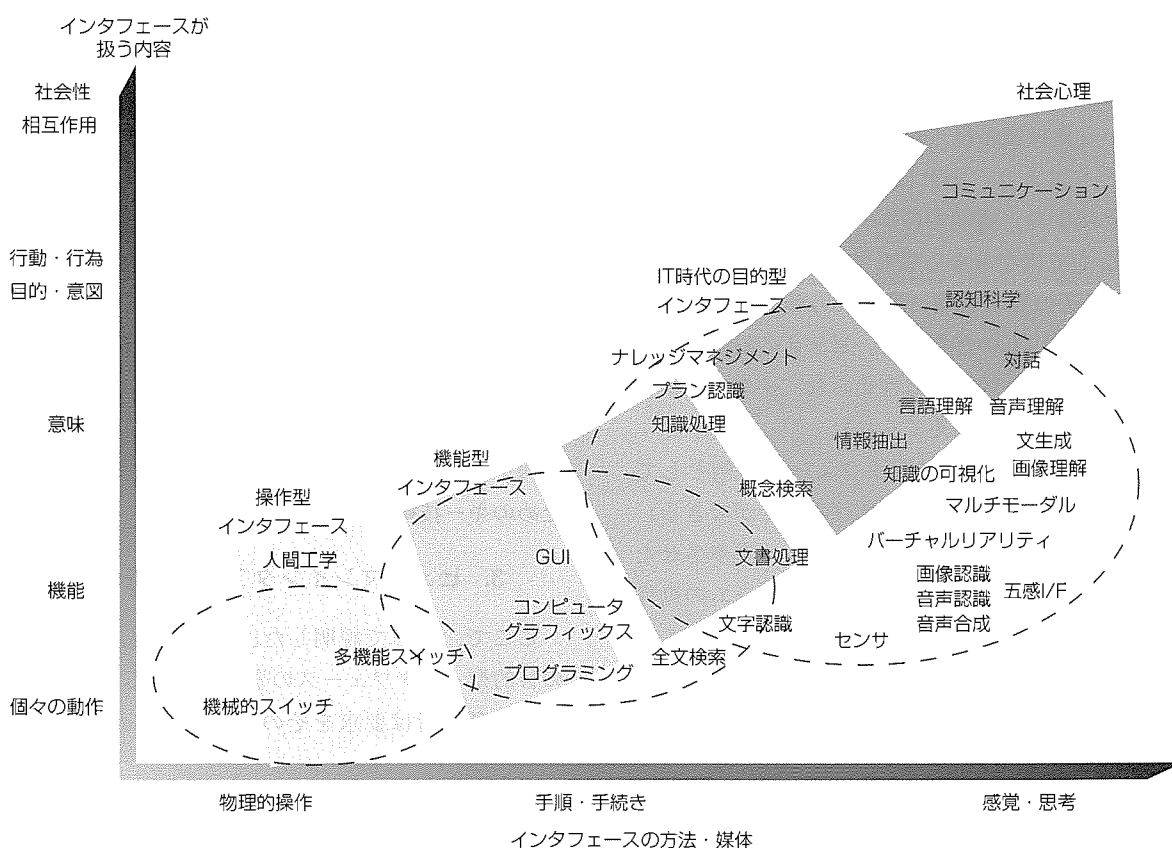


図1. 扱う内容と方法・媒体によるインタフェース技術の分類

係は、人間が個々の機械の動作を指定(入力)し、機械が指定の動作を人間に与えるというものになる。例えば、テープレコーダのスイッチを押すことで再生が開始する、冷蔵庫のドアを引くことでドアが開く、などがこれに当たる。

このような単純な操作のためのインタフェースでは、①通常1対1に対応付けられる人間の操作と機械の動作の関係が直感的に分かりやすいか、②操作が効率的に行えるか、などの観点がインタフェースデザインの基本的な観点である。したがって、このようなインタフェースでは、機器のデザイン、人間工学に基づく設計、評価などをその主な課題としていた。

(2) 機能型インタフェース

機械の進歩によって機械自体の機能や役割が増大してくると、人間と機械の関係は、機械の個々の動作を操作することから、複合的な動作によって達成される機能を利用者が要求し、機械がその機能を提供する関係へと進化した。このインタフェースに着目すると、ユーザーの要求機能を機械の操作の組合せていかに入力できるようにするか、また、制限された物理的な条件の中でどのように入力手段を配置するかなどが課題となった。

例えば、ビデオでは、あるボタンを押すと再生が始まり、あるボタンを押すと停止するといった操作型から、“録画予約”という機能を実現するためのインタフェースが求められるようになった。しかし当初のインタフェースは、録画予約をするためのビデオデッキにとって必要な情報を個々に決められた順序で入力するという構成であったため、多くの人にとり録画予約は極めて使いにくいものであった。この問題は、視点を変えれば、何時に録画を開始し何時に止めるのかという機械側の視点から、ユーザーに入力を求めていたためとも言える。それに対して、Gコード^(注1)が提案されることにより、番組をGコードへの変換するという手順を踏まなくてはならないにしても“この番組を録画する”という利用者の要求機能に近い入力のスタイルになり、大きく使いやすいものとなった。ビデオの録画予約には昔から多くの不満がお客様から寄せられ、インタフェースを考える大きなヒントを与えてくれた。

また、このようなタイプのインタフェースの例としてGUI(Graphical User Interface)を挙げることができる。GUIはコンピュータのインタフェースを大きく進歩させたと言われている。これは、コンピュータプログラムの動作(機能)を直感的に分かりやすいアイコンに割り付けたこと、複雑な入力を階層化して有限のディスプレイに提示しながら入力させることでユーザーの要求を入力しやすくさせることで成功したものである。

このようなインタフェースでは、ユーザーが要求する機能に対して、ユーザーが実際に行う手順がいかに少なく、

(注1) “Gコード”は、ジェムスター社の商標である。

直感的に分かりやすいものが大きなポイントとなる。ただし、いずれにしても、ユーザーは要求に対して機器の操作手順に置き換える必要がある。したがって、使い方に対する学習もインタフェース性能として考慮が必要となる。種々の機器で同一の操作手順であることや、見た目が同一であることが重要である。パソコンのソフトウェアが基本的には同一の操作性を持つGUIで作成されていることは、学習を容易にしたという点では大きなメリットを利用者に与えることとなった。統一性や一貫性もインタフェース開発の大きなポイントである。

(3) 目的型インタフェース

さらに、現在迎えているIT時代の情報機器、さらには将来のユビキタスコンピューティングの時代にはどのようなインタフェースでなくてはならないのだろうか。まず、機械が情報を扱うようになると、それまでの機能(処理)や動作といったレベルではユーザーの要求を示すことが困難になる。インターネットで情報を得ようとするときの人間の要求は、機械の動作や機能レベルで表現できるものではなく、“こういう情報がほしい”というユーザーの行為そのものになるのである。したがって、当然インタフェースもユーザーの目的のレベルで入力が行える必要がある。“三菱電機の新製品の情報がほしい”“この近くでATMのあるところを教えてください”“10万円以下のパソコンを調べたい”，IT社会で情報を扱う機械のインタフェースは必然的にこのような目的レベルの要求を扱わなくてはならないのである。もちろん、これは、Webのサーチに限らず、日常生活の電気製品、将来像としての情報家電でも、機械の機能が複雑になればなるほど、目的レベルでのインタフェースが望ましい姿となる。ビデオでは、録画機能と番組を入力するためのGコードへの変換という手順を踏まずに、“明日の野球を録画したい”という人間の視点での目的を直接扱えることが望まれるようになるだろう。

莫大な情報の中にいるユーザーがだれでも情報通信技術の恩恵を受けられるか、システムが利用者の目的・要求をいかに正確に把握しそれに適合するサービス・情報を提供するか、ヒューマンインタフェース技術がデジタル世界の情報を実世界で役立つ情報や知識として利用者に提供するためのキー技術になるのである。

3. ヒューマンインタフェースを支える技術

ここで、前章で説明したIT時代に要求される情報機器の目的型インタフェースの姿を整理すると、

- 利用者は要求をその目的のレベルで入力できること
- その入力は、いったん機器の操作手順へ変換することなしに直接入力できること。インタフェース操作のための学習や訓練が必要ないこと

- 情報機器の出力は、利用者の要求に応じた内容・形態であること

- 出力は利用者にとり受容・理解が容易であることが挙げられる。“目的型”のインタフェースは、利用者の行動と機械の動作の観点から特徴付けたものであるが、入出力の形態と人間の理解という観点から考えると、インタフェースは“人間の認知過程に適合”したものである必要がある。では、これを実現する個々の技術を見てみよう(図1, 図2)。

(1) 入力手段

入力手段として期待が高いのは言うまでもなく音声認識であるが、上に整理したように、単に音声を利用できるようにしたということだけで良いインタフェースが実現できるわけではない。ユーザーがその目的や要求を機器の操作に置き換えることなく直接入力できることが重要なのである。

この特集では、この技術として、“音声認識技術とその応用”“人物監視技術”を取り上げた。さらには、直感的に理解しやすく多人数の協調作業のためのユニークなインタフェースである“接触認識技術”を紹介する。

なお、人間の要求を機械が把握する上で重要な技術が画像処理・画像理解技術である。利用者の振る舞いを非言語情報としてとらえること、他の情報媒体とでマルチモデルインタフェースを構築することなどで期待が高い。

(2) 目的の推論・文脈の理解

人間がある行動をするとき、1回の行為でその目的が達成できることはあり得ない。複数の行為によって目的を達成する。また、行動目的も、通常は、個々の行動目的を達成することで上位の目的を達成するというように、階層的構造を持っている。人間であれば、相手の個々の要求から上位の目的やユーザーの嗜好(しこう)などまでを推論し、

適切な対応を行うであろう。個々のユーザーの入力からその意味・目的を抽出するだけでなく、一連の要求とその結果である動作、すなわち人間と機械の対話文脈からユーザーの目的を推論し、それに応じた動作をする必要がある。すなわち、インタフェースを考える上では文脈処理、対話管理技術が重要な技術となる。

この特集では、“音声対話技術”で対話処理技術を紹介している。また、“産業用ヒューマンインタフェース評価システム”は評価技術として紹介しているが、その基本にある人間の行動のモデル化は、この一連の動作と目的をモデル化するものであり、インタフェース設計の基盤として重要な技術である。

また、紙面の制約から取り上げることはしなかったが、利用者が明示的に入力しない情報の活用も文脈として活用されるべきである。各種のセンサなどをインタフェースとして活用することが今後の大きな課題である。特に、衛星からの電波を利用したGPS(Global Positioning System)の進歩により、今後の無線LAN環境では、ユーザーが今どこにいるかという位置情報が情報提供の上で重要な役割を果たすだろう。さらに将来は、高度な画像認識システムによって、ユーザーの置かれた環境や状況の把握が行われ、文脈情報として活用されるだろう。

(3) 情報処理技術

ユーザーと機械が目的のレベルで入出力を行うためには、情報を単に文字列や画像、音声信号として扱うのではなく、ユーザーの目的に応じてその内容や意味のレベルで扱えることが不可欠となる。IT時代のシステムでは、情報処理そのものが情報とユーザーの要求する内容や意味を結び付ける技術であり、インタフェース技術として考える必要がある。今回の特集では、その一部として“ネットワーク上の文書テキスト検索・利用技術”“情報化オフィスにおける

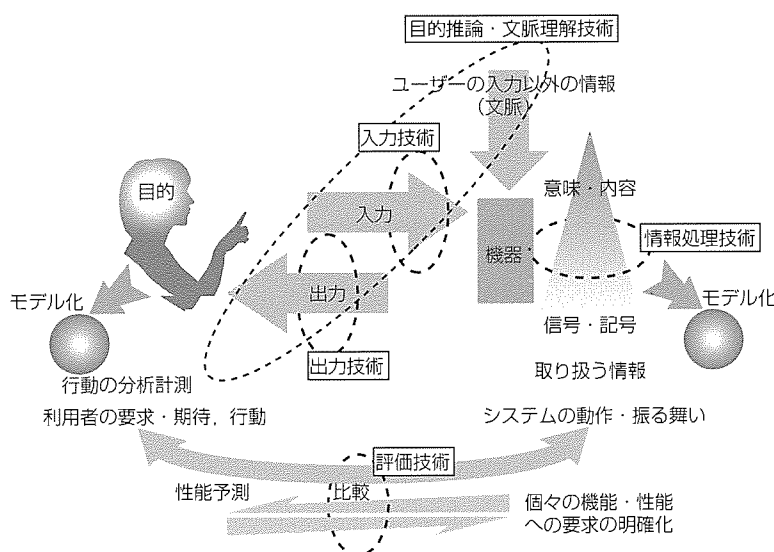


図2. インタフェースの構成に着目した技術分類

文書インタフェース技術”を紹介する。また、画像からの情報の抽出、画像の理解も、画像という信号を意味のレベルで取り扱う技術であり、その一例として“人物監視技術”を紹介する。

(4) 出力技術

ユーザーの理解を高めるために、直感的に分かりやすい形態で情報を提供する技術が要求される。そのためには、基本となる人間の情報認知過程の研究が必要であり、さらにはそれに基づいて実際に情報を提示する情報出力の基盤技術、加工技術、表現媒体の変換技術が必要となる。この特集では、“テキスト音声合成技術”を紹介するとともに、画像出力の基盤となる“携帯端末用“Z3D”グラフィックエンジン”“マルチプロジェクト技術”を紹介する。また、入力技術にも触れた応用の立場から“携帯電話のインタフェースとしての画像処理技術”を説明する。これらに加え、今後は、要約技術、文生成技術、知識の可視化技術への要求も高まるだろう。

(5) 評価技術

上記の技術を統合したインタフェースを人間中心の技術として確立するには、評価技術が極めて重要である。評価は、インタフェースの問題点を把握し、改良するために不可欠である。しかし、評価手法は確立しておらず、また評価尺度を考えること自体も困難な課題である。

ここでは、人間の行動分析に基づく評価と設計技術として“インタフェースデザイン評価技術”を、また、人間の行動を定量的に評価しインタフェースの開発に活用することを目的として“ヒューマンセンシング技術の開発と作業計測への応用”を、さらに、上記(2)でも取り上げた行動モデルに基づく事例として“産業用ヒューマンインタフェース評価システム”を紹介する。

これらの種々の評価手法を目的に応じて選択又は統合するとともに、インタフェース自体のモデル化を今後進めることで、定量的なインタフェース性能から個別の要素技術への要求機能などを明らかにすることが可能となり、インタフェースの開発は更に進歩すると考える。

4. む す び

ヒューマンインタフェース技術を概観し、IT時代の目的型インタフェースとそれを支える技術について述べた。ここでもう一度、情報通信技術そのものの将来を考えてみよう。現在のIT社会は、やや極端な言い方をすれば、高速通信網の整備とCPU、記憶媒体の進展により、“通信ができる”“莫大な情報が利用できる”だけの社会であり、人々の生活の質を向上させる社会基盤という観点ではまだ不十分であると言わざるを得ない。インターネット上にあるデジタル世界が現実世界で生きるためには、常に人間の要求を把握し、それに適合した情報が自然に提供されることが不可欠である。世界中の詳細な地図がインターネット上にはある。しかし、それだけではなんら有益とは言えない。私たちは、現に道に迷い、電車を間違える。“そこを左です”という情報が提供できる情報技術、さらに個人と個人をその要求や興味などで結び付け、場所や時間に束縛されない新しいコミュニティを創造できる情報通信技術、これら人間中心の視点で人と人との関係をより豊かにする技術こそが今後のITと言えるだろう。インタフェース技術はその核となる技術なのである。今後、当社は、更にこの分野の研究開発に積極的に取り組み、ユーザーの立場に立った人間に優しいインタフェース技術の開発に注力していく考えである。

音声認識技術とその応用

岩崎知弘*
花沢利行*

要 旨

音声は人間にとって最も自然な情報伝達手段であり、音声インタフェースは、直感的で分かりやすく利便性も高い。この音声インタフェースとして実用化されている代表的分野には、CTI(Computer Telephony Integration)サーバ、カーナビ、音声ワープロなどが挙げられる。

CTIサーバは公衆向け情報サービスに用いられ、最近、ボイスポータルとしての応用が注目を集めている。このCTIサーバの場合、音声は、電話網を通ることによって帯域制限され、騒音やひずみ加わる。このような電話特有の劣化に対応するための電話音声認識技術について述べる。

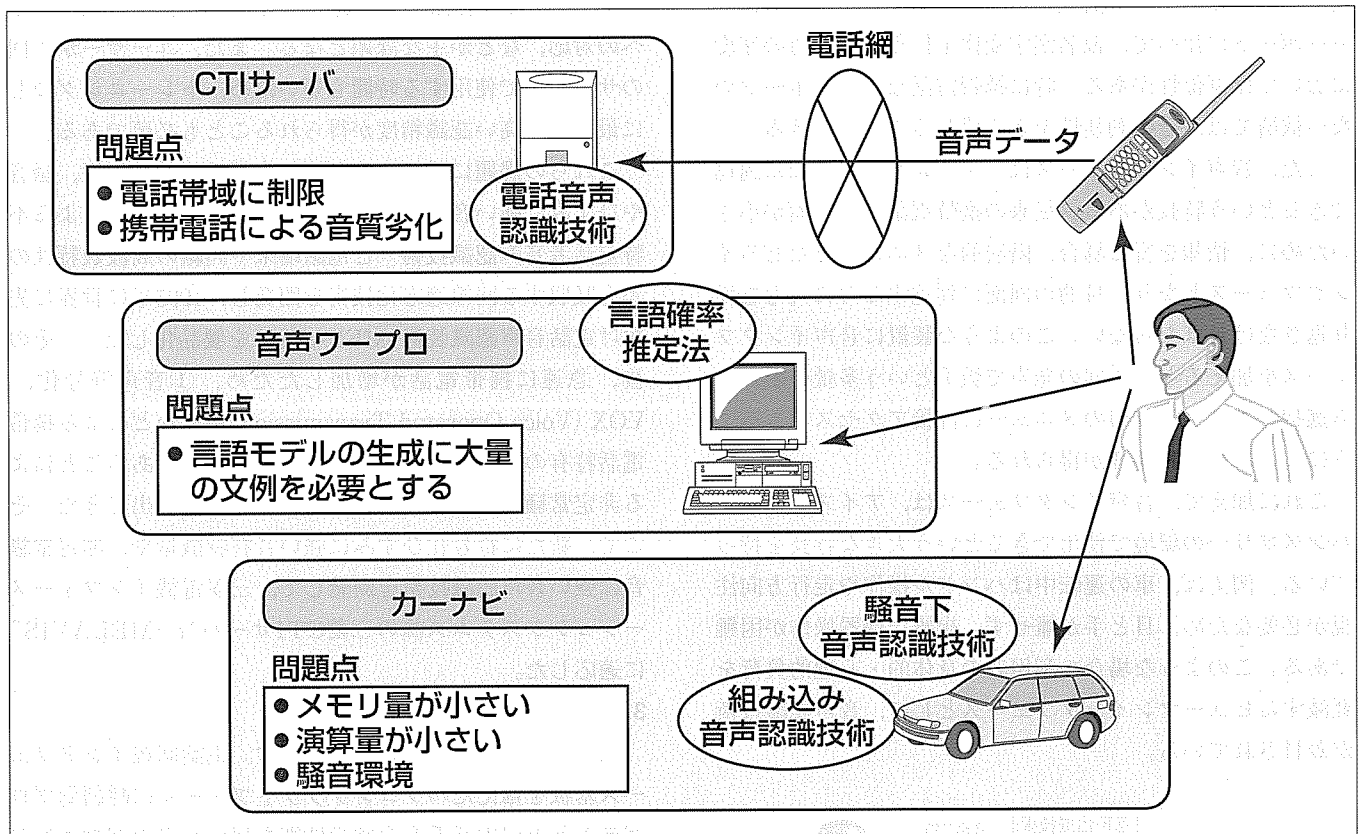
車の運転中は目と手が拘束されるため、カーナビ操作を音声だけで実現できる音声認識への期待は高い。しかし、カーナビは、目的地(住所)入力のための大語彙(ごい)の音声認識、高速走行騒音下の音声認識機能を、演算量やメモ

リが小さい環境で実現する必要がある。ここでは、これらの問題を解決する小さな演算リソースで動作する組み込み用途の音声認識技術と騒音下音声認識技術について説明する。

また、音声入力による文書作成への期待は依然として高い。この音声ワープロで不可欠の言語モデル作成のためには大量のコーパス(文例)が必要であり、従来は特定業務用に対応できなかった。

本稿では、少量のコーパスから容易に業務専用の言語モデルを作成できる言語確率推定法などのディクテーション技術について述べる。

三菱電機は、IT機器の高機能化に合わせて、今後、更に使いやすい音声インタフェースを提供していく予定である。



音声インタフェースの実用例

近年、様々な機器で音声インタフェースを備えるものが増加している。CTIサーバでは、電話音声認識技術を用い、電話網を通して送られてくる音声を認識することによって情報検索が可能となっている。最近増加しているカーナビでは、組み込み用音声認識技術と騒音下音声認識技術を用い、機器制御や目的地設定のために音声認識が用いられている。パソコン上でも音声認識機能によって音声ワープロを実現しているが、少量のコーパスから言語モデルを生成する言語確率推定法により、業務特化の音声ワープロを実現できる。

1. ま え が き

音声は人間にとって最も自然なコミュニケーション手段であり、“いつでも、だれでも、どこでも”使えるという大きな特長を持っている。この音声を用いたインタフェースは、コマンドや検索内容を普段使っている言葉で指定できるため、直感的で分かりやすい。このため、今後迎えるユビキタス環境において、人間の要求を機械に伝える手段として、音声は理想的なインタフェースであると期待されている。

この音声インタフェースは、図1に示すような技術から構成される。人間の発する音声の内容を音響的に識別してコード(認識結果)に変換する音声認識技術、認識結果から発話の意図を抽出する意味理解技術、過去の対話履歴から次に人間に問いかける内容を決定する対話管理技術、及び機械からの問いかけを自然な音声に変換して人間に伝える音声合成技術である。対話管理技術、音声合成技術は後の論文に譲り、ここでは最新の音声認識技術とその応用例について紹介する。

2. 音声インタフェースの特長

インタフェースは、使って利便性が感じられることが必要である。このような観点から、音声インタフェースは、キーボードに比べて、仮名漢字変換を伴う文字入力の手数においては優位性がある。特に携帯電話などキーボードのない機器では、その利便性をより感じる事ができる。

また、音声インタフェースは、ダイレクトに目標に到達できるという特長がある。従来の携帯電話では画面が小さいために、情報を得る場合、階層的なメニューをたどるインタフェースとなり、目的の画面に到達するのに入力を繰り返さなければならない。このような装置に音声インタフェースを加えると、1回の発声で数千という多量の項目から選択したり、下位層のメニューに直接アクセスできるようになり、高い利便性が得られる。

これに加えて、音声インタフェースは、アイズフリー、ハンズフリーの環境で使用できるという大きな特長を持っている。例えば、車の運転中はハンドル操作や走行方向注視が必要なため、目と手が離せず、複雑な機器操作が困難である。このような場合の利用者の身体的・心理的負荷を軽減するヒューマンインタフェースとして、音声認識技術が着目されている。

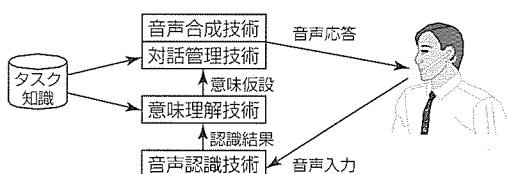


図1. 音声インタフェースの構成技術

3. 音声認識技術とその応用分野

音声認識については、上記の特長から高い期待が寄せられ、1970年代から多くの研究開発が行われてきた。特に、最近の技術の進歩により、音声認識を応用した製品が広く世の中に普及するようになった。代表的な例としては、CTIサーバ、カーナビ、音声ワープロ(ディクテーション)が挙げられる。当社も、これらの分野に向け製品開発を行っている。

以下、当社音声認識技術と応用製品について説明する。

3.1 CTIサーバ

インタラクションの多い情報の検索や予約業務などでは、音声インタフェースの有効性が大きい。このような理由から、人間のオペレータの代わりに機械がユーザーの声を認識し自動応答する電話系音声認識を搭載したCTIサーバを用いたボイスポータルが注目されている。ユーザーはボイスポータルに電話をかけて、天気予報やレストラン情報を音声で検索することができる。

3.1.1 電話音声認識技術

CTIサーバは公衆向け情報サービスに使われるため、音声認識機能としては、①電話回線を介した品質劣化音声(音声周波数帯域の300Hz~3.4kHz制限、電話特有のひずみ)への対処、②様々な話者への対応、③種々の電話端末への対応、などが主な課題となる。また、ユーザーが1回のサービスで使用する時間も短いため、トレーニングなしに最初から高い認識精度が得られることも必要である。

これらの課題に対処するため、当社は、帯域制限、騒音やひずみに強い音響特徴量を用いた音素片モデルによる不特定話者音声認識技術⁽¹⁾と電話端末や回線の周波数特性の差を吸収する回線適応化技術を開発し、1995年に世界に先駆け電話音声認識装置“MELAVIS”を製品化した⁽²⁾。その後、急速に携帯電話が増加したため、①音声符号化、VOX (Voice Operated Transmission)機能などによる携帯電話特有の音質の劣化、②使用環境が屋外であることによる非定常騒音、などの問題に対応する必要が出てきた。そこで、新たに符号化ひずみに強い音響特徴量や、非定常騒音に強い音声認識技術を開発して、三菱電機インフォメーションシステムズ(株)の三菱CTIサーバ“e-MELAVIS”に適応した。

3.1.2 CTIサーバ“e-MELAVIS”

図2にe-MELAVISの構成を示す。電話回線インタフェース基板を備えたパソコンプラットフォームに呼制御プログラムと上記電話系音声認識技術を用いた音声認識エンジンを実装している。1台のe-MELAVISで複数回線の同時認識が可能である。機能としては、同時5万単語認識や住所の連続発声を認識可能とする大語彙認識機能と、ガイダンスの途中で発話の割り込みを可能とするバージョン機能

を備えており、音声入力とプッシュボタン入力との併用も可能である。

現在、e-MELAVISは、携帯電話向け局地天気予報サービス、公共向け自動電話情報サービス、コールセンターの自動一次受付などに用いられている。

3.2 カーナビ

車の運転中は、目と手が拘束され、リモコンやGUIを用いた操作は限定される。この理由から、音声認識のカーナビへの搭載が一挙に加速した。

3.2.1 組み込み音声認識技術

カーナビでの操作性向上のためには、住所などからの目的地設定において、県名から丁目番地号までの住所の連続発声の認識を行うことが必要となる。しかし、CTIサーバとは異なり、カーナビは極めて限定されたメモリと100MHzクラスの組み込みCPUという演算リソースしか使えないという問題があり、大語彙の認識は困難であった。

当社は、この問題を解決するために、予測型探索法を開発した⁽³⁾。この方式では、住所の単語辞書を都道府県や市区町村で構造化した図3に示す木構造のネットワークで構成する。また、特定の県の市、特定の市の町などの単位ごとに約3,000のサブネットワークに分割して記憶している。さらに、メモリが限定されているため、使用頻度の高い前半部はメモリ上に、容量の大きい後半部は補助記憶(DVD)上に記憶している。認識時には発声に応じて必要なサブネットワークをメモリ又はDVDから読み出し、前後の単語の音素の並びを考慮しながら、1音節ずつメモリを確保して認識を行う。このように必要な部分のみをメモリに読み込むことによって少ないメモリで大語彙の認識を

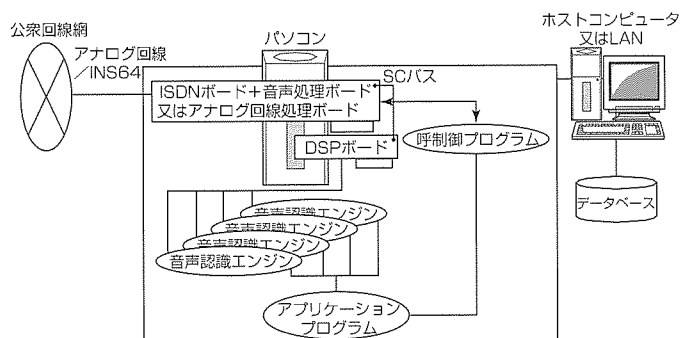


図2. e-MELAVISの構成

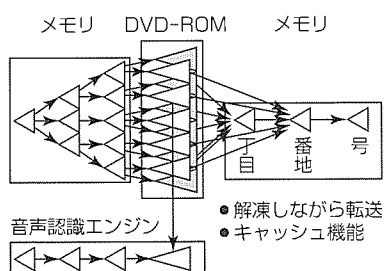


図3. 木構造ネットワークの単語辞書

可能とした。また、単語の間には自動的に無音モデルも挿入することにより、単語間に発声休止があっても認識できるように工夫している⁽⁴⁾。

3.2.2 騒音下音声認識技術

カーナビでは、騒音環境下で上記の大語彙認識を高精度に行う必要がある。このため当社は、車の走行環境に極めて強い騒音下音声認識技術を開発した。具体的には、指向性マイクと低域成分をカットするフィルタリングによって騒音(Noise)パワーに対する音声(Signal)のパワーであるS/Nをまず向上させる。続いて、ノイズレベルの変動をモニタしながら、S/Nに応じて音声区間検出のパラメータを最適化し、ハンドルに備えた発話スイッチ押下後に更に正確な音声区間検出を行う。なお、発声直前の定常騒音を記憶しておき、発声から騒音成分を引き去るスペクトラムサブトラクション法を改良した技術により、定常騒音への耐性を高めている。さらに、ハザードやワイパー音のような周期的に現れるノイズに対しては、周期性をとらえてノイズをカットする方式を開発した。以上のような騒音対策により、様々な走行環境下でも、従来難しかった大語彙の音声認識を可能とした。

3.2.3 音声認識カーナビ“CU-V7000”

上記技術を用いた業界最多の耐騒音型大語彙音声認識エンジンを開発し、当社DVDカーナビCU-V7000(図4)に適用した⁽⁵⁾。指向性マイクはバイザーの先端に取り付け、ハンドルには発話スイッチを備えている。

このカーナビでは、以下の機能を高速走行下で実現している。

- (1) 2,800万件の住所からピンポイントで目的地を設定可能である。県名から丁目番地号までの連続発話入力ができ、任意の位置での発話休止も可能である。
- (2) エリア施設検索では、同じ県内の駅、ホテルなどを施設名で直接発話できる。また県外の施設は“沖縄の万座ビーチ”のように県名に続けて施設名を発話することで、一言で日本全国の地点検索が可能である。
- (3) 住所、施設名、コマンドの同時認識が可能である。メ



図4. CU-V7000

ニューを切り換えることなく1回の発声で要求を実行できる。

3.3 音声ワープロ(ディクテーション)

音声入力による文書作成への期待は極めて高い。しかし、従来のディクテーション技術では、語彙辞書と文法(言語モデル)を新聞やWebなどの大量の一般文例から作成しているため、特殊な用語や言い回しを含む専門文章の認識は困難であり、また、単に専門用語を登録するだけでは認識性能が不十分であった。そこで、専門業務分野に対応した言語モデルを作成する必要があるが、一般に、専門業務ごとの文章蓄積が少なく、有効な言語モデルの作成は困難とされてきた。例えば、専門分野の少量文章から言語モデルを作成する方式の場合、文例に含まれないような表現、個人ごとの表現のゆらぎに対して、認識性能の劣化が著しいという問題を生じる。このようなことから、専門業務用ディクテーションの実用化は進んでいなかった。

この問題に対処するため、当社では、特定専門分野に適応し、しかも発話バリエーションに対して許容度が高い、言語モデルを効果的に生成する新技術“言語確率推定手法”など⁽⁷⁾を開発し、さらに、システム使用中に発生する未知の表現を学習する新方式“未知語学習機能⁽⁷⁾”により、発話の自由度を一層改善した。これにより、専門業務向けの高い認識性能と十分な信頼性を持ち、多種の業務への適応性に富むディクテーション技術を実現した。この技術により、医療、法曹、金融などの専門分野における記録、点検、原稿などの文章を、特別な訓練なく容易に入力できる業務用音声ワープロを実現できる。

4. む す び

音声認識技術は実用化時期を迎えたが、まだ単純な選択機能に音声インタフェースを利用しているに過ぎない。今後は機械も人間と同様に対話しながら問題を解決していくことが必要であり、自然な話し言葉を扱う音声インタフェースを実現する必要がある。現在、この目標に向かい音声認識・理解技術の研究を鋭意進めている⁽⁸⁾⁽⁹⁾。具体的には、自由発話に対応した音響モデル、統計モデルに基づく意味

理解方式などの開発を行い、従来方式では対応が困難であったユーザーの非文法的な会話を含む多種多様な言い回しを扱う音声インタフェースの実現を目指している。

今後、IT機器は更に高機能化し、ユーザーは意識しないで音声インタフェースを利用できるようになると思われる。当社は、IT機器の高機能化に合わせて、今後も使いやすい音声インタフェースを提供していく予定である。

参 考 文 献

- (1) Iwasaki, T., et al.: A Real Time Speaker Independent Continuous Speech Recognition System, Proc.ICPR, 663~666 (1992)
- (2) 青木芳秀, ほか: 三菱音声認識装置“MELAVIS”, 三菱電機技報, 69, No.11, 1010~1015 (1995)
- (3) Iwasaki, T., et al.: A Memory Management Method for a Large Word Network, Proc. EuroSpeech97, 171~174 (1997)
- (4) 岩崎知弘, ほか: 住所認識のための辞書分割法の検討, 音響学会講演論文集, 3-1-1 (1999)
- (5) 岩崎知弘, ほか: 大語彙音声認識技術を用いたカーナビゲーションシステム, 三菱電機技報, 76, No.1, 26 (2002)
- (6) 伍井啓恭, ほか: 音節認識結果における誤り傾向を考慮した未知語学習の一方式, 日本音響学会講演論文集, 3-Q-21, 159~160 (1999)
- (7) 阿部芳春, ほか: 認識誤り傾向の確率モデルを用いた2段階探索法による大語彙連続音声認識, 電子情報通信学会論文誌, D-II, J83-D-II, No.12, 2545~2553 (2000)
- (8) 岡登洋平, ほか: タスクの例文を用いた自由発話音声認識のための言語モデルの構築, 日本音響学会講演論文集, 2-1-13, 73~74 (2001)
- (9) 谷垣宏一, ほか: 最大エントロピー法による発話理解のための効率的モデル構築, 情報処理学会論文誌, 43, No.7 (2002)

テキスト音声合成技術

藤井洋一*
石川 泰*

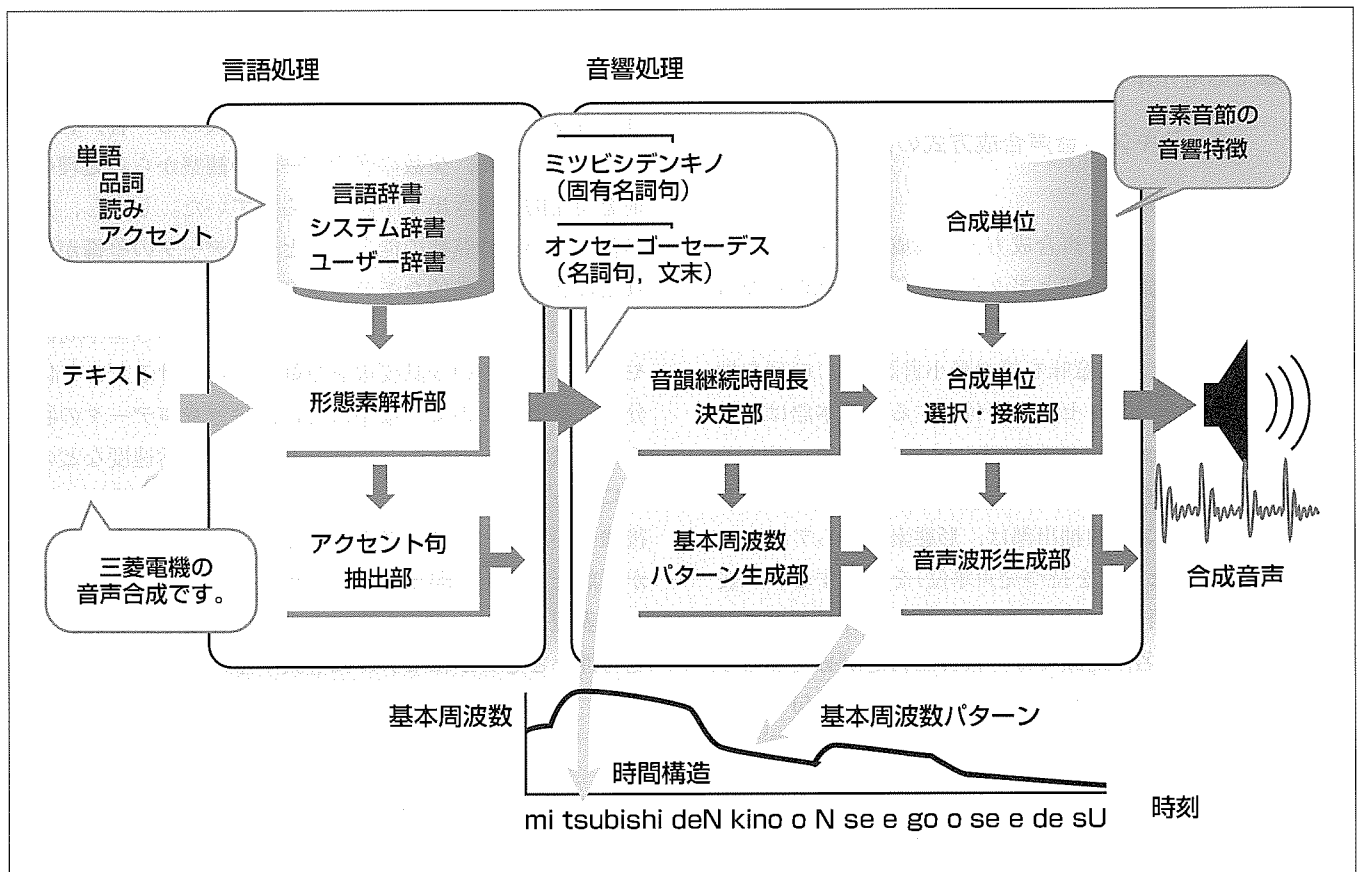
要 旨

任意のテキストを音声に変換するテキスト音声合成技術は、運転中で目が走行方向から離せないカーナビや、ディスプレイが小さい携帯端末向けのインターフェースとして注目されている。また、障害者や老人をIT社会における情報弱者としないためにも、ますますこの技術への要求が高まっている。しかし、テキスト音声合成の品質は、最近の研究の蓄積によって向上しているものの、今後の本格的実用化と利用分野拡大には、更に品質向上を果たす必要がある。

このテキスト音声合成の処理の概要を図に示す。処理は、言語処理と音響処理に大別される。言語処理では、入力文から、読みの単位となるアクセント句と、その読み、アクセント位置、文の構造を抽出する。音響処理では、まず、

基本周波数(声の高さ)のパターンを規則によって生成し、個々の音素の音韻継続時間を決める。次に、音素や音節の音響的な特徴を記憶している合成単位を接続し、合成音声を生成する。したがって、言語処理の性能は、合成音声の読みやアクセントの正確さを決め、音響処理は、自然性、肉声感、明瞭性を決める。このうち、音響処理の品質向上では、生成された音声の特徴がある特定の音声データに近いだけでなく、利用者の要求に応じて発話速度や抑揚のパターンなどの制御ができることが必要となる。

本稿では、聴覚特性と言語構造に着目し、自然性が高くかつ制御が容易な音韻継続時間長制御方式を中心に開発方式を説明し、今後の課題と、応用分野についての展望を述べる。



テキスト音声合成方式の構成

テキスト音声合成では、入力されたテキストを言語処理で解析し、アクセント句(発声上のまとまり)、その読みとアクセント位置、句の言語的属性を抽出する。これら情報から、音響処理は、個々の音素の時間長を決定し、文全体のリズムを生成するとともに、基本周波数(声の高さ)のパターンを生成し、文の抑揚を作成する。さらに、読みの情報から音声の基本的な単位の特徴を接続し、作成した声の高さで合成音声を生成する。

1. ま え が き

音声は人間の最も基本的な情報伝達手段であり、任意のテキストを音声に変換するテキスト音声合成技術への期待は高い。このテキスト音声合成技術は、人間が文を読み上げる過程の解明と模擬の問題であり、聴取・理解過程の研究課題である音声認識・理解と同様、極めて困難な問題である。しかし、生成した音声は人間が聴取するため、人間の言語的な予測などの聴取・理解の能力の高さによる歩み寄りがあり、音声認識よりは早い時代に実用的方式が提案された。ただし、この段階では音質が十分でなく、音声合成を必要とするシステムも多くなかったため、新聞の校閲システムや障害者の補助システムなど、ごく限られたシステムに適用されるにとどまっていた。しかし近年、合成音声の品質向上に加え、電子メールやインターネットなどの大量のテキスト情報が利用可能となるとともに、カーナビのようなディスプレイを注視できない装置や、小型のモバイル端末の普及により、音声出力の要求も急速に高まり、今ようやく本格的な実用時代を迎えつつある。今後、更なる応用分野の拡大のためにも、一層の品質向上が期待されている。

本稿では、2章で音声合成方式の概要と合成音品質について述べる。3章で今回開発した音声合成方式の特徴を示し、4章で応用分野など今後の展開について述べる。

2. テキスト音声合成方式の概要

2.1 合成処理の概要

要旨の図にテキスト音声合成方式の構成を示す。

言語処理は形態素解析部とアクセント句抽出部からなる。まず、形態素解析部は、言語辞書を使い、入力テキストから形態素(日本語として意味を持つ最小言語単位)の列を抽出し、品詞、読み、アクセントを付与する。日本語は分かち書きされず形態素境界が曖昧(あいまい)なため、多くの形態素候補から最も確からしい形態素を抽出する必要がある。次に、アクセント句抽出部は、形態素列からアクセント句(一つのアクセントを持つ発声単位)を抽出する。一般的には文節(自立語、又は自立語と付属語の連鎖)がアクセント句となるが、名詞連鎖では、複合語判定が必要となる。例えば、“音声合成”は複合して一つのアクセント句となるが、“東京横浜”は複合せず、二つのアクセント句で発声される。アクセント句抽出部では、複合語判定と、複合語化及び付属語の接続で生じるアクセント変化の処理を行う。

以上の処理により、言語処理は、入力テキストからアクセント句と、その読み、アクセント位置及び言語カテゴリとその関係を抽出し、音響処理に渡す。

音響処理は、音韻継続時間長決定部、基本周波数パターン生成部、合成単位選択・接続部、及び音声波形生成部か

らなる。音韻継続時間長決定部は、読みやアクセント句の言語カテゴリなどの情報から個々の音素(音声の最小単位)時間長を決める。基本周波数パターン生成部は、アクセント位置と言語カテゴリから基本周波数(声の高さ)のパターンを生成する。さらに、合成単位選択・接続部は、音声合成システムが合成単位として記憶している音素や音節などの単位の音響特徴パラメータを選択し接続する。最後に、音声波形生成部において合成音声生成される。

2.2 合成処理と品質

合成音声の品質は2.1節の各部分の処理性能に依存する。それぞれの処理と合成音声の品質の関係について示す。

言語処理の精度が低いと読みやアクセントに誤りが生じ、言語カテゴリに誤りが生じると文のイントネーションが不自然なものとなる。“前日本協会の”“にわにはにわの”など形態素の境界が曖昧な場合や、“1/2”などのように“二分の一”“一月二日”という複数の読みがある入力文を正しく解析するには、高度な意味処理や文脈処理が必要である。しかし、現状は、一般的な意味や文脈を扱う手法は確立されておらず、統語的な関係や統計的な確からしさのみでいかに正しい解析を行うかが課題である。

一方、音響処理は合成音声の品質を直接決めるものであり、継続時間長規則は文のリズム感、基本周波数パターン生成部は文のイントネーションや意味的なまとまりなどの自然性を決定する。また、合成単位の選択と基本合成方式は、合成音声の明瞭性、肉声感などを決める。音声合成の研究では、当初、少数の音声データの観察から各処理の規則を発見的に導き出す手法が採られていた。しかし、1980年代中盤から計算機的能力向上と音声データの整備が行われたことで、数理的なモデルを音声データによって学習する方法が一般的になった。しかし、生成すべき基本周波数や継続時間とその変動要因を単純に関連付けるだけでは十分ではない。このような手法では、学習音声データの特徴をよく表現するモデルが構築できるが、発話速度などの変更によって著しく品質が劣化するなどの問題が生じる。今後強く要求される利用者の好みや希望に応じた音声の出力が可能な方法を実現するには、単に要因と結果の数値モデルで表現するだけでなく、音声生成機構や言語構想、知覚特性などをいかにモデルに導入するかが重要な課題となる。

3. 開発方式

2章で述べた課題を解決し高品質の合成音声を生成するための言語処理と音響処理について、以下に述べる。

3.1 言語処理

言語処理のうち、形態素解析部では、入力テキストに対して抽出した形態素の接続可能性を評価して形態素列を決定する。具体的には、個々の形態素とその接続にコストを

定義し、コスト最小の形態素列を結果とする方式である。現在、コストは品詞と品詞連続に対して経験的な値を与えている。これにより、品詞の異なる異音同表記語の優先順位を決め、出現が稀(まれ)な付属語を適切に処理するようにし、さらに、コストの違いでユーザーが登録した単語を一般の辞書であるシステム辞書の単語より優先するようにした。また、複合語処理では、単語連続に対する統計的特徴を利用する方式等を検討し⁽¹⁾、これら知見に基づいて、品詞、単語カテゴリー及び単語レベルで複合語化の規則を記述し適用している。詳細な評価は今後の課題であるが、詳細な規則が記述できることで複合語判定精度の向上を達成している。

今後、それぞれの規則の詳細化により、複雑な意味処理や文脈処理なしに、言語解析精度の向上を図る予定である。

3.2 音響処理

音響処理のうち、主に自然性にかかわる韻律制御、すなわち音韻継続時間長と基本周波数の制御手法を説明する。

音韻継続時間長は、主に、以下の要因によって変動する。

- 当該音素、前後の音素環境、文節内位置
- 文節のモーラ(拍)数
- 文節の文内呼気段落内位置、文全体の発話速度

音韻継続時間長の制御方式は、①種々の要因と観測された値の関連を数理的に関連付ける方法と、②時間構造の単位を階層的にとらえ、各要因と階層的な単位の関連を求める方法、とに大別できる。①では、大量のデータから数量化I類などの方法によって比較的精度の良い予測が行えるが、発話速度の変化などへの対応が困難である。一方、②の方法として一般的なのは、モーラを単位とする方法である。日本語では、モーラの時間がほぼ等時間になると言われている。そこで、モーラ時間長を一定とした上で、上記の変動要因をモーラ又は音素継続時間に適用する。しかし、詳細な分析によれば、モーラの時間構造が特殊モーラなどで大きく異なることが分かった。さらに、日本語においては、2モーラを単位とするリズムが指摘されている。そこで今回、この構造に着目し、2モーラを単位とする音韻継続時間長制御方式を提案した。

図1に、提案する階層的な時間構造に着目した音韻継続時間長制御方式を示す。この方式では、文節の平均発話速

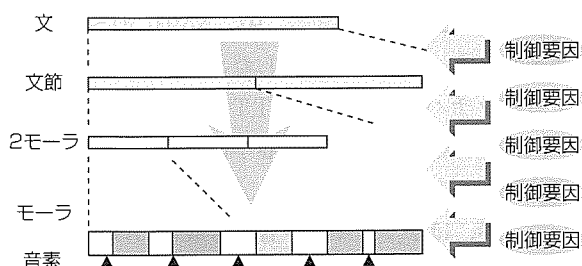


図1. 階層的な時間構造に着目した音韻継続時間長制御

度をその変動要因と考えられる文節のモーラ数、文節の呼気段落内位置、及び文全体の発話速度から求める。次に、文節内の時間構造単位として次の3種類の単位を考え、モーラの時間長を求める。

- (1) 促音、撥音(はつおん)を伴う2音節、長母音音節、“ai”の母音連鎖を含む音節(2モーラ特殊音節: U21と記述)
- (2) 連続する2音節(2モーラ2音節: U22)
- (3) 上記の音節によって孤立する音節(1モーラ1音節: U11)

例えば、“みつびしでんき”は/みつ(U22)/びし(U22)/でん(U21)/き(U11)/という単位の系列となり、“赤い絵の具”は/あ(U11)/かい(U21)/えの(U22)/ぐ(U11)/となる。U22の平均的な1モーラの時間長に対して、U21では短くなり、U11では伸張する明確な特徴があり⁽²⁾、さらに、個々の音素の時間長による変動が上記単位の中で吸収され、2モーラのリズムを維持する傾向が観察されている。

また、図2はアクセント句のモーラ数、呼気段落内文節位置と平均モーラ長との関係を示すもので、アクセント句が発話の時間構造単位であること、及び文レベルのリズム形成に特徴があることを示している。

これらの分析に基づき、提案方式では、アクセント句のモーラ数と、文内位置から句の平均モーラ時間長を決定し、句の読み情報から上記の2モーラを考慮する単位系列に変換し、それぞれの時間長を決定し、U21、U22ではその内分規則によって個々のモーラの時間長を決定する。最後に、子音長を規則によって決定することで、母音長を決定する。この方式は、自然音声の時間構造の特徴をよく再現するとともに、各階層での特徴を独立に制御できるため、自然性

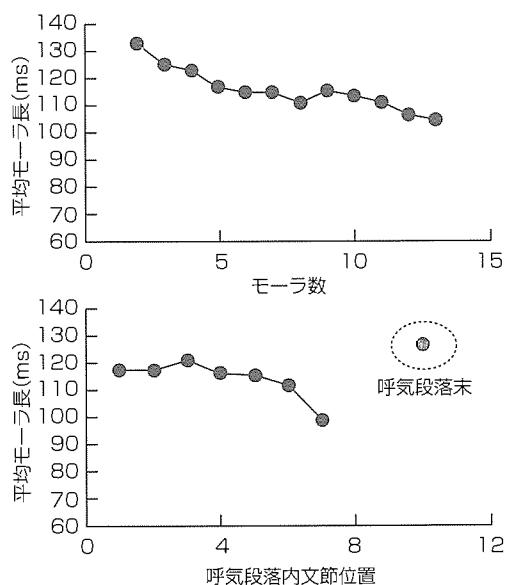


図2. モーラ数、呼気段落内文節位置と平均モーラ長

を損なわずに全体の発話速度や文のリズムなどの発話スタイルの制御が実現できる。

また、基本周波数制御では、アクセント句の言語カテゴリーの局所的な系列によって句の基本周波数を決定する手法を開発した。基本周波数のパターンは、個々のアクセント句のアクセント位置と、文の言語的情報とによって決定する。言語情報としては、文意や係り受け構造が最も重要な情報であるが、実用的な方式を考えると、

- (1) どのような入力でも処理することが不可欠であるため、複雑な解析、結果が得られない解析は行わない。
- (2) 個人性の学習などのため、音声データベースを用いて学習が可能な方式とする。

が必要である。そこで、図3に示すように、文の局所的なN個の句系列の言語カテゴリーから、N番目の句の基本周波数パターンパラメータを決定する手法を提案した⁽³⁾。実験からは、3個の句系列から比較的良好なパラメータ推定が可能であること、その際に、言語カテゴリーとして用法に着目したカテゴリーが品詞カテゴリーよりも良好な結果を示すこと、発声の休止位置推定も高精度で行えることを確認した⁽⁴⁾。

このように、開発した方式では、継続時間長と基本周波数パターン、すなわち、韻律情報の制御において自然音声の特徴をよく再現でき、自然性の高い音声の合成が可能である。さらに、発話構造のモデル化や、学習可能な方法であることなど、今後、発話スタイルの制御など多様な音声を生成するための規則化が容易に行えるという特徴がある。

4. 音声合成の今後の展開

4.1 技術的展開

音声合成技術は、大きく進歩しているものの、自然性や了解性、表現性の目標を人間のレベルとすれば、そのゴールは極めて遠い。しかし、マンマシンインタフェースとして考えれば、目的への適合性、他のインタフェースとの比較が重要であり、特に、情報システムでの出力用途では、了解性が高く、聴取が容易であることが要求される。さらに、聞き取りやすさなどの観点での評価技術の確立も重要な課題である。また、単にテキストを音声に変換するだけでなく、文の要約・生成技術についても開発し、音声インタフェースとしての技術の確立を図る予定である。

エンタテインメント分野などでは、擬人化エージェントなど映像情報との統合や、音質や発話スタイル、感情表現など音声の情報量の豊かさを活用できることが要求される。このような目標に対しては、必ずしも人間の発声を模擬するのではなく、音声合成だからこそできる多様な制御が重要となる。音質や非言語情報と物理的な特徴との関連についての基礎的な研究開発を進めることが基本ではあるが、この関係の解明は極めて困難であり、工学的には、了解性

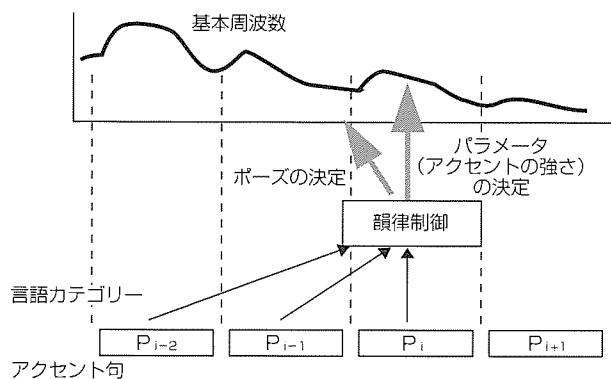


図3. 基本周波数パターン生成方式

や基本的な自然性を損なわずに、外部からの制御要因を増やす方式を開発し、クリエイターなどが自由に音質のデザインを行うことが可能な方式開発が重要である。

4.2 利用分野の展開

カーナビ、携帯情報端末での活用は今後ますます増加すると予想される。特に、ボイスポータルは、この数年のうちに急速にコンテンツを増やし、実用的なシステムとして利用されるようになるだろう。さらに、情報家電やゲーム機、パソコンやホームコンピュータの音声ブラウザ、マルチメディアコンテンツ作成現場や、教育システムなどにも音声合成が利用されるようになるだろう。

5. むすび

今後期待されるユビキタス情報社会では、利用者は、どこでも、だれでもが、その環境で情報技術による支援が受けられるようになるだろう。小型の端末を携帯していれば、行った先々で、親切な仮想的案内役がついていてくれる、そのような社会の実現のためにも音声合成の技術開発を更に進め、豊かな音声インタフェースの技術の確立を図る予定である。

参考文献

- (1) 藤井洋一、ほか：テキストコーパスを用いた複合語アクセント句の決定方法、日本音響学会2001年春季研究発表会、講演論文集分冊I、313～314 (2001)
- (2) 石川 泰：規則合成のための2モーラを単位とする音韻継続時間長規則 — 単位内時間構造の特徴 —、電子情報通信学会技術研究報告、97, No.93, 41～48 (1997)
- (3) 海老原 充、ほか：簡略化したネットワークモデルによるピッチパターン制御とTTSシステムへの組み込み、日本音響学会平成10年春季研究発表会、講演論文集分冊I、219～220 (1998)
- (4) 石川 泰、ほか：テキスト音声合成における統計的言語情報を利用したポーズ予測、日本音響学会2000年秋季研究発表会、講演論文集分冊I、197～198 (2000)

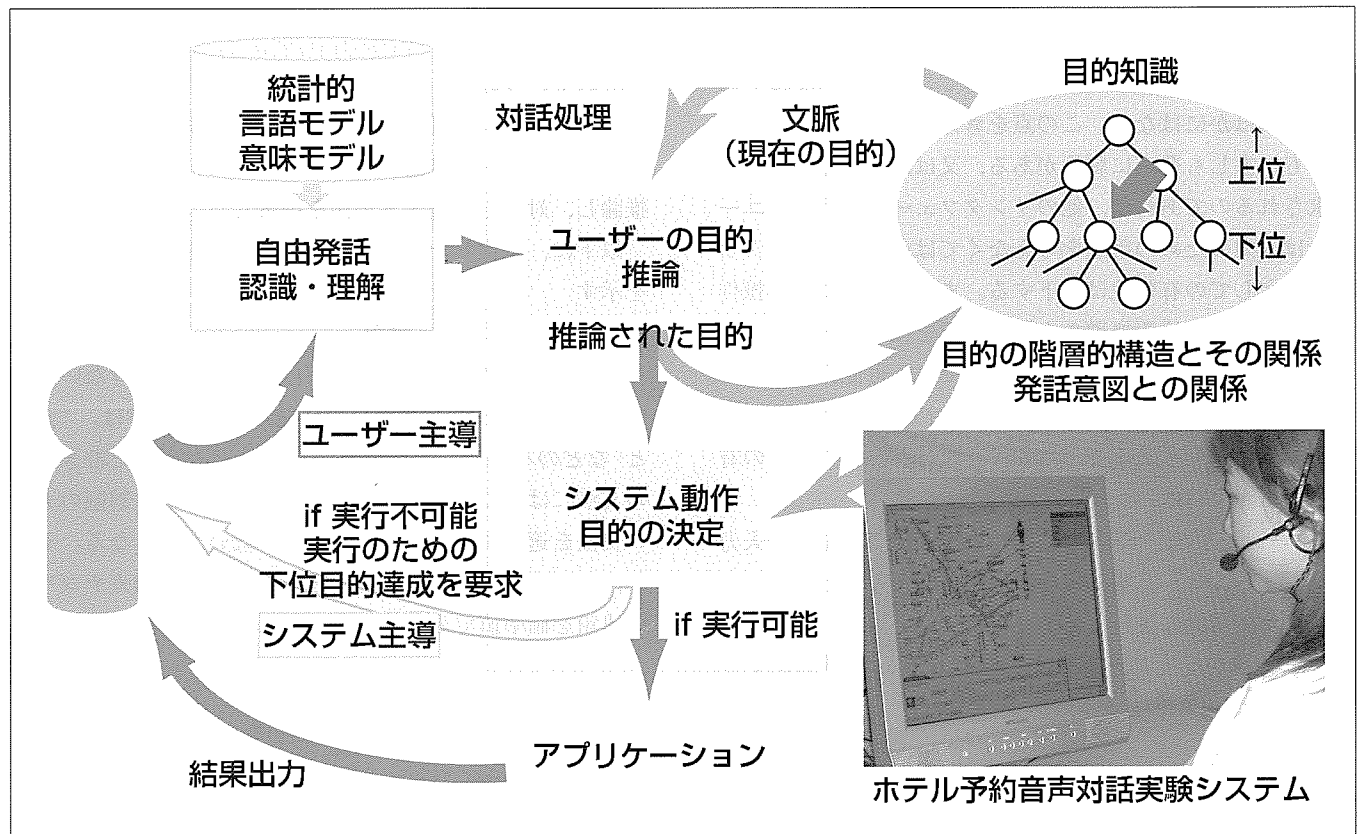
音声対話技術

要 旨

音声は人間にとって最も基本的な情報伝達手段であり、機械を利用する場合も、人間に話し掛けるように意志を伝え会話をしたいという要求は極めて自然なものである。この音声によるインタフェースを構築する上で重要な技術課題が対話処理である。対話処理とは、ユーザーとシステムとの間の入出力を管理し、ユーザーの目的を適切に達成させる対話の維持管理と、前後の発話やシステムの状況に依存したユーザーの発話意図を把握する文脈処理を行うものである。このうち、対話の維持管理では、システムが対話の主導権を持ち、ユーザーに決められた手順で発話を促す方法があるが、利便性が極めて低いものとなる。これに対し三菱電機は、基本的にユーザーに主導権を与え、かつ、目的達成を支援するとともに、文脈処理を実現する対話管

理方式を開発している。この方式では、個々の発話意図からユーザーの対話を通じての上位の目的を推論し、システムはその目的達成のための動作を行う。達成のために必要な下位の目的があれば、システムが対話の主導権をとり、下位目的を実行させることで、効率的に目的を遂行できる。さらに、目的の達成状態が文脈を表すため、文脈処理も同時に実現している。

本稿では、ホテル予約をタスク(対象業務)とする実験システムを例に、この方法を説明する。さらに、関連技術である対話エージェント“COLLAGEN”の紹介を交えて、音声対話技術の今後の課題と、対話を観点とする将来のマンマシンインタフェースについて展望する。



音声対話のブロック図とホテル予約実験システム

音声対話システムでは、自由に発話した会話文を理解するための統計的言語モデル・意味モデルが適用され、入力音声から発話意図とパラメータからなる発話意味が抽出される。対話処理部では、発話意図からユーザーの目的を推論し、それに応じたシステムの動作目的を決定する。システムが動作目的を達成できない場合、必要な下位の目的を、システムは、動作目的として、対話の主導権を得て対話を進める。

1. ま え が き

音声認識システムは、長年の研究により、電話系音声認識装置やカーナビなどで本格的な実用化時代を迎えた。しかし、“人と話すように”を音声インタフェースの目標とするのであれば、まだその道のりははるかに遠く、現状の技術の未熟さを感じている方も多いかもしれない。解決すべき問題は山積しているが、自由に機械との間で音声対話ができたとときの有効性には疑問の余地はなく、今後の情報社会ではその要求がますます強まるだろう。

本稿では、有効な音声インタフェース実現のための音声対話技術について、その課題を明らかにし、開発中の音声対話システムについて説明する。最後に、関連技術の紹介を交え、音声対話から見たヒューマンインタフェースについて展望する。

2. 音声対話システム

2.1 音声対話システムの特徴

音声認識によるインタフェースの特徴については、入力速度が速いことや、特別の訓練が要らないことなどが指摘されているが、最大の特徴の一つは、ユーザーの要求をインタフェースの操作手順に変換することなしに直接入力でき、自然言語で詳細な要求が表現できること、すなわち、情報量が多いことにある。ユーザーが思考を中断することなしにシステムに種々の要求を入力できるインタフェースは音声のほかにはない。この点を考えれば、音声認識を用いても、もしも発声に制約がある、又は、特殊な言い方が要求されるのであれば、他のインタフェースと同様、ユーザーは入力のための手順を考えなくてはならず、インタフェースとしての有効性は低下する。また、単純な機器操作ではなく複雑な情報処理や大量の情報扱うシステムでは、ユーザーの目的が1回の入力で達成できる可能性は少ない。何らかのインタラクションがあって、初めて目的が達成できる。これらを考えると、情報伝達媒体としての音声の有効性を生かすには、

- (1) ユーザーの個々の入力に制約がなく、要求を直接入力できること、基本的に自由発話扱えること
- (2) 複数の発話によって目的を実行するときに、その手順に制約がなくユーザーが主導権を持った対話ができること

が要求される。

2.2 対話処理技術

上記の要求を満足するためには、

- (1) 制約のない自由な会話音声の認識と理解
- (2) 発話の意味・意図の文脈依存性の解消
- (3) 人間と機械のインタラクションを維持管理し、ユーザーの目的達成を支援する協調動作の実現

が不可欠となる。

さらに、音声対話では、

- (4) 音声認識の誤りや、曖昧(あいまい)性に対する対話文脈の観点での妥当性の検証による対処も重要な課題となる。

このほかにも種々の課題があるが、この(2)、(3)、(4)が音声対話の主要な技術課題である。

対話管理を実現する簡単な方法が、状態遷移によるシステム主導対話である。これは、あらかじめ対話文脈を規定しておき、システムが主導権を持ち、文脈どおりに対話を行わせる方法である。規定された状態遷移がゴールに向かう以上、対話は曖昧性なくシステムの目的(ユーザーの目的)に到達する。対話の状態自体が文脈を表すため、文脈処理が容易で、状態ごとに発話が限定されることで認識率の向上も期待できる。しかし、システムが対話の主導権をとることは、ユーザーへの大きな制約となり、利便性が低いシステムとなる。目的が単純でユーザーの入力も規定しやすいタスクであれば問題は少ないが、複雑なシステムでは、①不要な入力を要求される、②システムの動作が予想できない、など問題は大きい。

図1に、音声対話の状態遷移記述に基づく制御の概念を示す。

3. 目的推論に基づく音声対話管理方式

そこで、今回は、ユーザーの個々の発話目的から対話全体を通じた大局的目的までを階層的に表現した知識源を構築し、ユーザーの発話に制約を設けることなくその目的を推論し、対話を管理する手法を提案した。

以下に、試作したホテル予約システムを例に、その概要を示す。

3.1 自由発話認識・理解

音声対話システムでは、発話自体に制限を設けないことが望ましい。ところが、自由に発声した音声では、“えーと”などの不要語、言い直し、言いよどみ、また書き言葉の文法には当てはまらない文が頻発するため、制限の大きい文法を適用すると認識率は極めて低いものとなる。そこで、会話文データを収集し、その単語及び単語カテゴリーの連鎖を確率値で表現する自由発話認識向けの言語モデル

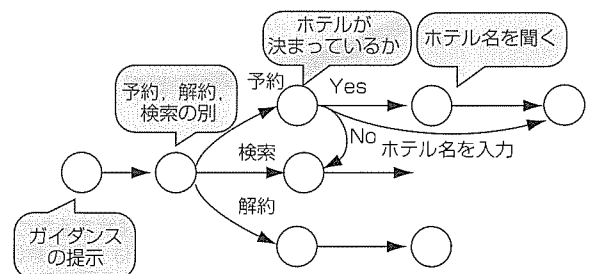


図1. 音声対話の状態遷移記述に基づく制御

を構築した。さらに、単語系列から意味を抽出する部分でも、同一の意味が多様に表現されることを考慮し、統計的な手法に基づく意味抽出法を開発した⁽¹⁾。この方法では、文に一つの単語又は単語の組が観察されたとき、その文がある文意を表現したものと判断することの尤度(ゆうど：確からしさ)を表すモデルを用いる。観測された単語系列にこのモデルを適用し、最も高い尤度で予測される文意を意味理解結果とする。モデルは、対話コーパス(対話文例集)から学習によって獲得する。すなわち、ホテル予約についての会話データを収集し、各文に人手によって文意とそのパラメータである意味表現を付与しておき、単語系列としての文と、意味表現の間の有意な関係を最大エントロピー法によって学習する。この方法で、多様な表現に対して頑健性の高い理解方式を実現できる。

3.2 対話処理

対話処理では、対話全体の管理と文脈処理が行われる。音声理解の結果得られた発話意図から、階層構造と相互関係を持つ知識として表現された目的に関する知識源を適用し、ユーザーの目的を推論する。図2にホテル予約タスクの目的の一部を示す。図において、上に記されているものが上位の目的であり、線で結ばれている目的は直接関係がある上位下位の目的を表し、さらに、その関係は目的達成のために必須(ひつす)の下位目的か、関連の下位目的かで分類されている。発話意図と目的の関係は、意図がその目的を表現するものとしての妥当性を表す値として与える。この値は、本来は対話コーパスの分析などから確率値として与えるべきであるが、現在は、表出するものか、しないものかの1/0の値を経験的に与えている。目的の推論は、目的の活性度を求め、一定の活性度以上の最も上位の目的がユーザーの目的であると判断する。

発話意図が入力されると、発話意図に直接対応する目的の活性度が増加するとともに、図3に示すように、上位の

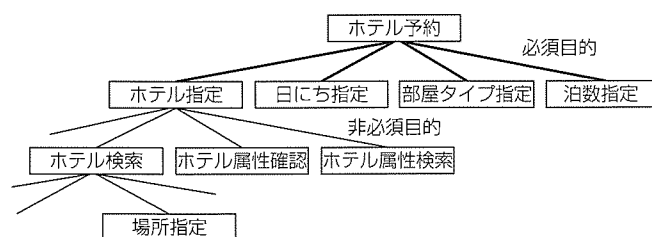


図2. ユーザーの目的に関する知識表現

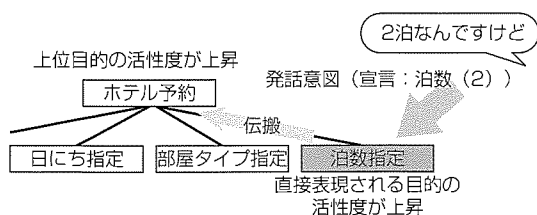


図3. ユーザーの目的推論のための活性度の伝搬

目的に活性度の伝搬が生じる。これにより、下位の目的の実行が続くと、上位目的の活性度が高まり、システムは、発話意図が直接表す目的ではなく、上位目的をユーザーの目的として認識し、その達成のための動作を実行する。これにより、例えば、“Aホテルなんですけど”という“ホテル名を宣言する”という発話は、他の目的の活性度が低ければ、直接の目的である“ホテル名の獲得”がシステムの動作目的となり、単に“Aホテルですね”又は“はい”とホテル名を受理する。一方、ホテル予約が目的として活性化されていれば、“予約”を動作目的とし、実行しようとする。さらに、このとき目的の達成状況を参照し、必須の下位目的が達成されていない場合は、下位目的の実行を動作目的とするため、システムは“ご宿泊はいつですか”という“日にちの獲得”を実行する。このように、この方法では、対話主導権が適切に交替し、目的達成のための協調動作を実現するとともに、発話の文脈依存性も解消している⁽²⁾。

3.3 システムの評価

システムの問題点を抽出するための小規模な評価実験を行った。システムを初めて使う被験者に、あらかじめホテル予約の状況を設定してもらい、予約を行わせた。定量的な評価は今後の課題であるが、以下の知見が得られている。

- (1) 想定した文意を表す音声では高い理解率が得られているが、学習コーパスにない文意が現れ、理解誤りの大きな要因となった。
- (2) 対話管理部で上位目的の推論を行わない場合、すなわち、活性度の上位伝搬を行わない場合は、対話達成率が低下し、手法の有効性を確認した。

4. 音声対話システムの今後の展開

4.1 今後の課題

自由発話の音声対話システムは、音声理解のためのモデル作成に目的のタスクのコーパスが必要であること、実用的理解率・タスク達成率が得られていないことなど、実用化には更に研究開発が必要である。課題を以下に示す。

- (1) タスク不依存の意味理解の枠組みとタスクの知識の記述・獲得方式
- (2) 目的の関連知識と発話と目的の相互関係、上位目的の推論方式の一般化と学習
- (3) 文脈妥当性検証による意味理解率の向上⁽³⁾
- (4) システムの評価手法の確立と、要求性能からの必要となる音声理解率や、対話管理部の性能の予測

4.2 関連研究

当社では、人間と機械の協調作業を支援する対話エージェントCOLLAGENの研究を進めている⁽⁴⁾。これは、図4に示すように、人間と機械の間の動作をエージェントが観察し、ユーザーのプラン(目的)を推論し、質問・要求にこたえるシステムである。プランは、レシピと呼ぶ知識源を

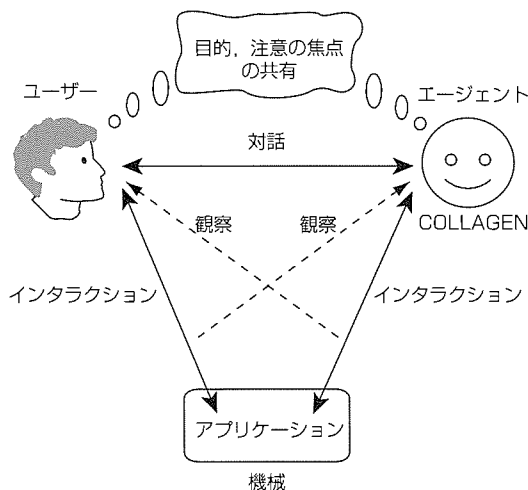


図4. COLLAGENの概念

用いて推論される。レシピには動作AとBとで、動作Cが実行されるといった論理構造でタスクの目的が記述され、認識における曖昧性はCOLLAGENがユーザーに質問することで解消する。COLLAGENは、ユーザーの操作についての質問に答える、又は、ユーザーに代わって操作を行うことができ、協調動作のシステムが実現できる。簡単な論理式でプラン認識を行うことなどから、余り複雑なタスクへの適用が困難などの問題があるものの、アプリケーションへの組み込みやレシピの記述が容易であり、今後、操作手順のガイダンスを行う支援エージェントや、訓練システムなどでの実用化が期待できる。

4.3 今後の展開

複雑な業務を制約のない日常会話で行うことができるシステムの実現には、まだ時間が必要である。しかし、音声対話は、今後の情報社会で、だれもがストレスなしにシステムを利用するための必須技術である。限定されたシステムから、徐々に制約をなくし、対象業務を拡大していきたい。さらに、対話処理技術は、音声インタフェースに限らず、複雑な処理を行うシステムのインタフェース設計技術・評価技術としても重要である。

例えばGUIは、対話の観点から見直せば、システム主導対話である。対話の状態、システムが実行できる機能を常に明示する自己表現性の高さが特長ではあるが、入力制約が大きく、インタラクションが増える。そこで、機能の

分類で配置されることが多いメニューを目的達成の過程の観点で見直す、又は、文脈処理を導入し、状況に依存したメニューの構築などによる改良も可能である。将来は、システムの動作や利用者の行動目的記述から最適な対話インタフェースを自動構築する設計技術への展開も期待できる。

一方、現状の音声認識応用システムでは、認識性能よりも“何を言っているかわからない”ために起きる“語彙(ごい)外発話”が問題となることが多い。すなわち、音声認識の導入で自由度は向上したが、自己記述性が低下している。自由度、自己記述性、一貫性、可学習性など、個々の特徴の定量化、さらにはそれらとインタフェース全体の性能の関係を定量化する評価手法が対話のモデル化によって開発できれば、設計者の考え方に依存している対話アプリケーション評価に基づく設計指針を導入できるだろう。

5. むすび

音声対話技術について紹介した。今後は、人間の認知過程、行動、ユーザーが期待する機械のモデル化などの観点からの検討も進め、自由発話音声対話システムの実現を目指す。さらに、エンタテインメント性など、目的達成以外の対話の意味を考えることも必要であり、種々の分野の技術を統合して人間中心の視点からより良いインタフェースの製品が提供できるよう研究開発を進める予定である。

参考文献

- (1) 谷垣宏一, ほか: 最大エントロピー法による発話理解のための効率的モデル構築, 情報処理学会論文誌, **43**, No.7, (2002)
- (2) 渡邊圭輔, ほか: ユーザの行動目的の推定による協調的な音声対話制御, 情報処理学会第54回全国大会, 講演論文集(2), 233~234 (1997)
- (3) 渡邊圭輔, ほか: ユーザの行動目的に基づいた音声対話システム, 情報処理学会第60回全国大会, 講演論文集(2), 137~138 (2000)
- (4) Rich, C., et al.: COLLAGEN: A Collaboration Manager for Software Interface Agents, User Modeling and User-Adapted Interaction, **8**, No.3/4, 315~350 (1998)

人物監視技術

ポール・バズリー* 鷺見和彦***
 レシエク・チプリンスキー**
 ポール・ピオラ*

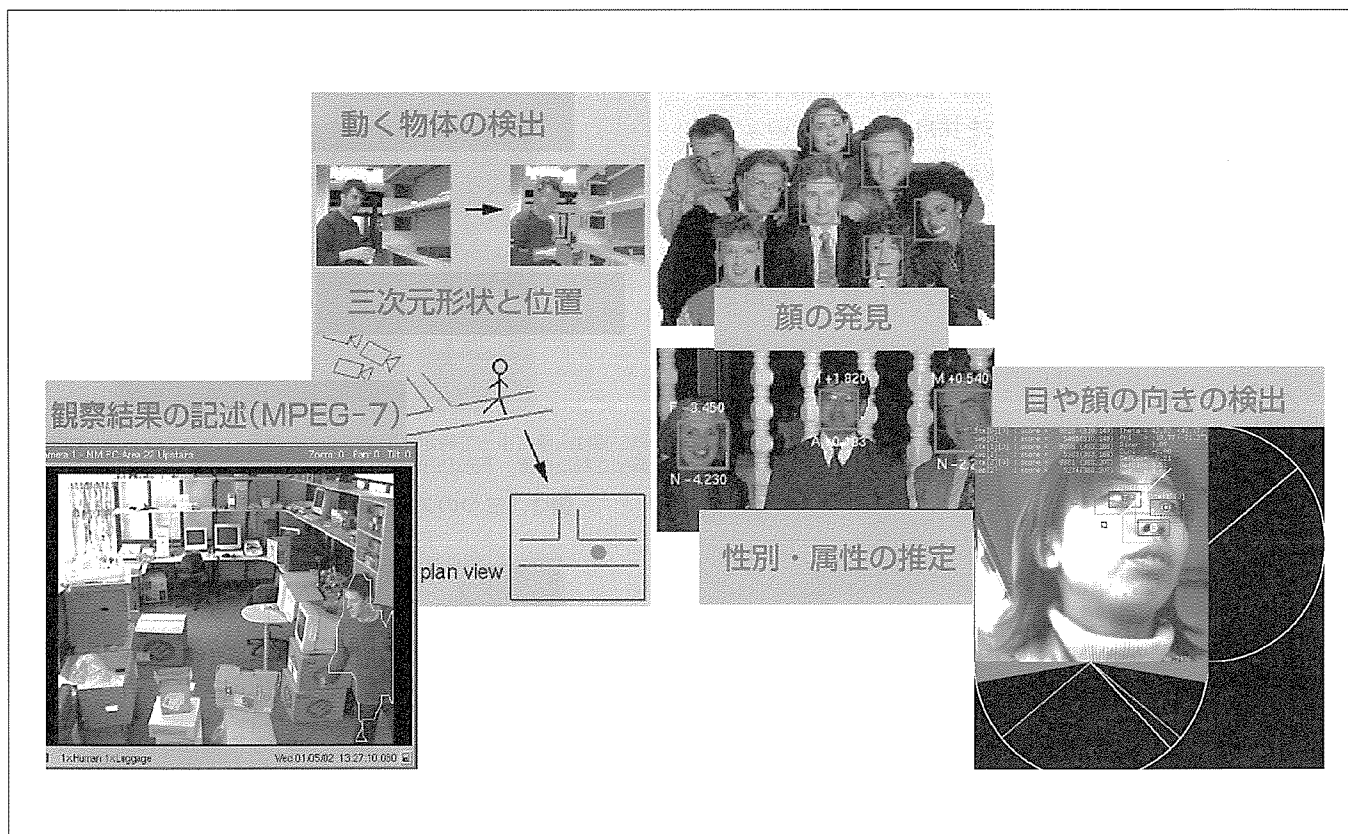
要 旨

カメラで撮影された人物を発見・識別し、その行動や意図を読み取ることは、機械から人への自発的コミュニケーションや人対人の非言語的コミュニケーションを自動化・支援する際に重要である。本稿では、このような機械が人を見るための技術を紹介する。

通常、人から機械への情報伝達では、人が意識的に通信を開始しなければならない。一方、人対人のコミュニケーションでは、通信の開始は相互に自発的に始められる。すなわち、受付や店舗では来訪者に対して挨拶(あいさつ)することから会話が始まり、警備や介護においても手助けが必要な状況をサービスする側が自発的に認識する。この

ような自然でさりげないコミュニケーションの開始を機械にも可能にするための技術を紹介する。

まず、1章では、人物の発見とその位置、形状と軌跡を高精度に認識するためのステレオ視覚システムを紹介する。次いで、2章では、マルチメディアの構造化と検索の標準MPEG-7に準拠した、人物の形状や色による特徴記述技術と検索技術を紹介する。3章では、より詳しく人物を観察するために映像から顔を抽出する技術について紹介する。最後に4章では、人の注意や意図を読むために目を発見し、注視方向を推定する技術を紹介する。



人物監視技術の構成要素

人物監視技術の構成要素には、動く物体の検出、顔と人物の検出、性別・属性の推定、形状と位置の解析、目・鼻・口などの顔の部分の検出及び人間の行動の記述などがある。

1. ま え が き

従来、ステレオビジョンは、複数のカメラを用いるため、コスト、演算量、消費電力が増大しやすく、監視システムにはほとんど使われてこなかった。しかし、電子デバイス技術の進歩によって問題が解決されつつあり、今後は普及していくと予想される。監視カメラの画像処理に求められる共通機能の一つとして人の検出と識別、行動の解析があるが、三次元の映像データから形状・身長・体型を計測することにより、単眼カメラには不可能な、正確な人の検出(車・動物等との区別)や詳細な行動の解析が可能となる。

1.1 ステレオシステム

ステレオ(立体視)を使用した監視システムの一例を紹介する。代表的なハードウェアは3台のカメラを用いるが(図1)、2台以上カメラがあれば構築することができる。

1.2 ステレオ法を用いた画像処理の手順

この節では、ステレオカメラシステムを用いて人を検出しその行動を解析する手順について説明する。

1.2.1 ステレオカメラ校正

実際の処理を始める前にオフラインでカメラ校正を行う。校正用チャートを視野の中に保持し、チャートの検出と校正を実行する。校正チャートは地面にも設定し、対象物の高さやカメラ基準座標の位置を計算する。

1.2.2 人物候補の抽出

特徴検出の目標は、人又は物を高い信頼性で検出することである。ここでは、背景除去法を採用して、背景のエッジマップと監視実行中の画像のエッジマップとの比較を行った。一度前景(新しく現れた物、又は動いたもの)のエッジが検出されると、それらは直接ステレオ再生アルゴリズムの入力特徴となる。図2に単一のカメラに対して検出さ

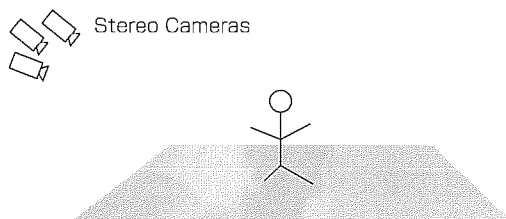
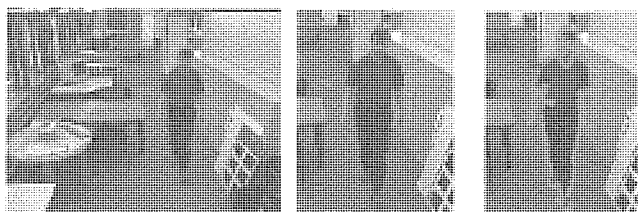


図1. 情景の三次元データを撮るステレオシステム



(a) 全体画像 (b) 対象物についてのズーム部分 (c) 特定された前景エッジ

図2. 映像中の動きの検出(単一カメラでの結果)

れた前景エッジの例を示す。

1.2.3 三次元形状の再構成

三次元形状の再構成は、傾きの方向やエッジの片側の色といったエッジの属性を使って、ステレオ画像間のエッジマッチングを行うことから始める。マッチしたエッジは次に3D再生のために用いる。

1.2.4 人の動作解析

この項では、再生した3Dデータの解析例を紹介する。図3は、動いている対象物の三次元的位置を背景の二次元マップ上に表示する応用例である。これは、立入禁止区域や危険な場所への侵入者有無などを調べるのに使うことができる。

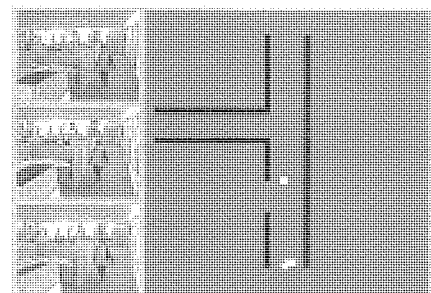
図4の上段は左から人物が立っている、座っている、横たわっている状態の画像である。同図中段は対応する3D再生、下段は再生3Dデータの正面への直交射影である。解析の結果、対象物の“直立性”が明らかにできる。これは、介助が必要な負傷者認識などに応用できる。

2. MPEG-7技術と監視システムへの応用

この章では、MPEG-7で標準化された技術を用いて人を発見したり映像を検索するセキュリティ映像レコーダについて紹介する。

2.1 MPEG-7技術

MPEG-7は、マルチメディアコンテンツを記述するた



黒い線は人手によって特定された映像中の境界、中央の白い点は対象物の算出された位置、下方の白い点はカメラの位置を表している。

図3. 三つのステレオ映像と再生映像の見取図

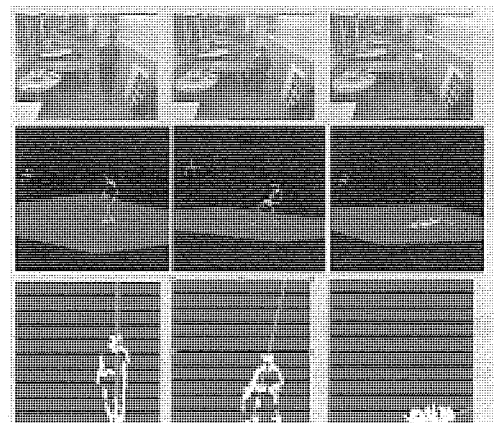


図4. 人間の姿勢の識別

めの世界共通の標準規格であり、マルチメディアコンテンツの構成要素を記述するメタデータと呼ばれる概念情報を画像に付加することにより、広範囲なマルチメディアのアプリケーションへの要望にこたえるものである。このうち、特にモニタリング及び監視の分野に関連する技術はVisualパートに含まれる。

この章では、MPEG-7の色、形、そしてマルチプルビュー(複数視点画像)の記述子における技術について説明し、これを利用したビデオ監視システムについて紹介する。

2.2 ドミナントカラー記述子の概要

ドミナントカラー(Dominant Color:代表色)記述子は、画像や画像領域の代表的な色をコンパクトに記述するものである。主な応用分野としては、色に基づいた画像データベースにおける視覚的類似による検索・ブラウジングなどがある。従来のヒストグラムに基づく記述子とは異なり、代表色は特定の色空間で固定されたものではなく、それぞれの画像から計算される。そのため、色の記述が正確かつコンパクトになり、大きなデータベースのインデックスとして有効に利用することができる。

ドミナントカラー記述子は、基本的には画像や画像領域を構成する代表的な色と、それらの色が存在する頻度の組合せで定義される。高度な類似性の一致が要求される場合には、距離類似度や色分散などの情報を併用して性能向上が可能である。ドミナントカラー記述子の構成を図5に示す。

2.3 カンターシェイプ記述子の概要

カンターシェイプ(Contour Shape:輪郭線形状)記述子は、物体や画像の一部の輪郭を基にその形状的な特徴をとらえるもので、曲率多重解像度空間表現が用いられる。この方法は、人間が形状を認識するときの情報処理によく似ており、特定のデータに依存することなく形状固有の特徴をとらえることができる。対象があいまいな動作をする場合や輪郭が部分的な閉空間である場合、又は遠近の変化を伴う場合への対応に優れる。

カンターシェイプ記述子は、曲率多重解像度空間において、対象となる曲線の一連のピークがなくなるまで低域フィルタを繰り返し適用し、その反復回数を得ることによって求められる。これらの反復回数が、対応するそれぞれの

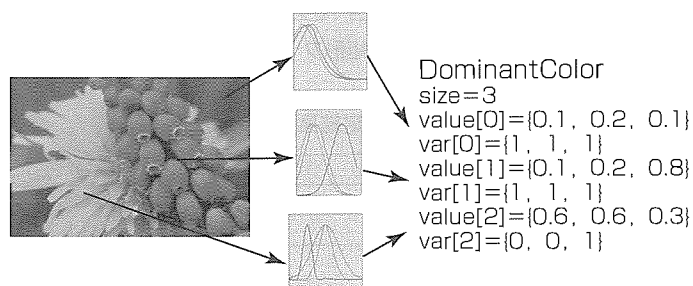


図5. ドミナントカラー記述子の例

ピークの位置とともに、カンターシェイプ記述子として保存される。

2.4 マルチプルビュー記述子の概要

マルチプルビュー(MultipleView:複数視点)記述子は、三次元の物体を様々な方向から見た場合の、視覚的特徴を表す二次元記述子の組合せによって構成され、物体の三次元視覚表現を形成している。ここで、二次元記述子は、カンターシェイプ(輪郭形状)、リージョンシェイプ(領域形状)、色、テクスチャ(模様)など、いずれも使うことができる。この記述子により、三次元の物体のマッチングを視覚的比較、又は二次元的な視点に立った表現法による比較という両面での比較が可能となる。

2.5 知的画像監視システムVISS

(Video Intelligent Surveillance System)

VISSは、MPEG-7の画像処理技術を応用して開発された画像監視のデモシステムである。このシステムにより、広範囲にわたる監視に必要な警備要員を削減することが可能となる。VISSは、複数のカメラからの入力画像をリアルタイムで解析し、侵入者が検出された場合は、監視者(警備員)に警報を発する。さらに、人間を犬、猫又は荷物などと区別することができ、誤った警報を最小限に抑えることができる。図6の右上の画面に、自動的に人間を検出する様子を示す。この装置では、侵入者が検出されると、画像データとともにMPEG-7の形状と色の記述子が抽出され記録されるので、監視者は、蓄積された長時間の画像データの中から、形と色の特徴に基づき、特定の侵入者の画像を素早く検索することができる。

3. 人の顔の検出と分析

3.1 統合的な人の顔の検出と分析

この章では、人の顔を自動的に検出・追跡する技術を紹介する(図7)。この技術によれば、一般的ノートパソコンにおいても毎秒10~15フレームの処理速度が得られ、インターネット上の様々な顔画像データを用いて処理性能を確認済みである。



図6. VISS監視システム動作中の画面

性別分類に関する従来技術としてはサポートベクトルマシン(Support Vector Machine : SVM)と呼ばれる方式があった。これを目と口を基準に顔の位置と大きさを補正したFERET(FacE REcognition Technology : 米国陸軍研究所による顔認識技術の開発プログラム)データでテストした結果、誤り率は3%であった。

従来技術の自動顔抽出法と組み合わせると、顔の位置と大きさの補正機能が不十分なため、多数の判定特徴(サポートベクトル)を投入しても顔認識誤り率は23%に上昇するが、今回の手法は、同等の性能(顔認識誤り率22%)を1,000倍以上高速に実現できる。

この技術の特徴は次のとおりである。

- (1) 顔の検出に有効な新しい画像特徴の集合を用いて、外観モデル(標準顔)を不要にしたこと
- (2) 画像特徴の組合せをAdaBoostという学習アルゴリズムで最適に決定したこと
- (3) 画像特徴の適用を直列化し、その順序を学習によって最適化させることにより、極めて高速な処理を実現したこと

この枠組みは、①顔の分析に有効な新しい画像特徴の集合、②AdaBoostに基づく新しい特徴選択アルゴリズム、及び③著しく処理を加速する学習と検出の直列構造を特徴とし、画像特徴とAdaBoostプロセスの両方を学習と特徴選択のために採用している。

3.2 フィルタと特徴

画像特徴は矩形(くけい)特徴と呼ぶことにするが、これは、多重解像度分析による特徴抽出を行う基底として知られるハール基底(Haar basis)に類似している。それぞれの矩形特徴は、しきい値 θ_i と矩形フィルタ $f_i(x)$ からなる二値しきい値関数であり、矩形フィルタ $f_i(x)$ は画像の線形関数である。

$$h_i(x) = \begin{cases} 1 & f_i(x) > \theta_i \\ 0 & \text{それ以外} \end{cases} \dots\dots\dots(1)$$

矩形フィルタには3種類ある。二矩形フィルタの値は、同じ大きさと形を持ち、水平方向又は垂直方向に隣接している二つの矩形領域内のピクセル値の総和の差である(図8の(A), (B))。三矩形フィルタの値は、中央の矩形内のピクセル値の総和から両側の矩形内のピクセル値の総和を差し引くことによって得られる(図の(C))。さらに、四矩形フ

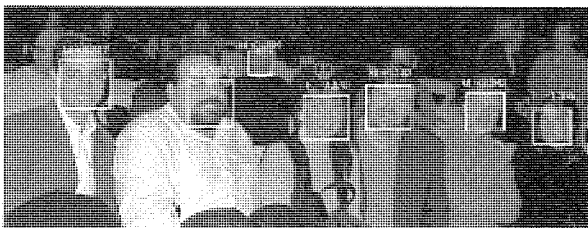


図7. 顔の検出と性別分類

ィルタは、対角位置にある2組の矩形内ピクセル値の差である(図の(D))。

統計的パターン識別器は、二値矩形特徴からなる。異なる矩形特徴の数は極めて大きく、それぞれのフィルタに対して多くの異なるしきい値が存在可能であり、結果としてそれぞれに対して違う特徴が出力される。純粋に機械学習に関する考察に基づけば、最終的に統計的パターン識別器がすべての異なる矩形特徴を含むことは不可能である。計算効率が大きな制約となるため、特徴の数を数百から数千に制限した。したがって、計算効率が高く、誤りの少ない識別器を作ることができる特徴の小さな集合をいかに選択するかが課題となる。

多くの特徴選択アルゴリズムが発表されているが、AdaBoostは、整合性のある有効な特徴選択のメカニズムを提供する。顔検出の識別器が単純な矩形特徴の識別器の組合せで実現されているとすると、AdaBoostの学習サイクル(ブースティング)1回ごとに最も有効な新しい矩形特徴が加えられていく。

実際にAdaBoostを数百回実行した結果、次式で与えられる最終分類子を得た。

$$C(x) = \sum_j a_j h_j(x) \dots\dots\dots(2)$$

ここで、 h_j は矩形特徴(式(1))、 a_j は各特徴に対する重み係数である。

男性1,600人と女性1,600人からなるデータ集合を用いてテストした結果、特徴学習では表1に示す誤差率を得た。比較のため、既存の方式SVMによる誤差率も表に含めた。このテストでSVMは非常に良い結果を示しているが、計算時間に顕著な差が生じる。特徴に基づく方式ではSVMより1,000倍高速に処理できる。

次に、様々な背景と照明下で撮影されたビデオ映像(ビデオシーケンス)によるテストを行った。一人の人間が9

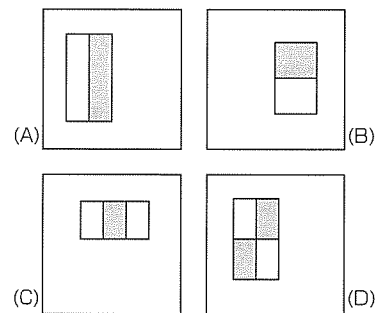


図8. 矩形フィルタ

表1. 性別判定結果(5-fold CV)

分類子	女性(%)	男性(%)	合計誤り(%)
50 features	24.7	24.3	24.5
100 features	23.7	23.1	23.4
300 features	23.7	21.2	22.4
500 features	21.9	20.1	21.0
SVM, rbf kernel	28.1	20.9	24.5

秒間登場する30編のビデオシーケンスを三つのサブシーケンスに分け、合計90編を用いた。ビデオの中に現れる顔の個々の性別についてサブシーケンス全体にわたって時間平均した結果、性別に関する最終誤り率は9%に減少した。

4. 目と視線の検出

この章では、照明変化の多い半屋外環境下で、着座した人物の顔方向を単眼視によって推定するために、顔面上の特徴点を追跡する手法を紹介する。この手法は、分離度フィルタと更新型テンプレートマッチングを用いて、それらの結果を重み付け統合する。この重み係数を適切に設定すると、高精度で見逃しの少ない追跡を実現できる。この追跡手法を用いて顔上の目と鼻の位置を追跡し、顔及び大まかな視線方向を推定するリアルタイムシステムを試作した。

顔や視線の方向を推定する手法は、大きく二つに分けられる。第一は目などの局所的な特徴を追跡する手法で、第二は顔全体の見え方から直接顔や視線方向を推定する手法である。前者は不特定の人物に適用しやすいが、顔特徴点を追跡中に見失わないことが課題であった。

当社で開発した手法は、分離度フィルタと更新型テンプレートという性質の異なるマッチングを並行的に行い、両者の位置を重み付け統合することでこの課題を解決した。この手法によれば、①探索が近傍領域で済むので計算が高速で、②両者のうちどちらか一方が追跡に失敗しても追跡が安定で、③高精度な追跡位置が得られる。

4.1 顔の特徴点(目と鼻孔)の初期検出

顔の特徴点を追跡するためには、まず追跡開始位置となる特徴点を発見する必要がある。この事例では、カメラは顔の正面下方から撮像していると仮定し、特徴として見付けやすい鼻孔と目に着目する。

まず、三つの灰色の短冊領域に注目して左側の鼻孔を検出する(図9)。画像上で、鼻孔より左の矩形領域に黒領域が存在し、かつ残りの二つの領域には黒領域が存在しない場合を左側鼻孔候補とする。右側の鼻孔検出も左右対称にして同様である。2値化で得られた鼻孔候補から選んだ任意の組合せのうち、領域間の距離がモデルに適合するものを左右鼻孔とする。さらに、鼻孔の位置から目が存在する

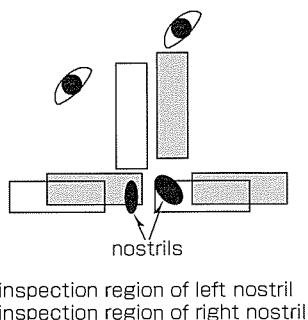


図9. 鼻孔の初期検出

領域を推定し、次節に示す分離度フィルタを適用して目を検出する。

4.2 2種類のテンプレートマッチングの統合による追跡

目と鼻孔の特徴点の追跡には、2通りのテンプレートマッチングを用いる。第一は、更新型テンプレートマッチングで、前回追跡に成功した特徴点を中心とする局所領域の画像をテンプレートとして、入力画像から最も相関の高い位置を探索する。この方法では、顔の個人差や環境の変化に影響されず、広い探索領域から検出が可能であるが、図10の(c)~(e)のように見え方が変化すると、徐々に追跡の中心がずれて最後には見失ってしまうという失敗に陥りやすい。

第二は、分離度フィルタと呼ばれる、顔の共通特徴をモデル化したテンプレートを用いる探索である。分離度フィルタは、図の(a)(b)に示すように、黒及び灰色領域で定義された2領域の平均輝度の分散を領域全体の輝度分散で正規化して出力するフィルタである。

図において、黒色の領域は周囲に比べて明るい画素を表しており、灰色の領域は周囲に比べて暗い画素を表している。また、背景色(白)は無視されることを示している。図の(a)(b)はそれぞれ目の中心(瞳(ひとみ))及び二つの鼻孔と、それをつなぐ領域をモデル化しており、個人や照明条件に依存しない共通の特徴である。ところが、顔の各部分には類似のパターンが数多く存在するため、探索領域を広くとると、目や鼻孔以外の特徴点を誤って検出する可能性がある。

そこで、これら二つの追跡手法の長所を生かしながら、以下のように両者の追跡結果を線形結合によって統合する。ここで、 z は統合された追跡点の位置、 x は更新型テンプレートマッチングによる追跡点の位置、 y は分離度フィルタによる追跡点の位置を示し、添字 i は第 i 番目の画像を示す。 k は1以下の正の係数である。

$$Z_{i+1} = Z_i + kx_{i+1} + (k-1)y_{i+1} \dots\dots\dots(3)$$

追跡一回当たりの更新型テンプレートマッチングの誤差は小さく、探索中心近傍では分離度フィルタの誤検出の可能性が低いため、一方が誤ったとしても他方が真の特徴点位置へ引き戻す働きをするため、長時間の追跡においても大きくずれたり見失ったりすることがない。

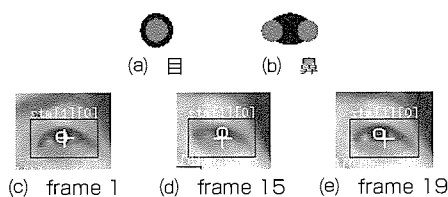


図10. 分離度フィルタの領域定義(a)(b)及び分離度フィルタと更新型テンプレートによる検出例(c)~(e)

4.3 視線方向の計測

上記二種テンプレートマッチング統合型の顔特徴追跡手法を用いて、運転中のドライバーの視線方向を推定するシステムを試作した。カメラはドライバーを斜め下60cmの距離から撮影しており、画像の解像度は640×480画素、画面上の顔の横幅は約250画素である。まず、三つの追跡手法それぞれについて目、鼻の追跡性能を、右目、左目、鼻の誤追跡及び追跡中断(システムが特徴点の追跡に失敗したと判断して特徴点の位置を検出し、追跡をやり直す)回数(カウント)を調べた。表2のように、二種統合型のこの手法は、誤検出率・中断回数ともに優れていることが実証された。

複数の人物に対して合計30分の画像シーケンスを用意し、正面から横向きの角度が50°を超えたかどうかによって脇見判定する実験を行った(追跡中の画面の例は要旨のページ参照)。追跡中断によるか検出が課題であるが、従来の脇見検出の性能を上回る87%という高い検出率を達成した(表3)。

5. む す び

監視システムへのステレオビジョンの応用、MPEG-7技術の応用、コンピュータビジョンによる人間の顔の検出、目と視線の検出について技術を紹介した。今後は更に監視システムの高度化が進むと考えられ、当社もこれまで開発した技術に加えて、様々な市場要求に適用し得る技術の開発を行っていく所存である。

参 考 文 献

(1) Beardsley, P., et al.: Classifying Pedestrians and Wheelchair Users using Stereo Vision (2002)
 (2) Viola, P., et al.: Rapid Object Detection using A Boosted Cascade of Simple Features, Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (2001)

表2. シーケンス(300秒)に対する誤追跡と追跡中断の割合

手 法		誤追跡			追跡 中断
		右目	左目	鼻	
更新型テンプレート マッチング	カウント	45.0	44.0	38.0	40.0
	時間(秒)	165.1	161.6	146.5	52.0
	誤認率(%)	55.0	53.9	48.8	17.3
分離度フィルタ	カウント	31.0	31.0	34.0	69.0
	時間(秒)	21.1	24.1	26.4	47.8
	誤認率(%)	7.0	8.0	8.8	16.0
統合型	カウント	4.0	4.0	6.0	30.0
	時間(秒)	2.2	2.0	8.0	20.0
	誤認率(%)	0.7	0.7	2.7	6.6

表3. 脇見判定精度

	真の脇見	正常検出	検出失敗		合 計
			見逃し	過検出	
数	15	13	2	11	28
率(%)		87.0	7.6	46.0	

(3) Viola, P., et al.: Fast and Robust Classification using Asymmetric Adaboost and A Detector Cascade, Advances in Neural Information Processing (2001)
 (4) Shakhnarovich, G., et al.: A Unified Learning Framework for Real Time Face Detection and Classification: 5th International Conference on Automatic Face and Gesture Recognition (2002)
 (5) Manjunath, S., et al.: Introduction to MPEG-7: Multimedia Content Description Interface. John Wiley, Chichester (2002)
 (6) Cieplinski, L.: MPEG-7 Color Descriptors and Their Applications, 9th International Conference on Computer Analysis of Images and Patterns, 11~20 (2001)
 (7) Bober, M., et al.: An MPEG-4/7 Based Internet Video and Still Image Browsing System, SPIE 4209 (Multimedia Systems and Applications III), 33~38 (2001)

接触認識技術 “ダイヤモンドタッチ”

ポール・ディーツ*
ダレン・リー*

要 旨

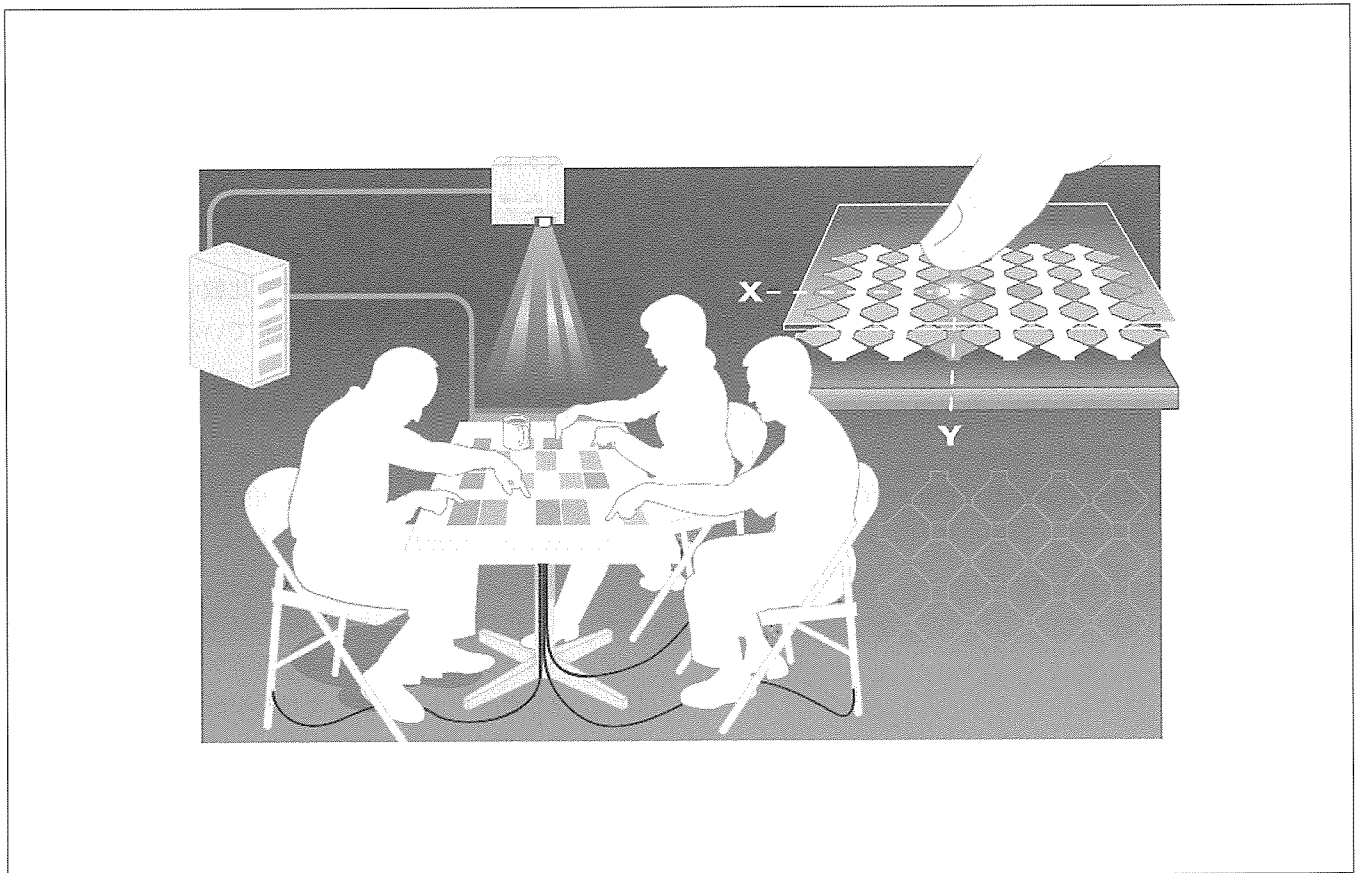
人間がコンピュータと対話する直感的な手段として、タッチスクリーンが様々な分野で採用されている。今回、複数のユーザーによる協調作業を可能とする桌上ディスプレイ付きマルチユーザータッチスクリーンシステム“ダイヤモンドタッチ(DiamondTouch)”を開発した。

- (1) 接触感応型の入力デバイスにより、ユーザーごとに接触情報を独立に検知し、分類・優先順位付けなどを行うことができる。
- (2) 複数のユーザーが同時にコンピュータと対話することが可能である。
- (3) タッチスクリーン表面に置かれた異物は検知されない。

- (4) 特殊ペンなど専用のデバイスが不要である。

このシステムにおいて用いる入力デバイスには、接触表面上に多数の微小なアンテナが配列され埋め込まれている。ユーザーが接触表面に接触すると、そこに位置する微小アンテナとユーザーの身体との間に微小な電流が流れるので、椅子(いす)を介して受信すれば、ユーザーごとの接触情報を独立かつ同時に検出することができる。

ダイヤモンドタッチは、人間とコンピュータの対話や、複数ユーザーによる協調作業の新たな形を提案するキーデバイスとして発展することが期待される。



ダイヤモンドタッチシステム

ダイヤモンドタッチシステムは、テーブル型の前面投射型ディスプレイを使用している。ディスプレイには多数のアンテナが配列され埋め込まれており、ユーザーと椅子を介して微小な電流が流れる。これにより、複数ユーザーが同時にコンピュータと対話し、どの接触がどのユーザーによるものか特定することが可能となる。

1. ま え が き

タッチスクリーンは、人間が機械やコンピュータと対話する直感的な手段として、様々な分野で採用されている。単に“触れる”という動作は、キーボードやマウス操作とは異なり、ユーザーに対して習熟の面倒を感じさせることがなく、ごく自然に受け入れられてきたのである。

今回、複数のユーザーによる協調作業を可能とするディスプレイ付き卓上マルチユーザータッチスクリーンシステム“ダイヤモンドタッチ”を開発したので、技術の概要とプロトタイプシステムについて紹介する。

一般に、複数のユーザーが同時に操作するマルチユーザータッチスクリーンは、次のような条件を満たす必要がある。

- (1) マルチポイント：複数の接触を同時・独立に検出可能であること
- (2) 同定：ユーザーごとにどの場所に接触しているかを検出し、分類や優先順位付けが可能であること
- (3) 異物の無視：表面に置かれた物体は検知しないこと
- (4) 耐久性：壊れにくく、頻繁な校正が不要であること
- (5) 簡便性：特殊ペンなど専用のデバイスが不要であること
- (6) 低コスト：低価格でランニングコストが低いこと

ダイヤモンドタッチは、テーブル上に置かれたタッチスクリーンと、上方からこのタッチスクリーンに向かって投影する前面投射型プロジェクタで構成される。タッチスクリーンの周囲に座っている個々のユーザーは、タッチスクリーン表面に投射される映像を見ながら、それに接触することを通じてコンピュータと対話する。

タッチスクリーン表面には多数の小さなアンテナが配置されており、さらに、薄い絶縁層に覆われて接触面を形成している(図1)。ユーザーが座る椅子は受信機に接続されており、ユーザーが接触面に触れると、その場所に配置されたアンテナ、ユーザーの身体及び椅子の容量結合によって微小な電流が流れるので、接触を検知することができる。また、接触位置は、微小電流が流れるアンテナの位置によって特定される。

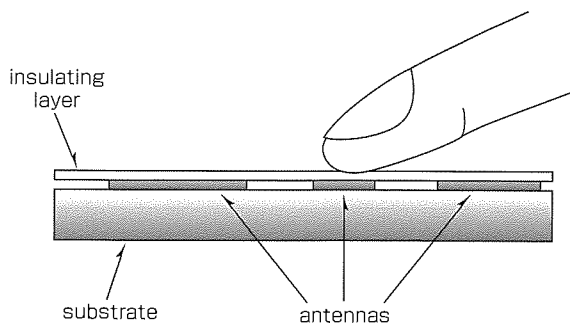


図1. タッチスクリーン断面図

2. ダイヤモンドタッチシステム

2.1 容量結合による接触検出

ダイヤモンドタッチは、タッチスクリーン、ユーザーの身体及び椅子が容量結合することによって作動する。タッチスクリーン表面に配置されている個々のアンテナの面積は、薄い絶縁層を介してユーザーの指との間に良好な容量結合を得る形と大きさでなければならない。

ユーザーの身体と椅子との間の容量結合は、金属製の椅子や導電マットを用いれば、ユーザーの指とタッチスクリーンとの間の容量結合よりもはるかに大きくなるので、ユーザーの指とテーブルとの間の接触検出がこのタッチスクリーンを動作させる上で重要である。

2.2 接触位置の特定

ユーザーがタッチスクリーンに接触する場合、近接した複数のアンテナと同時に容量結合すると考えられるので、複数のアンテナを介して得られる信号が交じり合った形で検出される。そのため、検出された信号がどのアンテナを介して得られたのか特定する必要がある。これを解決する一つの方法は、以下に示すような多重直交信号を用いることである。

- (1) 周波数分割多重(Frequency-Division Multiplexing : FDM)
- (2) 時間分割多重(Time-Division Multiplexing : TDM)
- (3) スペクトル拡散(Spread Spectrum : SS)

開発したプロトタイプシステムでは、回路設計が比較的容易なTDM方式を採用した。個々のアンテナを時分割に駆動して受信信号とのタイミングを比較することにより、アンテナ位置(=接触位置)を特定している。

2.3 ユーザーの識別

複数のユーザーが同時に操作するので、ユーザーごとに接触状況や接触位置情報を独立に検知する必要がある。このシステムでは、ユーザーごとに検出回路(タッチスクリーン~ユーザーの身体~椅子~受信機)を設けて複数ユーザーの操作に対応している。

2.4 アンテナ素子の形状と配置

タッチスクリーン表面に配置されるアンテナ素子の形状は、アプリケーションによって様々な形状とすることが可能である。例えばON・OFFのような単純なアプリケーションであれば、大きなボタン型のものであっても構わない。

まず、多数の小さなアンテナ素子を配置する代表的な例として、フルマトリックスと呼ばれる長方形の格子状パターンが考えられる。このパターンでは、各々独立に位置付けられたアンテナ素子により、ユーザーが接触した位置を厳密に特定することができる。しかし、これは接続線数と

周辺回路の規模が膨大になり、コストの増大を招きやすいという欠点がある。

開発したプロトタイプシステムでは、アンテナ素子パターンを重層構造の行と列のパターンとして、ユーザーが接触した場所を行と列に関する情報から特定できるようにした。このパターンを用いれば、フルマトリックスに比べて接続線数を大幅に削減することができ、周辺回路規模の増大も問題とならない。しかしながら、単純な長方形の行列パターンでは、パターンの重なり部分において、上層パターン(列パターン)は下層パターン(行パターン)に対する遮へい物として作用し、行に関する感度が低下するので、例えば図2に示すような連結菱形(ひしがた)パターンを採用して、ユーザーとアンテナ間の容量結合を最大化しながら、行アンテナと列アンテナ間の不要な遮へい効果や容量結合を抑える必要がある。

実際の操作においては、通常1回の接触で複数の行と列に関する情報が検知される。そのため、複数の行と列にまたがって情報が検知された場合は、それらが小さな領域に含まれるときには一点の接触とみなし、検知された信号から中心位置を算出して接触位置を特定する。一方、検知された領域が互いに離れて複数存在する場合には、複数の接触とみなし、それぞれの中心位置を算出して接触位置を特定する。

この処理を応用すれば、あるユーザーがタッチスクリーン上の2点を指定した場合、その2点間を結ぶ線を対角線とする長方形を描画することが可能となる(バウンディングボックス機能)。バウンディングボックス機能は、タッチスクリーン上である範囲を指定する場合など、極めて効果的に使える機能である。

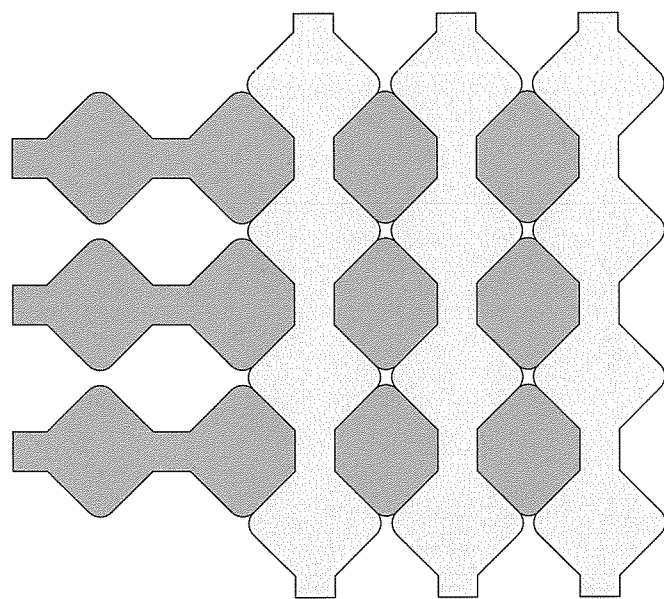


図2. 連結菱形アンテナパターン

3. プロトタイプシステム

ダイヤモンドタッチの機能検証のために、20cm×20cm(80アンテナ)から48cm×80cm(256アンテナ)まで数種のサイズのプロトタイプシステムを開発した。タッチスクリーンは、プラスチックシートと誘電体からなり、導電性のインクでアンテナパターンが印刷されている。個々のアンテナは、0.5mm間隔で配置されており、ユーザーの指の接触は少なくとも2行2列のアンテナにまたがるようになっている。

タッチスクリーンからコンピュータまで、振幅15V、100kHzの信号が10サイクルごとに時間分割多重(TDM)で送信されている。ユーザーがタッチスクリーンに接触すると、ユーザーの身体及び椅子を通じて信号が受信され、同期復調及び8ビットのアナログ-デジタル変換の後、コンピュータに送られる。そして、前面投射型プロジェクタを用いて、タッチスクリーン表面に対話すべき映像を表示する。

ダイヤモンドタッチのシステムを動作させるソフトウェアは、WindowsとLinux両方のOSに対応するよう開発した。これにはデータ獲得、フィルタリング、ユーザー動作解釈などの基本的ルーチンが含まれており、接触点特定については、補間法を用いて識別分解能0.3mmを実現している。

図3は、実際のダイヤモンドタッチシステムを使用する様子を示している。実験では、映像の投射されたタッチスクリーン上で複数のユーザーがそれぞれ複数の箇所と同時に接触しても、互いに干渉することなくユーザーごとに接触状況を検出できることを確認した。

また、タッチスクリーン上に放置されたコーヒーカップ又はこぼれたコーヒーなどは、このシステムの動作に全く影響しないことを確認することができた。

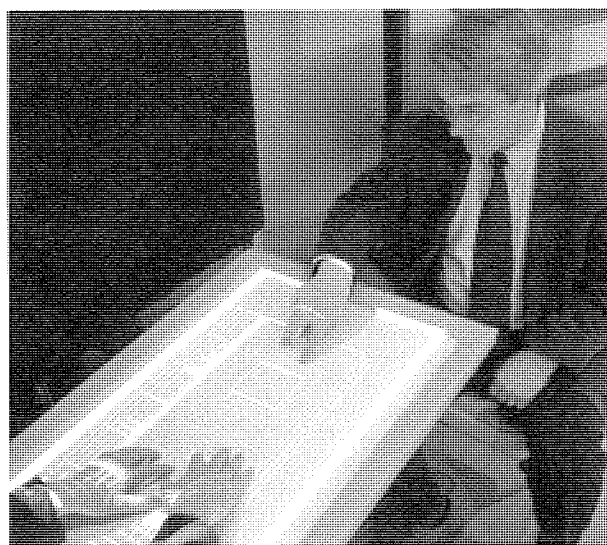


図3. ダイヤモンドタッチシステム使用状況

4. む す び

複数のユーザーによる協調作業を可能とするディスプレイ付き卓上タッチスクリーンシステム“ダイヤモンドタッチ”を開発した。接触感应型の入力デバイスにより、ユーザーごとに接触情報を独立に検知し分類・優先順位付けを行うことができ、複数のユーザーが同時にコンピュータと対話することが可能である。

今後は、インターネットを介した協調作業、店舗における注文システム、マルチプレーヤー用ゲームなど、多くの応用分野が考えられ、実用化に向けて更なる技術開発が必要である。

- (1) スクリーンの大型化
- (2) 後方投射型プロジェクタの採用
- (3) 用途開発

ダイヤモンドタッチが、人間とコンピュータの対話や、複数ユーザーによる協調作業の新たな形態を提案するキーデバイスとして発展していくことを期待する。

参 考 文 献

- (1) Dietz, P., et al. : DiamondTouch : A Multi-User Touch Technology in Proceedings of the ACM UIST 2001 Symposium on User Interface and Software Technology, 219~226 (2001)



携帯端末用“Z3D” グラフィックスエンジン

亀山正俊* 河合浩行***
藤本仁志* 児玉幸夫†
岩崎建樹**

要旨

近年、携帯電話を始めとする携帯端末は、表示画面のカラー化や大型化などにより、単なる電話機からマルチメディア情報端末へと進化してきている。このような携帯端末では使いやすく高度なマンマシンインタフェースや楽しさなどが求められ、3Dグラフィックス表示も必要になってくる。

三菱電機は、携帯端末で迫力のある3Dイメージを生成することのできる3Dグラフィックスエンジン(“Z3D”グラフィックスエンジン：以下“Z3D”という。)を開発した。

Z3DのLSIコアに内蔵される3Dレンダリングパイプラインは、ジオメトリエンジン、レンダリングエンジン、ピクセルエンジンで構成され、シェーディング、テクスチャマッピング、光の計算などを高速に実行することができる。また、2Dエンジンも内蔵し、文字描画やスクロールも高速に実行される。30万ゲートのロジックと2.3Mビットの

SRAMで構成され、約30mm²に集積することができた。3D表示のピーク性能は23万頂点/秒、5.2Mピクセル/秒である。また、クロックの制御によって3D動作時の消費電力を54mWと低くすることができた。

組み込みソフトウェア及びJavaアプリケーションから3Dグラフィックス描画機能を利用できるように、API(Application Program Interface)を規定した。このAPIは、PC(Personal Computer)における3DグラフィックスAPIと親和性が高い点、市販のモデラで作成した3Dオブジェクト(1体のキャラクタなど)を一つのディスプレイリストとして扱える点、アニメーション動作を簡単に制御できる手段を提供した点など、コンテンツ作成が容易になるよう配慮されている。これにより、3Dキャラクタのアニメーションや3Dゲームなどのアプリケーションプログラムを容易に作る事ができる。



“Z3D”グラフィックスエンジンを使用した携帯電話D504i

本格的な3Dグラフィックスエンジンである“Z3D”グラフィックスエンジンを搭載した当社の携帯電話D504iを示している。リアルな3Dキャラクタアニメーションや3Dゲームを実行することができる。

1. ま え が き

近年、携帯電話は、画面のカラー化・大型化・高画質化によって高精細なカラーの静止画やビデオを表示できるようになり、電話機からマルチメディア化された情報端末へと進化してきている。また、携帯情報端末(PDA)においても同様にカラー化と高画質化が進展している。このような携帯端末では、コンテンツの通信、蓄積、検索ができるだけでなく、使いやすく高度なマンマシンインタフェースや楽しさなどが求められ、PCがたどってきた進化と同様に、リアルタイムの3Dグラフィックス表示も求められてきている。

本稿では、このような高度化された表示画面を持った携帯端末で迫力のある3Dイメージを生成することのできる“Z3D”とそのソフトウェアについて述べる。

2. 開発コンセプトと機能概要

携帯端末の表示画面はPCに比べれば小さく画素数も少ないので、優れたマンマシンインタフェースを実現することが重要である。そのために、高速の3D表示を実現し、“見やすく”“分かりやすく”“楽しい”画面を提供することを基本的な目標としている。また、3D表示であっても、携帯であるために、低消費電力、IPコアの小型化、CPUに負荷をかけないことが求められる。

携帯端末で3D表示され得るコンテンツとして、3Dのキャラクターアニメーションや3Dゲーム、3D地図や建物のウォークスルーなどが挙げられる。これらコンテンツのリアルで迫力ある表示を実現するためには、滑らかな陰影付け、バイリニア補間によるテクスチャマッピング、複数光源による光の計算がなされた3Dイメージを7フレーム/秒以上の性能で表示することが必要となる。

また、3D表示機能にとどまらずグラフィックス全般の高速化・高機能化が求められており、文字描画や画面スクロール、矩形(くけい)塗りつぶしなどの2D機能の高速化も不可欠である。

図1にZ3Dと制御ソフトウェア、コンテンツ開発環境の構成を示す。

3. 3Dレンダリングパイプライン

一般に3D表示では、3D物体の表面を連続したポリゴン(三角形)で表現している。3D物体を二次元の画面に表示するときには、3D物体を2Dの画面に投影して2Dの画面でそれぞれの三角形を塗りつぶすことにより、3Dイメージを生成する。

3Dイメージを生成するための一連の処理をハードウェアで高速に実行する3Dレンダリングパイプライン⁽¹⁾は、パソコンやゲームマシンなどで使われているものと基本的

には同様である。3Dレンダリングパイプラインは、ジオメトリエンジン、レンダリングエンジン、ピクセルエンジンで構成される。

3.1 ジオメトリエンジン

ジオメトリエンジン⁽²⁾⁽³⁾は、固有の座標系にある3D物体を2Dの表示画面の座標、色情報に変換する処理を行う。3D物体を構成する三角形の頂点ごとに処理が行われる。具体的な処理は次のとおりである。

(1) 座標変換

3D物体固有の座標系から拡大・縮小、移動、回転をし、2D平面への透視投影を実行

(2) 光の計算

光源データや視点データ、3D物体を構成する各頂点の座標データ、法線ベクタ、色データからその頂点の光(色)の強さを計算する。光源の種類には平行光線、点光源、スポットライトがあり、複数の光源を定義することが可能

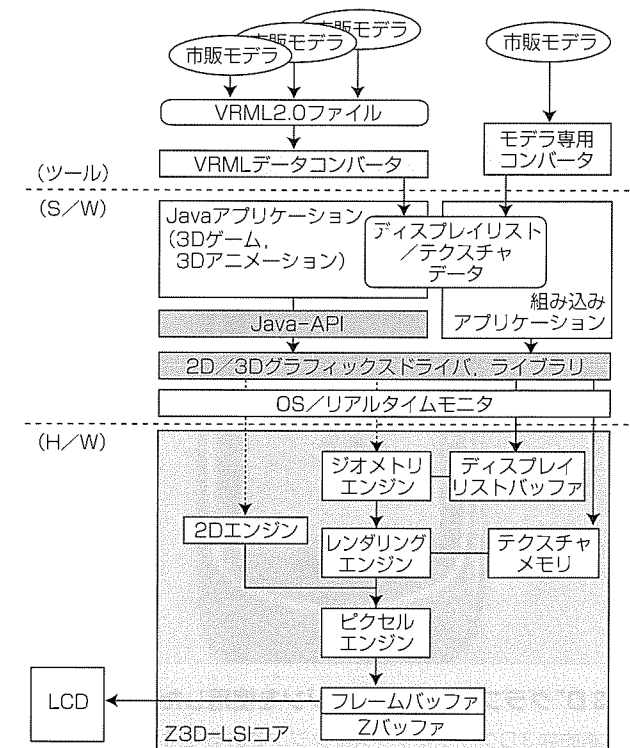
(3) クリップ処理

三角形の表示される部分だけを切り取る処理

(4) 頂点補間

3Dキャラクタのアニメーションを高速に実行するために頂点補間の機能も実現

このようにジオメトリエンジンでは数値計算を高速に実行する必要があり、図2に示すように浮動小数点演算器2個と整数演算器を1個搭載した専用プロセッサで構成される。二つの浮動小数点演算器は、並列に動作するSIMD



VRML : Virtual Reality Modeling Language(3Dイメージの記述言語)

図1. Z3Dの構成

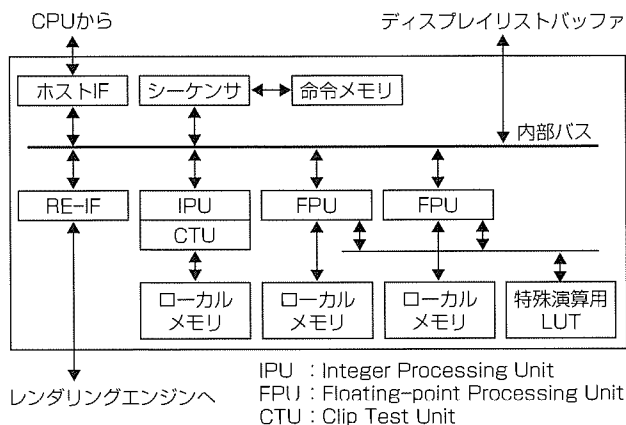


図2. ジオメトリエンジンの構成

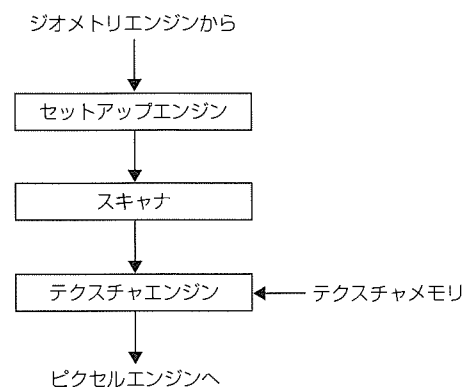


図3. レンダリングエンジンの構成

(Single Instruction stream Multiple Data stream)型プロセッサである。座標変換のための積和演算以外に除算や開平、累乗、クリップテストのための専用ハードウェアを内蔵しており、書換え可能なマイクロコードに従って幾何学計算が高速に実行される。

3.2 レンダリングエンジン

レンダリングエンジンは、ジオメトリエンジンで計算された画面座標系での頂点座標や色データから、三角形の塗りつぶしや、直線や点の描画処理を行う。主な特徴的機能を次に述べる。

(1) シェーディング

三角形を塗りつぶす際にカラー値を補間することにより、滑らかな濃淡を付け、三次元的な陰影を付ける。

(2) テクスチャマッピング

あらかじめ決められた模様を三角形の面に張り付ける処理で、より三次元的で現実感のあるイメージを作ることができる。テクスチャマッピングでは、三角形の傾きや遠近の差による模様の変形をバイリニア補間(4点補間)し、より滑らかな動きが実現できる。フルカラーだけでなく、少ないデータで効率的な3Dイメージが生成できるインデックステクスチャも提供する。

(3) アンチエイリアシング

画面上で斜めの線を引いたときに生じるギザギザを取り去る直線アンチエイリアシングの機能を搭載している。

レンダリングエンジンは図3に示されるようなパイプライン構造をハードワイヤードロジックによって実現しており、高速に演算処理することができる。セットアップエンジンは三角形の塗りつぶしやテクスチャマッピングのためのパラメータを計算する部分、スキャナは陰影付けされたピクセルデータを生成する部分、テクスチャエンジンはテクスチャメモリからデータを読み出し、テクスチャマッピングされたピクセルデータを生成する部分である。

3.3 ピクセルエンジン

ピクセルエンジンは、レンダリングエンジンが生成する

ピクセル(画素)データと、フレームバッファ/Zバッファにあるピクセルデータとの間で演算を行う。具体的には次のような機能を実現する。

- (1) Zバッファ法による隠面消去(遠い物体が近い物体によって隠れる)
- (2) ブレンディング機能により、半透明表示、直線アンチエイリアシング、マルチテクスチャなどを実現
- (3) ラスターオペレーションによるピクセル間の論理演算

4. 2Dエンジン

Z3Dには、3Dパイプラインだけでなく、2Dエンジンも実装されている。2Dエンジンは表示画面上での処理を実行するもので、具体的には次のような機能を高速に実行することができる。

(1) 文字描画

あらかじめ特定の文字のフォントパターンを2Dエンジン内部のメモリに書き込み、起動をかけると、打ち抜きで文字パターンを高速に描画する。

(2) 矩形塗りつぶし

画面上の矩形領域を特定の単色で高速に塗りつぶす。

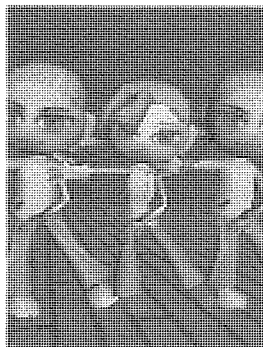
(3) 矩形転送(BIT Block Transfer)

画面のイメージデータを矩形単位で移動し、画面のスクロールやウィンドウの移動を高速に実行する。

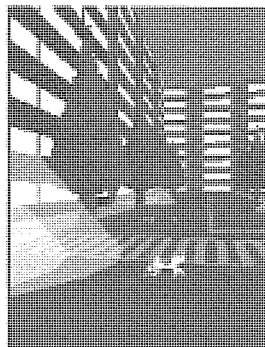
2Dエンジンは、ピクセルエンジンと協調して動作することができる。

5. LSI化と性能

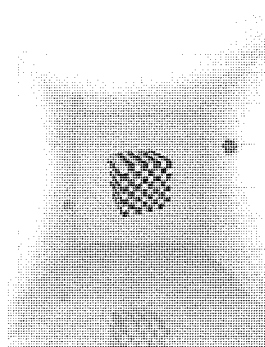
Z3DのLSIコアは30万ゲートのロジックと2.3MビットのSRAMで構成され、当社プロセスで約30mm²に集積することができた。また、3Dイメージ生成をこのようなパイプライン構造によって実行することにより、ピーク性能で23万頂点/秒、5.2Mピクセル/秒を実現することができた。なお、Z3Dの消費電力は3D表示時に約54mWで、低消費電力を実現している。



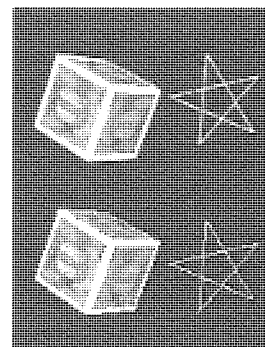
(1) ダンシングシーン
複数のキャラクターが躍動感あふれた動きをすることができる。



(2) 市街クルージング
市街地を歩いたりドライブができる。



(3) 複数の点光源と反射
赤、青、緑の点光源によって物体表面の色が異なる。床には物体の反射が映っている。



(4) アンチエイリアシング
上図の斜め線のギリギリが小さい。

図4. 3D表示画面例

6. 制御ソフトウェア

Z3D制御用ソフトウェアは、ハードウェアリソースを管理する2D/3Dグラフィックスドライバと、3Dグラフィックスライブラリ(Z3D-lib)で構成される。組み込み用のAPIとほぼ同等の機能を提供するJava-APIを規定し、広くJavaアプリケーションから3Dグラフィックス機能を利用可能とした。このAPIは、PCにおける3DグラフィックスAPIと親和性が高い点、汎用モデラで作成した3Dオブジェクト(1体のキャラクターなど)を一つのディスプレイリストとしてまとめて扱える点など、コンテンツ作成が容易になるよう配慮されている。

7. コンテンツ開発ツール

3Dキャラクターアニメーションや3Dゲームソフトウェアを開発する上で、市販のモデラを使用して作成した3DオブジェクトをZ3Dで実行可能なディスプレイリストの形式に変換するコンバータや、データ確認用のビューなどのツール類の開発も行った。より多様なモデラに対応するため、最も一般的な3DデータのフォーマットであるVRML2.0からのコンバータを準備し、さらにモデラ独自の機能を取り込むために、専用のコンバータを順次開発していく方法をとった。

図4は携帯端末上のコンテンツによる3D画面表示例を示している。図の(1)(2)はアニメーションの1シーンを、

(3)(4)は特殊効果の例を示している。

8. む す び

本稿では携帯端末用の3Dグラフィックスエンジン“Z3D”の概要を述べた。携帯端末とそのマンマシンインタフェースの発展に伴い、今後とも優れた3Dグラフィックス表示技術が求められており、より小型で高性能、低消費電力なLSIコアが必要となる。機能面においても、単にPCの技術に追従したリアルな3Dイメージを生成するだけでなく、“楽しさ”や“分かりやすさ”を持ったグラフィックス表示技術を追求していくことやカラーイメージ関連機器との融合を検討していく必要がある。

参 考 文 献

- (1) 亀山正俊, ほか: 三次元グラフィックス技術と3Dチップセットへの応用, 三菱電機技報, 72, No.9, 724~727 (1998)
- (2) 井上善嗣, ほか: 三次元グラフィックス向浮動小数点累乗演算器の開発, 信学技報, ED99-57, SMD99-31, ICD99-39, 17~22 (1996-6)
- (3) Kawai, H., et al.: A Programmable Geometry Processor with Enhanced Four Parallel SIMD Type Processing Core for PC-Based 3D Graphics, IEICE Trans. Electron, E85-C, No.5, 1200~1210 (2002-5)

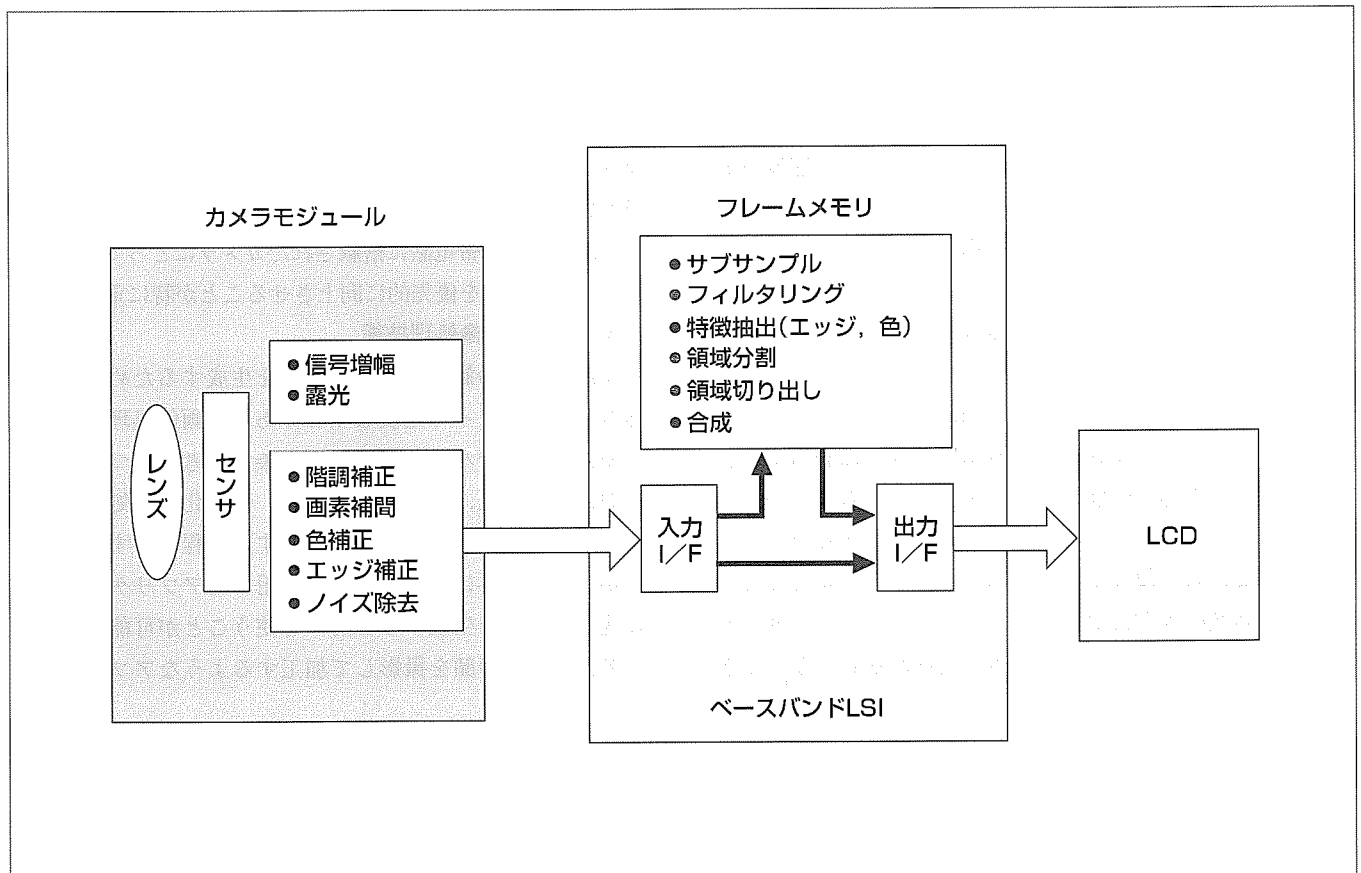
携帯電話向け画像処理技術とその応用

的場成浩*
 田中正二**
 ミロスロー・ポーバ***

要旨

携帯電話機は、ここ数年で表示デバイスがモノクロからカラー化へと一気に進み、表現能力が大きく向上してきた。これにより、従来の文字系中心のマンマシンインタフェースから画像系中心のマンマシンインタフェースへとウェイトがシフトするとともに、画像関係のアプリケーションや周辺機能が次々と搭載されつつある。その一つがカメラ機能である。一足先に登場したデジタルスチルカメラが、画像をデジタル化して容易に扱えることから、パソコンを核として、急速に市場を広げてきた。このデジタルスチルカ

メラとしての機能が、携帯電話機にも取り込まれつつある。しかし、携帯電話機は、普及層の広さ、携帯性の高さなどデジタルスチルカメラとは少し異なった面を持っており、アミューズメント性や操作性面への考慮が重要となる。つまり、電話機でありながらカメラ機能を備え、さらにはその応用機能まで必要となる。そこで、これらのニーズに対応するため、今後携帯電話機に要求されるカメラ機能やカメラを用いたアプリケーションの実現に向け有効と考える幾つかの画像処理技術を紹介する。



携帯電話機におけるカメラ画像のデータフローと処理

携帯電話の使用シーンの多様性に適応したカメラ制御と画像処理の実現と、携帯電話の中の限られた資源の中でカメラ画像を用いたアプリケーションを実現することが重要となる。

1. ま え が き

携帯電話における画像処理の機能や性能はここ数年で急速に向上し、カメラを搭載した携帯電話が徐々に普及しつつある。これら携帯電話に搭載されるカメラは、携帯電話に付加されていることから、使用シーンや用途の面で一般のデジタルスチルカメラとは少し異なった想定が必要となる。

例えば、携帯電話は多くの人が常に持っており、どこでもすぐに使えることから、撮影される場所や撮影条件の多様化が挙げられる。また、デジタルスチルカメラとは違い、撮影した画像をすぐにその場で相手の携帯電話やパソコン等に送信したり、相手の携帯電話で撮った画像を受信して見るといった音声以外のコミュニケーション手段としての位置付けで使われることも多くなる。

したがって、単なるカメラとして使うのではなく、撮り方や撮った画像の取り扱い方などをも考慮したアプリケーション面からの仕掛けや工夫も必要となってくる。

本稿では、これら携帯電話に搭載されるカメラに必要な技術や、カメラ画像を用いたアプリケーションを実現する際に有効になるとと思われる画像処理技術について述べる。

2. 携帯電話用カメラ

2.1 携帯電話用カメラの現状

携帯電話が使われる多様な環境は、カメラにとって厳しいものとなる。電子的な撮影を行うデジタルスチルカメラでは、銀塩カメラで用いられるフィルムに比べて感度やダイナミックレンジが狭いCMOSセンサやCCDセンサ等の撮像素子を用いるため、適切な撮影条件となるようにきめ細かい制御を行うことが要求される。

しかし、携帯電話に搭載されるカメラの場合、実装スペースや消費電力といった面でデジタルスチルカメラに比べ実装条件が厳しく、また、カメラ制御に制約がかかることが多い。また、第一世代の携帯電話用カメラは表示用液晶の色再現能力や応答性能が低く、画面サイズも小さかったため、画質の面ではそれほど厳しい要求はなかった。

しかしながら、現在では、液晶の性能が向上したことから、撮像画像の画質に対する要求がデジタルスチルカメラ並みになり、トータル性能面からも条件が厳しくなっている。

2.2 携帯電話用カメラの構成

携帯電話にカメラを搭載する場合の構成としては様々考えられるが、一般的には、図1に示すように、電話機能用のベースバンドLSIの前段に一つのまとまったカメラセット機能として配置し接続する構成をとることが多い。カメラセット側でカメラにかかわる制御や信号処理はすべて行い、処理後の画像をベースバンドLSIへ渡す構成である。

この構成は、カメラ機能の追加がベースバンドLSIへ与

える影響が最も少なく、携帯電話本体側では、基本的に、カメラ画像データの転送やフロー制御部分に特化することができ、カメラ画像をアプリケーション用データとして独立して扱える利点がある。

2.3 カメラ制御技術

携帯電話に搭載されるカメラは、コミュニケーション手段としての用途の多様さから、静止物体よりは動きのある人や動物が被写体として選ばれる場合が多い。また、撮影方法も、携帯電話の構成上、片手で持つという不安定な状態で撮影されることが多い。さらに、撮影される場所も飲食店やホールなど暗い照明や特殊な照明の下で使われる機会が多く、撮影条件が厳しい場所で使用される場合が多い。そこで、これらの条件に対応可能なカメラ制御技術が特に重要となってくる。

カメラの使いやすさを測る上で重要な要素として、液晶ビューファインダのフレームレート(表示速度)がある。動きのある被写体を撮る場合には、このフレームレートが高いほど撮影しやすい。また、撮影環境に応じて適正な露光制御や色バランス制御を行うAE(Automatic Exposure)や、AWB(Automatic White Balance)機能がある。

AEやAWBは、定常的に被写体の明るさや色等の物理量をモニタし、適正な画像信号レベルが得られるように制御を行う機能である。被写体の明るさや様々な照明条件に対応するには、これらの機能が適切に動作する必要がある。さらに、AEやAWBでは動作の安定性と追従性(速度)をバランス良く制御することが、使いやすさだけでなく画質面でも常に安定した画像を撮影可能とすることにつながる。したがって、携帯電話に搭載されたカメラにとっては、これらの性能や機能を優先的に向上させることが特に重要となる。

2.4 カメラ画像処理技術

撮影のための制御と並び、画像を生成するための画像処理もカメラにとって重要な項目である。画像処理は、センサや液晶等のデバイス特性や撮像条件を考慮しつつ、画素補間、階調補正、色補正、先鋭化、平滑化等の処理を行う。これらの画像処理は、画像を生成する過程で必ず必要となるが、それ以外に、カメラ画像を用いたアプリケーションに適した画像を生成する目的にも使うことが可能となる。

例えば、人の顔を撮影して加工するようなアプリケーション

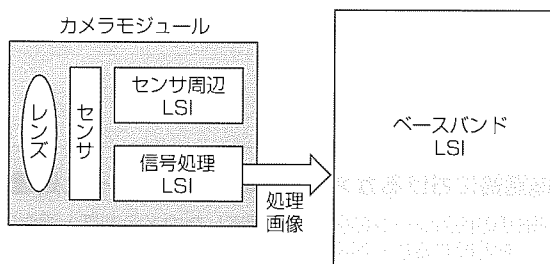


図1. 携帯電話用カメラの接続構成

ョンの場合、周辺の画像からターゲットとなる顔部分の画像に適した条件の画像処理を事前に施すことで、アプリケーション側での画像処理の負荷を低減させたり、認識処理能力等を向上させる効果も出すことが可能となる。

以上、携帯電話用カメラとして必要となる主要な技術に関して紹介したが、次章以降では、カメラ画像における顔画像を用いたアプリケーションの例を挙げ、その目的や技術的特長及び性能向上に有効と思われる手法に関して紹介する。

3. 小型携帯情報端末におけるリアルタイム 頭部領域抽出及び顔部位検出法

3.1 概要

PDA(Personal Digital Assistant)や携帯電話を始めとした小型携帯情報端末が急速に普及している。そのため、近い将来、個人情報保護のための様々な認証技術が必要になると考えられる。そこで、個人認証技術の中でも特に顔画像を用いた認証技術に着目し、小型携帯情報端末のハードウェアでも搭載可能な手法に関して紹介する。

前述のとおり、小型携帯情報端末をターゲットとしているため、開発するアルゴリズムは下記の条件を満たす必要がある。

- (1) 限られた計算資源でのリアルタイム処理
- (2) 任意の使用環境におけるロバスト処理

この章では、顔画像による個人認証の前処理に相当する技術として開発した、画像からの頭部領域抽出法及び顔部位検出法に関して述べる。ここで提案する手法は、上記(1)及び(2)の条件を満たすものであり、任意背景においても精度良く頭部領域を切り出し可能で、かつ両目や口、鼻などの顔部位の検出も可能となる。また、携帯電話のような計算資源が制限されたハードウェアにおいてもリアルタイム処理が可能となる。

3.2 リアルタイム頭部領域抽出及び顔部位検出法

図2にアルゴリズムの概要を示す。このアルゴリズムは、均等拡散モデルを用いてモデル化された肌色モデルを用いて顔領域の候補を求め、その結果と領域分割結果に基づき、顔の輪郭に沿って顔領域を抽出する。

そして、顔領域の上部に接している領域を髪毛領域として求め、抽出した顔領域と髪毛領域を合わせて頭部領域として抽出する。

また、抽出した顔領域から両目や口などの部位に相当する領域を抽出し、各部位を特定する。

3.2.1 均等拡散モデルを用いた肌色領域抽出

この手法で用いる肌色モデルは、画像データベース分野において開発されたものである。前述のとおり、この肌色モデルは、均等拡散モデルを用いることにより、物体形状及び光の影響を除き、撮影条件が未知の画像に対してもロバストに肌色領域を抽出可能である特長を持っている。

具体的には、次式で示す C_1-C_2 正規化色相空間において肌色をモデル化したものである。

$$C_1 = c_2 / c_1 \dots\dots\dots(1)$$

$$C_2 = c_3 / c_2 \dots\dots\dots(2)$$

$$c_1 = \arctan\left(\frac{R}{\max\{G, B\}}\right), c_2 = \arctan\left(\frac{G}{\max\{R, B\}}\right),$$

$$c_3 = \arctan\left(\frac{B}{\max\{R, G\}}\right) \dots\dots\dots(3)$$

3.2.2 色とエッジ情報を用いた顔領域の切り出し

前述の肌色モデルを用いた肌色抽出結果のみを用いて顔領域を特定する場合、ハイライトや影の影響から領域の一部に亀裂などの欠損が存在するため、必ずしも顔の輪郭で切り出せない問題がある。そこで、顔領域境界を正確に特定するために、画像領域分割結果を用いる。

画像の領域分割手法には様々な手法が提案されているが、なかでも色による領域分割が最も単純である。しかしながら、色による領域分割では、クラスタリングの色数を少なく設定してしまうと過少分割を起こす問題がある。逆にクラスタリングの色数を多くすると、過剰分割を起こす問題がある。そこで、この手法では、色数を少なく設定して領域分割を行った結果、図3に示すように、エッジ抽出結果を用いて過少分割を起こしている領域を修正することにした。

ここで、色による領域分割は、色の量子化手法を用いた。また、エッジ抽出は、Cannyエッジ抽出法を用いた。

3.3 頭部領域の特定

以上で求めた肌色領域抽出結果と領域分割結果に基づき頭部領域を抽出する。まず、肌色領域抽出結果から、大き

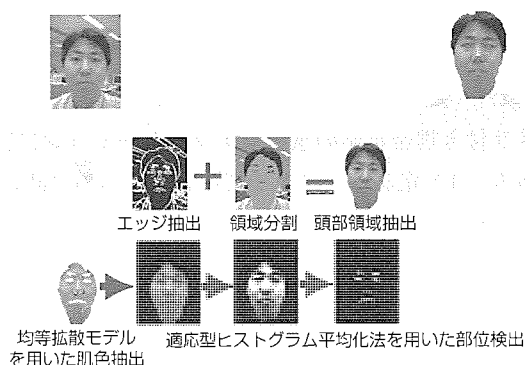


図2. 頭部領域抽出及び顔部位検出法

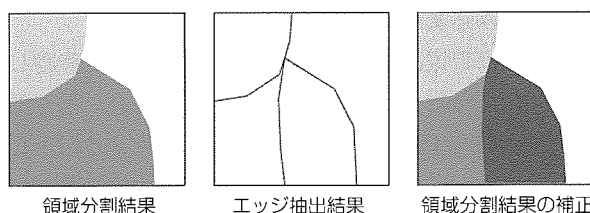


図3. 画像の領域分割手法

さが最大の肌色領域を顔領域候補とする。そして、領域分割結果から、顔領域候補と重なる領域を求め、それを顔領域とする。

次に、顔領域の上部に接している領域のうち、あらかじめサンプリングした数種類の髪毛色と領域の平均色との色差が閾値(いきち)内の領域を髪毛領域とする。

最後に、顔領域と髪毛領域を合わせて頭部領域とし、領域内部に発生した穴を埋めて抽出する。

3.4 リアルタイム顔部位検出

両目や口、鼻孔などは、肌領域に比べて輝度が低いため、輝度を手掛かりに検出することが可能である。しかしながら、固定の閾値を用いた場合、照明条件の変化に伴う輝度の変化に対応することができず、安定した検出が困難となる。

そこで、この手法では、3.3節で検出した顔領域全体に適応型ヒストグラム平均化法を適用することにより、固定閾値でも安定して部位領域候補を検出可能とした。

適応型ヒストグラム平均化法は、通常の平均化処理が画像全体の輝度分布を平均化するのに対し、あらかじめ設定された小領域内の局所的な輝度分布を平均化する手法である。つまり、場所によって異なる輝度分布に適応して平均化処理を行うため、画像に依存せず良好に画像のコントラストを向上させることができる特長がある。

実験の結果から、照明条件が異なる場合においても、固定の閾値で各部位の候補領域をほぼ抽出可能であることが分かった。

以上の処理により、両目や口、鼻の候補領域を抽出した後、3.3節で抽出した頭部領域の大きさから推定される各部位の位置に基づき、各部位に対応する領域を特定する。

4. 顔画像の安定化と追跡の高速アルゴリズム

4.1 概要

カメラ付き携帯電話の重要なアプリケーションにTV電話がある。TV電話は、これまでの音声のみの対話に話者の顔を相互に見ることができるため、音声のみの対話に比べ、よりお互いの理解が深まる効果がある。そのため、携帯電話にTV電話機能を搭載することにより、これまでの音声通話の制約を超え、ビジネス及びプライベートにおけるコミュニケーションをより円滑にすることが可能となる。しかしながら、TV電話機能を搭載することで操作が煩雑になってしまうことがあってはならず、これまでの操作系に自然に組み込まれる必要がある。

ここで、携帯電話に搭載されたカメラを用いて移動しながら被写体を撮影する場合を考える。この場合、携帯電話を持つ手は絶えず動いており、ユーザーはカメラが正しく顔の方向に向いているかどうかを常に意識する必要が生じ、会話への集中を妨げ、動画送信によって会話効果を向上しようというメリットを打ち消しかねないものとなる。ま

た、カメラと顔の距離が変わると顔が画面に対して大きすぎたり小さすぎたりするという問題もある。図4の上部にこれらの問題の例を示す。

これらの問題を解決するために、英国の三菱電機Visual Information Laboratoryでは、顔画像の安定化及び追跡(Face Stabilisation and Tracking : FST)技術を開発した。

このシステムでは、カメラで撮影された映像からリアルタイムに顔を検出し、かつ、追跡することにより、常に被写体が中央に位置するように撮影することを可能にするものである(図の下部)。

従来、顔検出及び追跡に関する様々な手法が提案されている(参考文献(1)に顔検出及び追跡手法のサーベイが紹介されている)。しかしながら、開発した提案手法は、従来手法に比べ次の利点がある。

まず、提案手法は照明条件の変化及び被写体の肌色の違いに対してロバストである。次に、従来手法に比べ、約10倍程度計算量が少ない。最後に、メモリ使用量が1Kバイト以下と非常に少ない。これらの特徴は、携帯電話に搭載するための条件に適している。

以下に、この手法の詳細を述べる。

4.2 FST技術の詳細

FSTは、映像からの色ベースの顔領域検出と追跡から構成されている。図5にアルゴリズムのフローチャートを処理結果の例とともに示す。

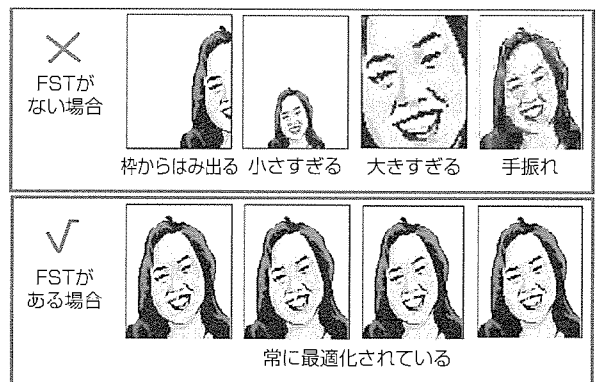


図4. FSTの実施例

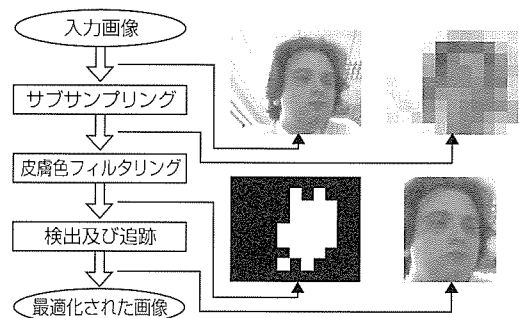


図5. FSTの処理手順

このアルゴリズムでは、まず、キャプチャしたビデオフレームをある大きさのブロック単位にサブサンプリングし、以降の処理にかかる計算量を減らしている。それと同時に、肌色以外の顔領域、例えば口ひげや眼鏡等からこのアルゴリズムが受ける影響を軽減することができる。

次に、肌色画素をサブサンプリングした画像から抽出する。肌色は、顔検出にとって非常に重要な特徴量であるが、照明条件の変化によって肌色の範囲が変化する問題がある。そこで、画像から取得した肌色情報(RGBの値)を肌色抽出に最適な別の色空間にマッピングする。これは、対数の色空間でLogRGと呼ばれるもので、照明条件の違いを吸収することができる特長がある。

肌色抽出では、サブサンプリングした画素値をLogRGに変換し、あらかじめLogRG空間でモデル化した肌色モデル(ヒストグラムによって表現)と比較することで、画素が肌色か否かを判定する。さらに、提案アルゴリズムでは、肌色モデルを使用者の肌色や使用時の照明条件に合わせて最適化するキャリブレーション機能を持っている。

最後に、肌色抽出結果(肌色マップ)を基に顔領域を検出し、追跡する処理を行う。

それにはまず、雑音除去のために、このマッピング情報に対して膨張・収縮処理を行う空間フィルタにかけられる。次に、連続成分分析を行い、肌色を持つ画素がグループ分けされ個別の肌色領域が作成される。次に、一連の条件に従って、どの肌色領域が一つ前のビデオフレームの顔領域に相当するかが決定される。これらの処理により、顔領域のみが符号化され、受信側に送信される。このシステムはそのほか、顔領域追跡の速さと滑らかさを調整する機能も持っており、また使用用途によっては、動画像情報の中に様々な顔がある場合、送信画像が常に一番顕著な被写体であるように、検出対象を自動的かつスムーズに切り換えることも可能である。

FSTアルゴリズムは、加算とごく少数の除算を含む整数演算のみで実行できるため、携帯電話に使用されている計算資源の乏しいプロセッサへの適用も可能となっている。また、前述のとおり、メモリ使用量も1Kバイト以下と非常に少ない。さらに、ビデオフレームをメモリに格納することなく、パイプライン処理によってアルゴリズムを実行することも可能である。

広範なテストを行うために、パソコンベースのリアルタイムFSTデモ機を開発した(図5の画像は、このデモ機によって作成した)。このアルゴリズムをパソコン(Pentium 4, 1.8GHz)上に実装したデモ機では、QCIFサイズ画像の処理にCPU時間の10%未満しか消費しておらず、その処理時間の大半もデータのハンドリングと表示に費やされている。顔画像追跡としてQCIFビデオ画像を1秒当たり10フレームの速度で処理させるために、このアルゴリズムの全体演

算に必要な性能は大よそ10MIPSである。

ベンダーに依存しない一般的なVHDLモデルを使用したFPGA(Field Programmable Gate Array)ベースの実装に関するFSTシステムの設計とシミュレーションも行った。このシステムは多様なFPGAへのインプリメントが可能であり、ASIC(Application Specific IC)への移行のためのフィジビリティスタディにも使用可能である。設計では、Altera APEX 20K 1000FPGAデバイスの8%以下の領域で実現可能であり、33MHzのクロックスピードにおいて、1秒間に400フレームの処理が可能である。

4.3 結 論

提案した顔画像の安定化と追跡技術を使えば、携帯電話の使用感を飛躍的に向上させると考える。すなわち、被写体を常に中央に撮影するための操作を必要としないことから、ユーザーは会話に集中することができる。顔画像が自動的に追跡されるので、画面から顔が外れることなく、常に中央に大きく表示できることから、表示サイズが小さい携帯電話には最適であると言えることができる。また、FSTシステムを使ったアプリケーションでは、MPEG-4やH.263の符号化効率も向上している。

この章では、今回開発されたリアルタイムの顔画像の検出と追跡を簡単な計算で行う技術を紹介した。これまでの実験により、このアルゴリズムが様々な照明条件下及び異なる肌色に対して確実に動作することが分かっている。このように、この技術は、将来の携帯電話に新しい独自の機能を加えるものとなるであろう。

5. む す び

携帯電話において、画像に関する機能や性能の向上は目覚ましいものがある。本稿では、その一例としてカメラを備えた携帯電話を例に挙げ、カメラを活用したアプリケーションを考えていく際に考慮すべきカメラ制御やカメラ画像処理技術のポイント、及びカメラ画像を有効に生かすアプリケーションを実現させる技術やアルゴリズムについて紹介した。

これまではカメラを搭載した携帯電話やPDAが市場に出始めた市場の立ち上がり期であったが、今後一気に普及が進むと予想される。その普及期におけるニーズをあらかじめ予測し対応できる技術をシステム全体から総合的に構築していくことが重要である。また、既存の機器が組み合わせられたとき、本来の機器とは異なる機能や用途が新たに発生することを意識して技術開発を進めることも重要であろう。

参 考 文 献

- (1) Yang, M. S., et al.: Detecting Faces in Images-A Survey, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24, No. 1, 34~58 (2002)

マルチプロジェクタ技術

ラメシュ・ラスカル*
 ジェロエン・ファン・バー**
 芦崎能広***

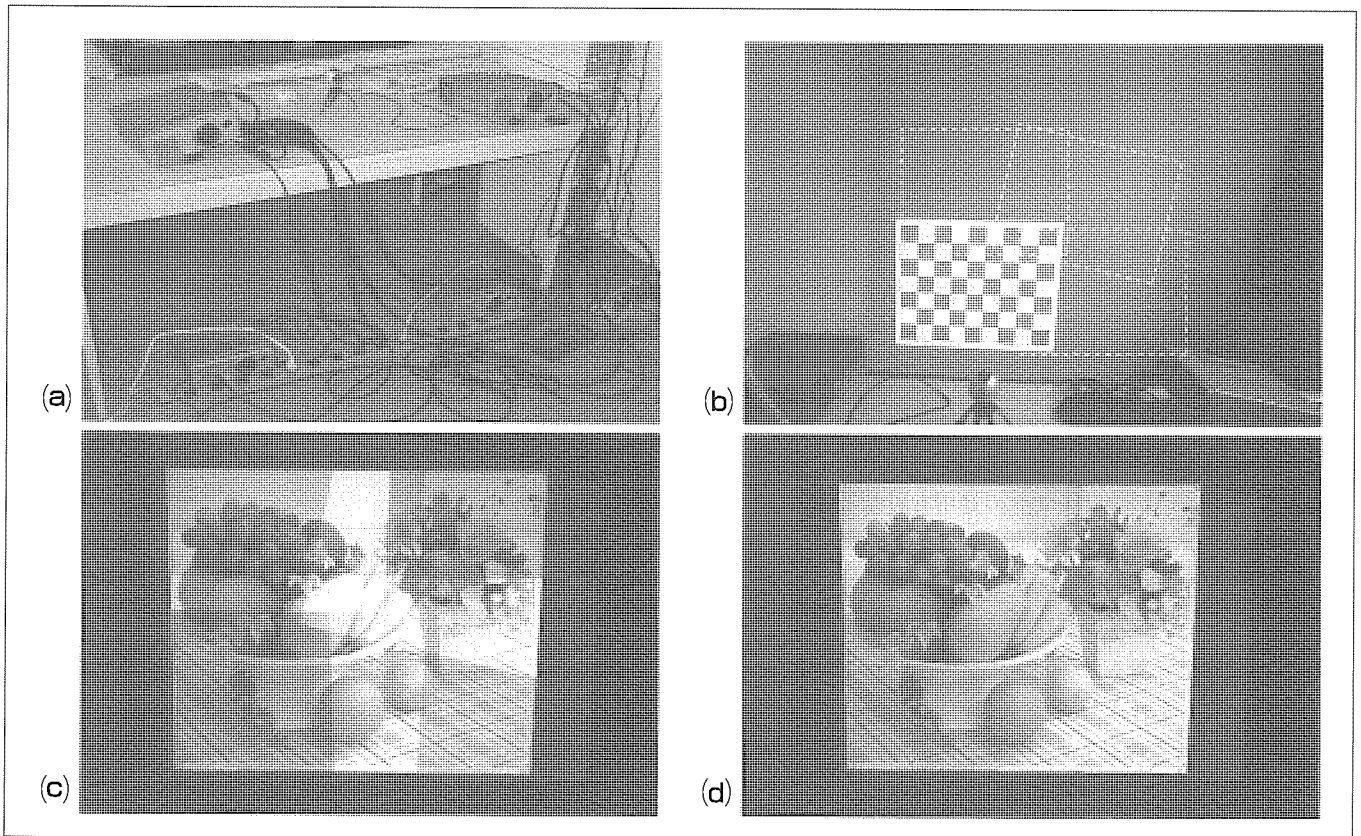
要 旨

今日、プロジェクタを複数台並べて高精細な大画面表示を実現するマルチプロジェクタシステムは、基盤設備と保守が高コストなため、シングルプロジェクタよりも単体当たりの価格がかなり高額である。これは、マルチプロジェクタシステムが、多くの場合、隣接する個々のプロジェクタ画面の表示を違和感なく連続させるために、手動で微妙な補正をする必要がある、通常は、熟練者によってこの調整が行わなければならないためである。今般、この補正コスト削減を目的として、大雑把に設置された重なりのあるプロジェクタ画面を使って、繋(つな)ぎ目のない大画面表示システムを作り出すマルチプロジェクタの技術を開発した。

開発した技術により、プロジェクタの画面と画面を繋ぎ合わせる画像調整を完全に自動化することができる。2 ×

2に配列された四つのプロジェクタからなるシステムの画像を補正するのに要する時間は20秒以下と短く、しかも、セットアップが容易である。また、この技術により、精密な支持構造は不要となり、手動補正を不要にするのと併せて低価格化を実現している。

このマルチプロジェクタシステムでは、複数のプロジェクタ間の相対的位置関係を自動的に計測するためにデジタルカメラを利用している。システム全体の補正では、表示画面上に幾何学パターンを投影し、画面間で重なり合った画像の位置合わせと輝度の混合を行っている。表示画像を幾何学的に変形させるための効率的なレンダリング法についても述べる。また、表示画面上に繋ぎ目のない矩形(くけい)画像が形成される様子を実験結果とともに示す。



システムの外観と自動調整中の画像

(a)は繋ぎ目が重なるように大雑把に設置された複数プロジェクタの外観、(b)は幾何学パターンの表示である。これにより、ひずみを除くためのパラメータを決定する。表示画像の輪郭は黄色で示されている。(c)はプロジェクタの各画面を重ね合わせた単一矩形ディスプレイ画像、(d)は重ね合わせた部分に輝度のクロスフェーディングを行った繋ぎ目のない表示画像である。

1. ま え が き

複数のプロジェクタをタイル状に並べたマルチプロジェクタシステムは、解像度が高く、明るい画像が得られる現実的な解決策であるため、広く用いられている。

現状の比較的廉価なマルチプロジェクタシステムでは、各プロジェクタに精密な機械的調整を行うための装置がついており、画面と画面との連続性を手動調整する。また、廉価な背面投射型のプロジェクタの場合、隣り合う画面の繋ぎ目には小さいが目に見える目地がある。最近のシステムの中には、前面投射型のマルチプロジェクタのみならず、背面投射型でも、各プロジェクタの画面と画面とを重ね合わせて画像の位置合わせやクロスフェーディングを行うことにより、目地を見えなくしているシステムがある。このシステムには、プロジェクタの画面を調整するために純粋な機械的装置ではなくて電気機械的調整が可能な装備が備えられており、システムをセットアップする際には、各プロジェクタで表示される画面の容易な位置調整を行うことができる。しかし、スクリーンへの投影の正確性を調整するのはやはり手動であり、これでは、依然としてかなり面倒である。おそらく、この手間が大画面表示システムを構築する上で、最大の問題点になっていると考えられる。この調整プロセスの自動化技術を開発することにより、セットアップの煩わしさを軽減することができる。

2. マルチプロジェクタ

本稿で紹介するマルチプロジェクタは、自動調整機能付きプロジェクタであり、所望の縦横比を持った単一矩形画像を生成することができる。それぞれのプロジェクタは、スクリーン上で任意に多少傾いて設置されていてもよい。要旨のページの図(a)のように、市販の標準的なプロジェクタ4台と低価格のデジタルスチルカメラ(以下、単に“カメラ”という。)を用いて、試作機を開発した。

2.1 概 要

我々のマルチプロジェクタシステムは、1台の固定されたカメラで得られる画像と、異なる位置に置かれた複数のプロジェクタによる画像との間の射影により、繋ぎ目を自動的に調整することができる。射影は3×3のホモグラフィを用いて記述される(3.1節)。図1と図2はマルチプロジェクタ内での処理の流れとアルゴリズムの概略を示したものである。次節では、この処理の原理を数学的に解説する。

2.2 相対的配置

画像座標間の関係の定義は一对のホモグラフィによって定義される。まず、プロジェクタの画像を P_i で表す。ここで $i = 1, \dots, n$ は各プロジェクタに付けられた番号であり、試作装置では $n = 4$ である。また、カメラの画像を C で表す。

同次座標で表される座標値を考えると、 $u = (u, v, 1)^T$ で与えられるプロジェクタ画像 P_i のピクセルとそれに対応する $x = (x, y, 1)^T$ で与えられるカメラ画像のピクセルとは3×3のホモグラフィ $H_{c,i}$ によって次のように式で関係が表される。

$$u_i \cong H_{c,i} \cdot x_i \quad i=1, \dots, 4 \dots\dots\dots(1)$$

これにより、任意の隣り合う二つのプロジェクタ間のピクセルマッピングが次のように決められる。

$$u_j \cong H_{c,j} \cdot H_{c,i}^{-1} \cdot u_i \dots\dots\dots(2)$$

ここで、 u_i と u_j はそれぞれプロジェクタ P_i と P_j に対応するピクセルを表している。結局、もし表示画像が繋ぎ目のない画像として映し出されるなら、表示画像の画面座標はカメラの入力画像の座標 x に等しいことになる。

3. 補 正

繋ぎ目のない画像を作るためには、①プロジェクタの表示画面が幾何学的に重なり合うこと、②重なり合う部分の輝度が重なり合わない部分の輝度と同じに見えること、の2点が必要となる。

3.1 幾何学的アラインメント

ここでは、プロジェクタのピクセル間の関係を計算する方法について述べる。同じ関係は、投影された画像の位置合わせを確実にするためのレンダリングでも使われる。

まず、最初にそれぞれのプロジェクタから順次、格子状の幾何学パターンが投影され、次に1台のカメラによってディスプレイ表面上に投影された画像が記録される(図3(a))。二次元のカメラ画像から特徴点を抽出し、それに対

<p>前処理段階</p> <ul style="list-style-type: none"> ●各プロジェクタに対して <ul style="list-style-type: none"> -幾何学パターンを投影 -カメラ撮像によって幾何学パターンから特徴を抽出 -カメラとプロジェクタ間のホモグラフィを計算 ●重ねて投影された映像の中から使用可能な矩形表示部分を決定 <ul style="list-style-type: none"> -プロジェクタと画面表示領域の間のホモグラフィを計算 -輝度をクロスフェーディングするための重みを計算 <p>レンダリング処理段階</p> <ul style="list-style-type: none"> ●各プロジェクタにおいて <ul style="list-style-type: none"> -ホモグラフィを使って入力画像を幾何学ひずみ補正 -テクスチャマップを用いたクロスフェーディングによる輝度補正
--

図2. 平面上に単一画像を繋ぎ目なしで表示するためのアルゴリズム概要

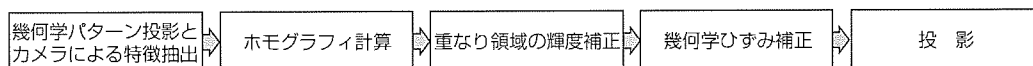


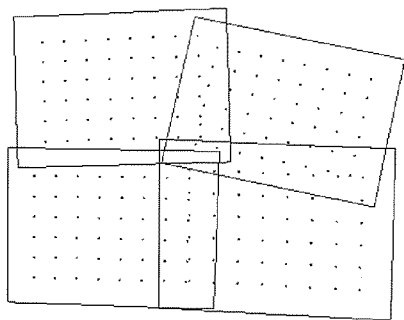
図1. マルチプロジェクタシステム構築のための処理の流れ

応する点を二次元のプロジェクトパターンから得ることにより、式(1)におけるカメラと個々のプロジェクトとの間の 3×3 ホモグラフィが決定される。ホモグラフィは線形最適化を使って計算する。

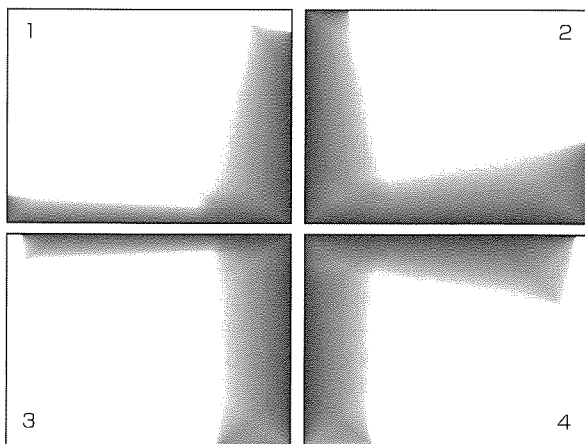
3.2 輝度混合

複数のプロジェクトによって重なり合って映し出された部分の映像はより明るく見え、要旨ページの図の(c)のように目立つ。そこで、繋ぎ目のない重なりを実現するために、一般にはクロスフェーディングとして知られた輝度混合の技術を採用している。各プロジェクトに対し一つのマスクを作成し、そのマスクはプロジェクトの各ピクセル u の輝度に対して $[0.0, 1.0]$ の範囲の重みを付ける。既にどの部分が重なり合っているか分かっているため、最終出力画像の各ピクセルに対し、そのピクセルが幾つのプロジェクトでカバーされているのか分かる。ピクセル u_i の重みは総和が1になるよう調整される。図3では、この装置で使われた四つのプロジェクトの配置が一例として示されている。プロジェクト画面の縁に近いピクセルにはゼロに近い重みを、重なり部分の境目に近い部分には1に近い重みをかける。この方式により、異なるプロジェクトにまたがる画像を繋ぎ目なしで連続させることが可能になる。

現実的な問題として、計算上の小さな残差や時間経過と



(a) カメラ座標系に記録された四辺形の映像。点はチェッカーパターンの特徴を表している。



(b) 四つのプロジェクトに対する輝度の重み。暗い領域はゼロに近い重みを持ち、明るい領域は1に近い重みを持つ。

図3. 輝度混合された各画面

ともに起きるプロジェクトの位置の乱れ、電気機械的振動などにより、ある時間を通して完全にプロジェクトを調整することは極めて困難である。

4. レンダリング

最終出力は、普通に配置された各プロジェクトからの投影が斜めに、つまり正射影になっていない状況でも、矩形入力画像からひずみと繋ぎ目を取り除いたものでなければならぬ。矩形表示画像を得るために、各プロジェクトは幾何学変形された入力画像が投影される。表示されないピクセルは黒のままにしておく。レンダリングのステップは、各プロジェクトで同じである。

投影された画像に必要な幾何学ひずみ補正は、プロジェクト画像と原画像のピクセル座標間のホモグラフィによって決められる。このホモグラフィは、それぞれのプロジェクト i に対して $H_{c,i}$ として既に計算されている。幾何学ひずみ補正は、平面状の表示表面に斜めから射影することによって生じるひずみの逆に処理することと考えることができる。

各プロジェクトにおいて、入力画像を幾何学変形するためのレンダリングプロセスは独立に行われる。入力画像はテクスチャメモリに格納される。次に、このテクスチャマップを xy 平面上の単位長方形に当てはめ、レンダリングの際にはホモグラフィ $H_{c,i}$ を透視投影行列とする。入力画像の幾何学ひずみ補正の変換は手近なグラフィックスハードウェア、すなわちパソコングラフィックスカードで行うことができる。

次のステップは、重なり合った部分の輝度補正である。図3(b)に示したようなピクセルごとの輝度の重みを第2のテクスチャマップのアルファチャンネルに格納する。このアルファチャンネルに対する色は黒に設定しておく。同じ矩形がこのテクスチャで塗り直される。アルファチャンネルは透明度として働き、変形した入力画像の輝度に掛け合わされる。

5. 応用と構成

自動調整機能付きマルチプロジェクトは広い範囲で応用可能であり、その範囲は監視室の大画面モニタからホームシアターにまで及ぶ。博物館、学校、移動展示などにも、非常に使用が簡単なこの装置は利用できる。次の3点が同程度に大きい画像を生成する単一プロジェクトシステムに勝る大きな利点である。第1は、投影距離が短縮されるので、店舗や狭いオフィスのような小規模の設備でも収容することができる。第2に、マルチプロジェクトでは、表示可能な輝度(総ルーメン値)とピクセル解像度が単一プロジェクトに比べて高い。正確なサブピクセルレベルでの位置合わせを実現しているため、更なる輝度が必要な場合、2

表1. マルチプロジェクタの構成とこの技術の適用

例	システム構成	適用
1	背面投射型のマルチ大画面で、画面間の繋ぎ目があり、重なりがない場合	サブピクセル単位での調整に適用可能
2	背面/前面投射型のマルチ大画面で、画面間の重なりがある場合	繋ぎ目なしで輝度混合が可能
3	背面/前面投射型のマルチ大画面で、画面間にひずんだ矩形の重なりと多少の傾きがある場合	繋ぎ目なしで輝度混合、幾何学ひずみ補正が可能
4	重なりのある非平面マルチ大画面の場合	任意の画面構成での繋ぎ目なしで輝度混合が可能

台又はそれ以上のプロジェクタ画面を重ね合わせることも可能である。最後に、マルチプロジェクタは、任意の縦横比を持つ画像や矩形以外の形の画像を生成することが可能であるという利点がある。

調整保守のための精密な支持構造が不要となるため、システムコストの大幅削減をもたらす。また、表1に様々な形態でのプロジェクタ利用の特長を示す。

このアルゴリズムは、手動で行うと非常に煩雑になる調整を自動化するのに適しており、任意の数、配列のマルチプロジェクタの補正に対しても用いることができる。

6. む す び

幾何学パターンの投影と、幾何学ひずみ補正及び輝度の混合により、高画質で繋ぎ目のないマルチプロジェクタシステムの技術を開発した。この技術は、装置を設置する際に必要となる調整の時間を、ウォーミングアップ時間と同程度に短縮した。デジタルカメラやTVと同じように、ウォームアップ時間が10秒以下で、調整プロセスが完全自動であれば、ユーザーはこのような大画面で映画やゲームを楽しむことができる。また、単に大きな卓上ディスプレイ

として使うことも可能である。卓上や壁面に即座にプロジェクタ画面を映したい場合など、プロジェクタの配列を迅速かつ自動的にセットアップできることは使い勝手の上で不可欠の要件である。

この技術開発の主要な成果は、①位置合わせパラメータと輝度調整用の重みを迅速かつ効率良く決めることができる技術、②三次元グラフィックスハードウェアを使い、ホモグラフィを利用して完璧(かんぺき)な幾何学ひずみ補正を行う技術である。これらの成果により、低価格でセットアップ操作も簡単なシステムの構築が可能となる。

(注) このプロジェクトのホームページ:

<http://www.merl.com/projects/ProjectorMosaic/>

参 考 文 献

- (1) Raskar R., et al.: A Low Cost Projector Mosaic with Fast Registration. Proc., Asian Conf. on Computer Vision (2002)
- (2) <http://www.3d-perception.com/>
- (3) <http://www.comview-vs.com/>



ネットワーク上の 文書テキスト検索・利用技術

鈴木克志*
高山泰博*

要 旨

文書の電子化が進み大量の情報がネットワーク上に存在する現状は、情報過多又は情報過負荷(Information Overload)と呼ばれる。このような状況においては、計算機に結合されたネットワーク全体の文書資源を活用する文書の検索機能が、IT(情報技術)における重要な課題となってきた。

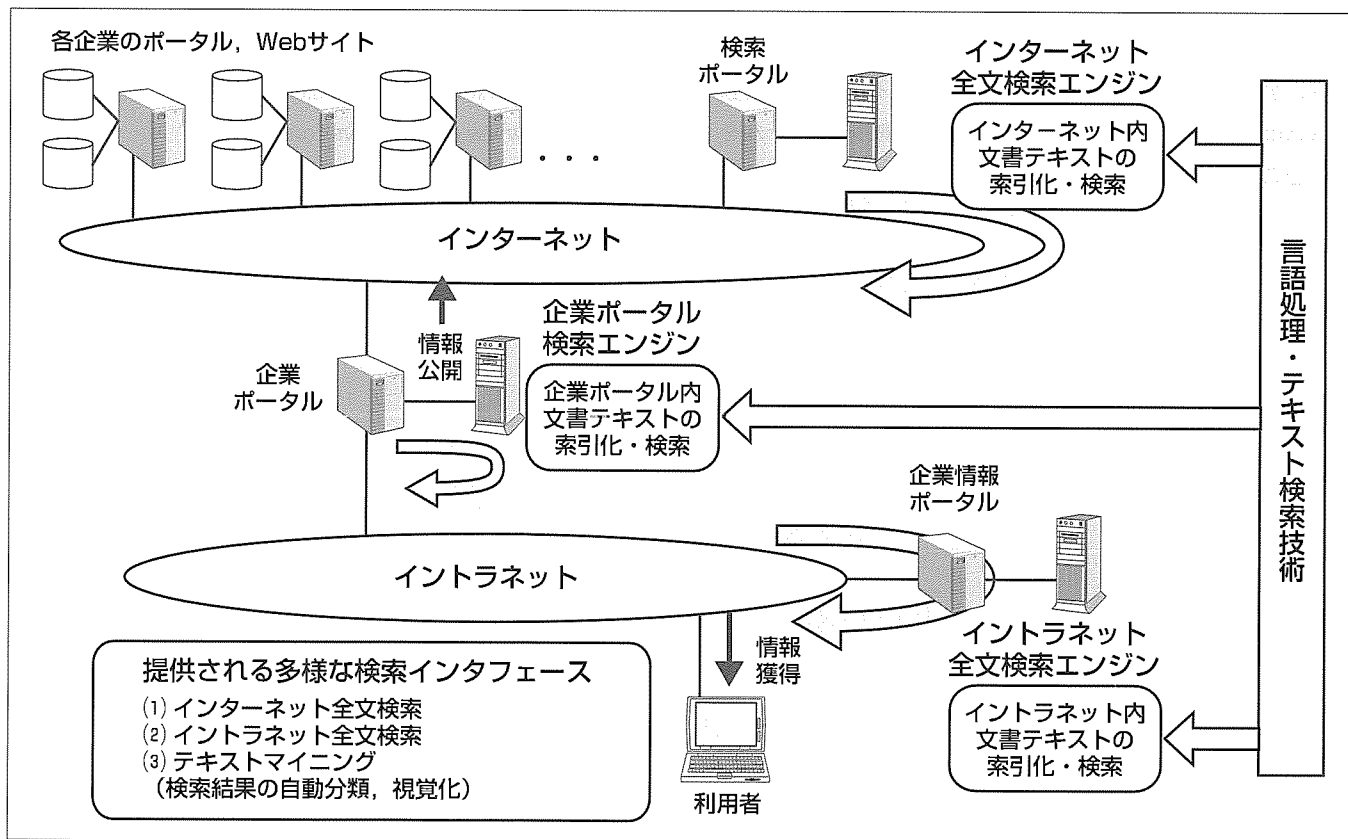
一方、ヒューマンインタフェースの観点から見ても、文書の検索機能は、ネットワークと人間を介するインタフェースの重要部分を占めていると考えられる。

文書検索ソフトウェアは、利用者から見て計算機システム全体のバックエンドに位置するプロセッサである。しかし、近年、バックエンドに位置付けられる検索エンジンの呼出しが、フロントエンドでの操作機能として重要な位置を占めるようになった。すなわち、最近では、オペレーティングシステムでの不可欠機能としてフロントエンドイン

タフェースに全文検索の入力窓が搭載されるようになり、Webの画面上に検索入力窓を設けているポータルも多くなっている。

さらに、検索結果を業務で利用する新しい応用が生まれてきており、そこでは、業務に応じた検索結果の後処理方式やフロントエンドインタフェースにおける検索結果の表示方式などの工夫が行われている。

そこで、本稿では、ネットワーク全体の文書資源を活用するための文書テキスト検索技術の重要性に着目し、文書テキスト検索の基本原則、検索方式の分類、及び重要な基盤技術としての形態素解析について述べる。さらに、検索ポータル、企業ポータル検索エンジン、インターネット情報収集、及びテキストマイニングなど、三菱電機の言語処理技術やテキスト検索技術によって実現可能となった先進の応用事例について紹介する。



ネットワーク上の文書テキスト検索ソリューションの概念

先進の言語処理技術とテキスト検索技術により、従来困難であったネットワーク上の大量文書の活用が正に可能な時代を迎えた。利用者は、種々の検索エンジンが提供する検索ソリューションによって、多様な検索機能を利用することができる。検索ポータルは、インターネット全文検索エンジンによって大量のWeb文書の検索機能を提供する。また、企業情報ポータルは従業員に企業内のテキスト検索を提供し、企業ポータルはポータル訪問者にコンテンツへの検索機能を提供する。これらの検索結果は、目的に応じたテキストマイニング機能によって業務に活用される。

1. ま え が き

文書の電子化が進み大量の情報がネットワーク上に存在する現状は、情報過多又は情報過負荷と呼ばれる。この状況において、計算機に結合されたネットワーク全体の文書資源を活用するために、文書の検索機能が非常に重要な課題となってきている。そこで、本稿では、ネットワーク上の文書資源を活用する文書テキスト検索技術について、検索方式や基盤技術の内容を述べ、当社の提供する先進的技術と新しい応用事例を紹介する。

2. ネットワーク上の文書テキスト検索技術

2.1 文書テキスト検索の基本原則

近年、高速検索のための索引作成技術が進展し、記憶装置の低価格化・大容量化に伴って、大規模な索引の実現と、それによる大量文書の高速度検索機能実現が可能となった。

ネットワーク上の大量の文書テキストを高速に検索するためには、テキストに対する索引をあらかじめ文書登録時に作成し、文書検索時には索引を用いて高速にアクセスする仕組みが必要である。索引を作成せずにテキストをファイル先頭から後方に向けて入力文と照合する方式もあるが、計算時間がかかるため、1台のパソコン内の文書などに用途が限定される方式であり、大量の文書に対しては実用的ではない。

2.2 検索方式の分類と動向

主な文書テキスト検索方式を機能と索引作成の種類によって分類するとすれば次のようになる。最近、(2)が普及し、(3)も目的に応じて実用化段階に入った。

(1) 統制語検索機能

古典的方式であり、人手で文書に付与した統制語(キーワード)を索引として文書を検索する。

(2) 全文検索機能

テキスト中の任意の箇所を検索する。次の二方式がある。

(2-a) n-gram索引：任意のn文字列を検索する。

(2-b) 単語索引：任意の単語を検索する。

(3) 類似文書検索機能

内容が類似する文書を検索する。次の二方式がある。

(3-a) 単語-文書索引：文書に出現する単語を要素とするベクトル形式で索引を作成する。

(3-b) 単語次元圧縮索引：ベクトルの次元を圧縮する。

2.3 形態素解析技術

日本語文書に対する単語全文検索(2-b)や文書検索(3)では、形態素解析技術が不可欠な技術である。形態素解析とは、分かち書きされていない日本語文章を単語の列に自動分割して索引を作成する処理のことを言う。図1に、形態素解析を用いた全文検索方式の原理を示す。

形態素解析ソフトウェアは、辞書を参照しながら入力文

を単語の列に分割する。索引登録処理は、形態素解析ソフトウェアの出力結果の中から索引として有効な索引語を選択し、文書の識別子(URL等)とともに索引データベースに登録する。

形態素解析の精度は、検索精度に大きな影響を与える。日本語の場合、精度向上のために単語辞書を充実させる必要があるが、単に語数を増やすだけでは十分でなく、単語間の接続関係を豊富に持つことが重要である。図2に、複合語の形態素解析が誤っていたために検索できない例を示す。

誤りの原因は、辞書を参照して単純に分割すると多数の分割可能性が生じることにある。複数の正解候補から最適解を選択するため、辞書、文法、及び統計情報など、長年蓄積された研究成果やノウハウを活用する必要がある。

2.4 三菱日本語形態素解析エンジン

当社は、20年以上に及ぶ日本語処理の研究開発の蓄積を基に業界トップクラスの性能を持つ形態素解析ソフトウェアを実用化している。表1にその仕様を示す。このソフトウェアは高速性と高精度を両立しており、インターネット全文検索など当社の応用システムに差別化技術として適用されている。

3. テキスト検索の応用

3.1 インターネット検索ポータル

3.1.1 検索ポータルの役割

検索ポータルは、インターネット全文検索エンジンを備

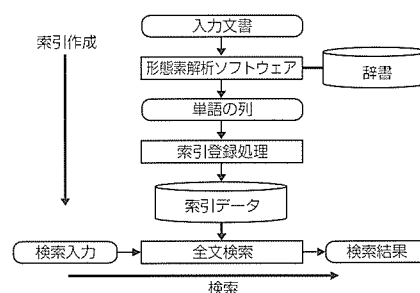


図1. 全文検索方式の原理

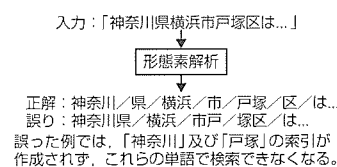


図2. 形態素解析の例

表1. 形態素解析ソフトウェアの仕様

項目	仕様・特徴
処理方式	独自の最適解探索アルゴリズム
速度	400Mバイト/時間以上(CPUクロック 550MHz)
精度	単語分割精度99%以上
辞書	基本語辞書、固有名詞辞書、複合語辞書(30.5万語)
制御	マルチスレッド対応

え、様々な情報サービスを提供することで利用者の便宜を図り、利用者がインターネットで最初に訪れるサイトとなることをねらう。利用者の数が多ければ多いほど広告収入増加が見込めるので、検索ポータル事業者にとっては、魅力的な検索エンジンを搭載することが事業に不可欠である。利用者にとっては、検索ポータルサイトがインターネットへの入り口として、フロントエンドインタフェースの役割を果たす。

3.1.2 検索ポータルとインターネット検索エンジン

インターネットには20億ページ以上のWeb文書が存在すると言われ、しかも、その量は日々増大している。従来、検索ポータルは自主開発の検索エンジンを搭載しているサイトが多かったが、インターネット上の文書量の増加につれて検索エンジンの拡張や保守コストが増大したため、市販の検索エンジンを導入するケースが増えてきている。

検索ポータルに提供されるインターネット検索エンジンは、①これらの大量の文書に対する索引作成と管理、②多くの利用者からの検索要求に対する検索レスポンス確保、及び③種々雑多なWeb文書の検索結果の中から検索結果としてふさわしいものを選択する等、高度な技術集積によって実現可能である。索引方式としては2.2節の(2-b)の単語索引が主流であり、索引作成のための形態素解析ソフトウェアの性能が重要である。

当社の関連会社であるTOCC社(トラフィック・ワン・コミュニケーションズ, <http://www.tocc.co.jp/>)は、Inktomi社との協業によって検索ポータルに対する検索エンジンのサービス事業を行っている。これには2章で述べた当社の形態素解析ソフトウェアが使われている。

3.2 企業ポータル

3.2.1 企業ポータルの役割

企業ポータルは、今後伸びが期待されるB2BやB2Cの場を提供する企業のインターネット上の入り口となるWebサイトである。いわば、企業ポータルは、インターネット時代における企業と顧客のインタフェースである。

3.2.2 企業ポータル検索エンジンの特徴

企業ポータルでは、ポータル内のコンテンツに対する検索機能が必要である。従来の企業ポータルにおける検索機能は単語入力による全文検索が中心であり、ポータル内の情報が増加してくると、一覧表示された検索結果から所望の情報を取り出すのが困難になってくる。また、ポータル内の情報はディレクトリとして階層的に整備・管理されているが、全文検索結果の選択に十分活用できていない。顧客が検索したときに入力に関連する複数のディレクトリが表示されれば、関連する製品やサービスが容易に把握できるようになる。

例えば、個人向け、法人向け両方の製品に関する情報を提供している企業ポータルにおいて、“テレビ”という入力に対して“カラーテレビ”及び“映像・画像システム”等が分

類して表示されれば、個人ユーザーであれば“カラーテレビ”の欄を、法人ユーザーであれば“映像・画像システム”の欄を選択すればよいことになる。全文検索ではこれらの文書が混在して一覧表示されるが、検索した文書をディレクトリごとに分けて表示すれば、全文検索で埋もれたり混在したりしていた文書を探しやすくなる。

3.2.3 三菱電機新Webサイト

当社Webサイト(<http://www.MitsubishiElectric.co.jp/>)では、関連するディレクトリの優先的検索を提供している。これは、入力文に関連するディレクトリ優先付けを2.2節(3)の類似文書検索技術によって実現し、全文検索結果を各ディレクトリごとに分類することにより、所望の検索結果を得やすくすることに特徴がある⁽¹⁾。このために、ディレクトリ全体の単語出現頻度傾向を代表するベクトル索引を用意する。構成を図3に示す。

3.3 インターネット情報収集

3.1節で述べたインターネット検索ポータルでは、漏れなくWeb文書を検索することが最重要であるため、非常に大量の検索結果が見付かる。そのため、所望の検索結果を獲得するためには、入力単語を追加し、AndやOrのオペレータで結合することによって検索結果を絞り込むことが必要となる。しかし、一般の利用者にはこの操作は煩雑な場合がある。ここでは、2.2節(3)の文書ベクトル検索によってこの課題を解決したインターネット情報収集システムの例を示す。文書ベクトル検索を利用し検索結果を絞り込むためのインタフェースとして、以下の工夫を行っている⁽²⁾。

- (1) Webページの構文情報チェックを併用する。
- (2) 適合性フィードバックを利用する。
- (3) ベクトルの重みを利用者が編集可能とする。

(1)は、URLのチェック、リンク集かどうかのチェック、掲示板かどうかのチェック、日付の利用などである。

また、(2)は、複数の検索結果に対して適否の判断を利用者が与えることで、システムが不要ページの順位を下げ必要ページの順位を上げるようにベクトルの重みを自動調整する機能である。

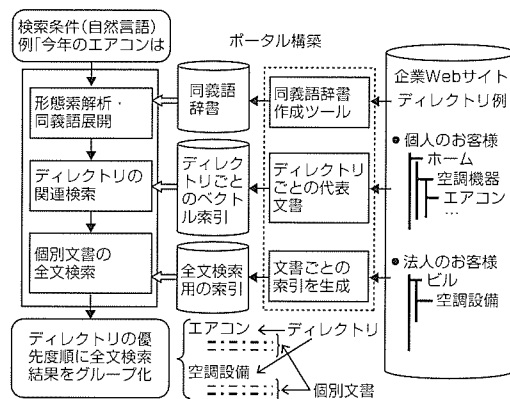


図3. 三菱電機Webサイトでの高度検索方式

(3)は、ベクトルの各要素(個々の単語)の重みを利用者が修正できるようにすることで、通常の全文検索エンジンの利用のように単語及びそのAnd/Orの組合せを考えながら操作するよりも負担が少なくなる。図4に、インターネット文書収集システムの構成を示す。

インターネット文書収集システムでは、まず利用者がキーワード(単語)を入力してインターネット全文検索結果を求める。その中から、検索結果としてふさわしい文書をサンプル文書として利用者が指定する。関連文書検索ソフトウェアは、2.2節(3)の類似文書検索機能によってサンプル文書と類似する文書を全文検索結果の中から抽出する。これにより、大量の全文検索結果から所望の文書を抽出する作業が著しく効率化される。

3.4 テキストマイニングと検索結果の視覚化

3.4.1 テキストマイニングの概要

テキストマイニングとは、固定的なアルゴリズムの呼称ではなく、テキスト処理や言語処理の要素技術の組合せによって特定の目的に応じて行う“大量のテキスト中から目的に有効な情報を自動抽出する処理”の総称である。システムの構成としては、テキスト検索が検索条件を与えて検索結果を求める処理であるのに対し、テキストマイニングでは業務目的に応じた情報抽出処理を行う。また、テキスト検索の結果は見付かった文書が優先度順にソートされたものであるのに対し、テキストマイニングの結果は目的に応じて様々な内容・形式となる。

3.4.2 テキストマイニングの具体例

以下のような具体例がある。

(1) テキスト情報又は検索結果の視覚化

WWW情報の視覚化やテキスト検索結果の視覚化により、利用者に情報フィードバックを与え、段階的な検索要求の改良を支援する。また、新たな知識発見のきっかけを与える。

(2) 文書の内容に応じた自動分類

ファイリングにおける文書の登録支援、コールセンターやポータルにおける顧客からのメールの自動分類、営業報告の自動分類、インターネット上のWWW文書自動分類、及び検索結果の自動分類などがある。

(3) 目的に応じた特定情報の自動抽出

インターネットや配信ニュース記事からの特定トピック抽出(例：企業動向、新製品情報)、インターネット掲示板からの企業風評抽出などがある⁽³⁾⁽⁴⁾。

3.4.3 実用的テキストマイニング

テキストマイニングは、形態素解析ソフトウェアを基盤とし、構文解析や言語パターン抽出と呼ばれる高度な言語処理技術を組み合わせて実現する。これらは新しい技術であり実用化の課題もあるが、当社では、単なるキャッチフレーズにとどまらない“実用的なテキストマイニング”として、下記のような実用化を行っている。

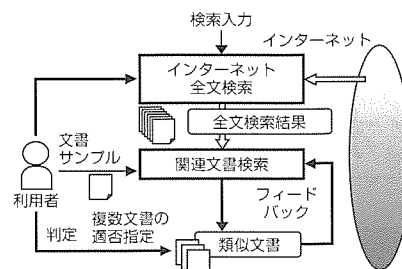


図4. インターネット文書収集システム

(1) ポータル検索結果の自動分類

3.2節で述べた企業ポータルの検索結果の自動分類技術は、ポータルを差別化する機能として適用されている⁽¹⁾。

(2) 顧客からのメールの自動分類

当社CRM(Customer Relationship Management)ソリューションで顧客クレーム抽出機能として活用されている⁽⁵⁾。

(3) 不正WWW文書の自動フィルタリング

TOCC社のインターネット検索ポータルサービスでアダルト文書の自動検出に活用されている。

(4) インターネットの類似文書自動抽出⁽²⁾

3.3節で述べたインターネット情報収集は、大量の全文検索結果から類似文書だけを抽出するマイニングを実現している。

4. む す び

本稿では、ネットワーク上の大量の文書テキスト情報に対する検索技術について述べた。従来のキーワードによる検索では大量のネットワーク資源の有効活用が不可能であり、先進の言語処理・テキスト検索技術によって初めて膨大なインターネット資源の活用が可能となる。その実例として、当社が提供している様々な文書検索ソリューションについて紹介した。今後は、更に日々増大を続ける情報洪水現象に対し先進の技術を提供していく予定である。

参 考 文 献

- (1) 高山泰博, ほか: 企業ポータルにおける関連ディレクトリ検索, 情報処理学会第64回全国大会, 4Y-01 (2002)
- (2) 永井明人, ほか: 文内の単語共起照合に基づくクレーム抽出方式の性能評価, 情報処理学会第64回全国大会, 2X-03 (2002)
- (3) 増塩智宏, ほか: 情報抽出におけるHTML文書の表示構造の解析, 情報処理学会第64回全国大会, 2Y-01 (2002)
- (4) 高山泰博, ほか: インターネット全文検索と関連文書検索を用いたポータル構築支援, 2001年電子情報通信学会情報・システムソサエティ大会 (2001)
- (5) 佐々木 誠, ほか: e-ビジネスCRMソリューションへの取組, 三菱電機技報, 75, No.4, 293~296 (2001)

情報化オフィスにおける文書インタフェース技術

岡田康裕*
 亀代泰三*
 平野 敬*

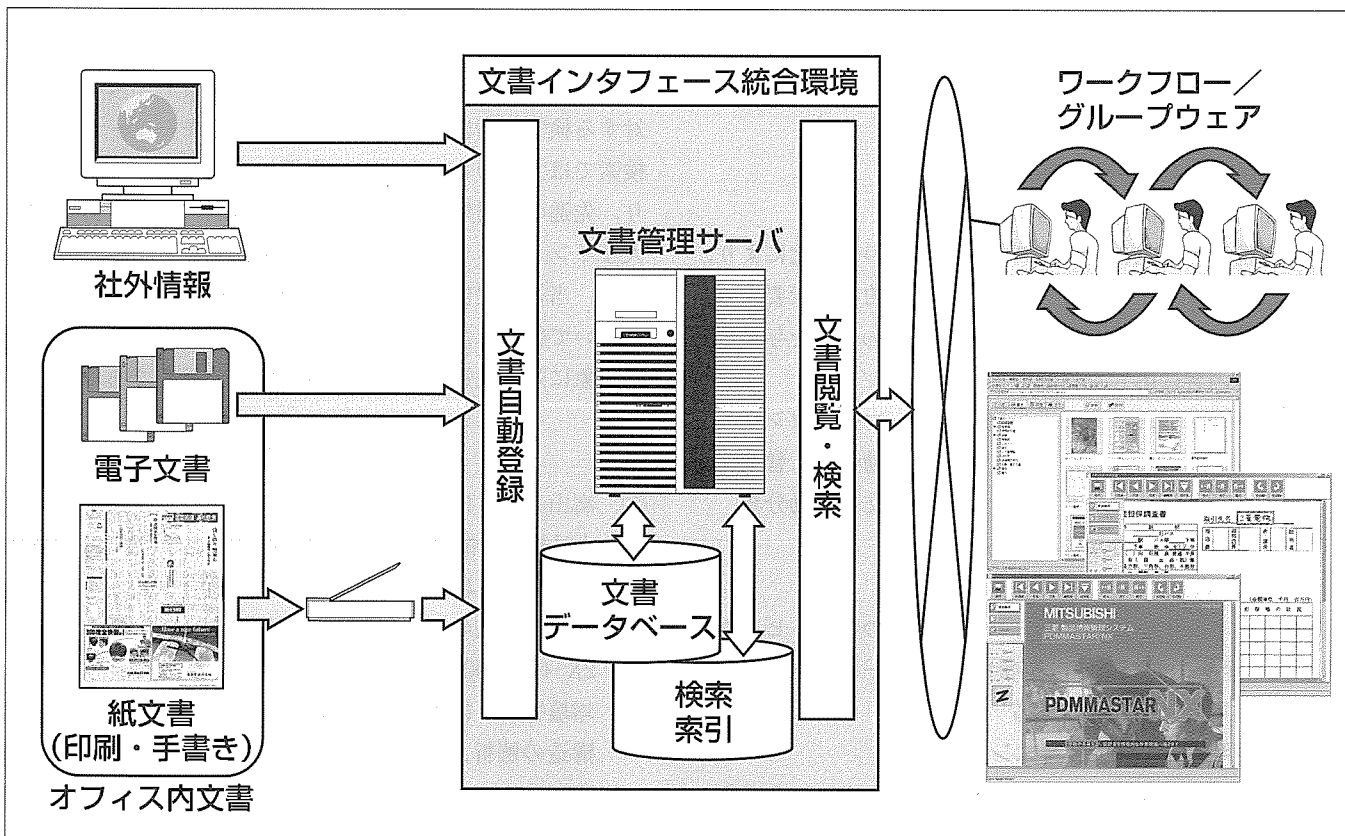
要 旨

過去から現在に至るまで、オフィスにおける人と人、人と組織間のコミュニケーション手段は、文書による情報伝達がその主役を担っている。文書はオフィスに欠くことができない存在であり、文書を用いたインタフェースはオフィスにおけるヒューマンインタフェースの一部であるとも言える。近年では、社内情報の活用と業務効率化のためにグループウェア、ワークフローシステムなどの導入が進み、文書の電子化・共有化が行われるようになった。一方で、依然として紙文書による情報流通も避けられず、情報化オフィスでは様々な種類・形式の文書が散在している。これらの文書資源を有効活用するためには、各所に散在する紙文書や各種電子文書をシームレスに管理し共有する文書インタフェース技術が必要となる。

このニーズにこたえるため、三菱電機では、電子文書・

紙文書双方に対して自動的に検索索引を作成する文書自動登録技術と、一覧性の確保、原文に忠実な文書閲覧、視認性の高い検索結果表示を実現する文書閲覧・検索技術を開発している。特に紙文書の登録・検索においては、当社独自の形状特徴併用検索方式によって紙文書をスキャナ入力するだけで文書イメージに対して高精度に全文検索することができる。

上記技術は、三菱電機インフォメーションシステムズ(株)の図面文書管理システム“FINALFILING”と、三菱電機アプリケーションサービス(株)がASP事業としてサービス提供している文書図面管理サービスに適用されている。今後は、文書情報を有効活用し企業活動の原点である新たな価値創出に資する文書情報活用技術を開発していく予定である。



文書インタフェース統合環境を用いたオフィス文書の電子化・共有化

オフィスにおける人と人、人と組織間の重要なコミュニケーション手段である文書を用いたインタフェースを円滑に行うための文書インタフェース統合環境を実現した。電子文書・紙文書双方に対して自動的に検索索引を作成する文書自動登録技術により、オフィスに散在する文書を社外情報とともに文書管理サーバに登録する。文書の閲覧・検索は、一覧性の確保、原文に忠実な文書閲覧、視認性の高い検索結果表示を実現した。既存のワークフローとグループウェアとの協調動作により、オフィス内の文書資源の有効活用を図ることができる。

1. ま え が き

過去から現在に至るまで、オフィスにおける人と人、人と組織間のコミュニケーション手段は、文書による情報伝達がその主役を担っている。文書はオフィスに欠くことができない存在であり、文書を用いたインタフェースはオフィスにおけるヒューマンインタフェースの一部であるとも言える。近年では、社内情報の活用と業務効率化のためにグループウェア、ワークフローシステムなどの導入が進んでおり、文書の電子化・共有化が盛んに行われるようになった。しかし一方で、依然として紙文書による情報流通も避けられない状況が続いている。

本稿では、上記のオフィス文書を取り巻く環境に着目し、オフィスにおいて重要な資源である文書情報を有効に活用するための文書インタフェース技術とその応用製品について解説する。

2. オフィス文書の電子化・共有化とインタフェース技術

オフィス業務で用いられる文書は多岐にわたり、図1に示すように、様々な文書が様々な目的で活用されている。これらの文書を効率的に作成・流通・蓄積・再利用するとともに、企業内外で作成された文書資源を活用し新たな価値を生み出すことが、現代の企業経営において重要な事項であることは言うまでもない。以降では、情報化オフィスにおいて電子化された文書を取り扱う上での課題を述べた後、企業内の文書資源を活用するために必要となる当社開発の文書インタフェース技術について解説する。

2.1 情報化オフィスにおける文書を取り巻く課題

この節では、オフィスに散在する文書を有効活用する上での課題について解説する。

2.1.1 文書の電子化・共有化の目的

オフィスにおける文書の電子化・共有化の目的には、下記のようなレベルがある。

- 目的レベル1：文書の保管管理に要するコスト削減
ストック型の文書を対象として、書棚のスペースコストや文書管理に要する人件費削減などを目的とするもの
- 目的レベル2：文書を探し出す作業の効率化
大量の文書を高速検索できる機能を活用して、文書探索

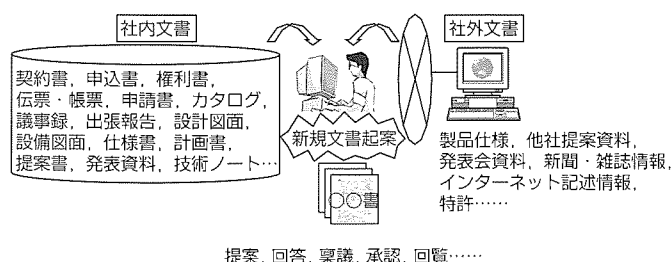


図1. オフィス業務で活用される文書の例

に要する時間・手間を削減することを目的とするもの
目的レベル3：文書による社内外情報伝達の効率化
文書の閲覧・承認・通知など、フロー型の文書管理を電子的に行い、フローに伴う人手作業と時間を省き、業務の効率化を目的とするもの

目的レベル4：文書資産の有効活用

電子化・共有化を図った個々の文書情報を有効活用して、新たな価値創出のために資することを目的とするもの

目的レベル1～3に対しては、その目的に応じて文書ファインディング、ワークフロー、グループウェアなど様々なシステムが製品化されている。システム構築に当たっては各目的レベルに合った個々のシステムを組み合わせるが、運用が煩雑にならないように留意する必要がある。今後は、目的レベル4を実現するための技術開発が必要となる。

2.1.2 文書の多様化

(1) 文書種類の多様化

スキャナやネットワークの普及によって新たな文書入手経路が開拓され、オフィス内で取り扱う電子文書の種類は多様化の一途をたどっている。従来からのワープロ文書に加え、カタログ、メール文書など多岐にわたる文書が電子化されている。また、設計・製造分野においては、図面や写真も文書情報として取り扱う必要がある。このような様々な文書を有効活用できる技術・システムの実現が望まれている。

(2) 文書ファイル形式の多様化

電子化される文書の増大に伴い、文書ファイル形式も多様化している。オフィスにおいて、表1に示すようなファイル形式の文書を対象として情報活用を図る必要がある。

2.2 企業内文書資源を活用するためのインタフェース技術

2.1節で示した課題を解決するために、当社では、種々の文書を自動蓄積し自由に閲覧・検索できる技術の開発に取り組んでいる。以下、具体的な内容について解説する。

2.2.1 文書の自動登録技術

種々の文書に対して検索索引を自動生成する技術を開発している。図2に示すとおり、紙文書をスキャンした画像、ワープロ文書、HTMLファイル内に画像表現された文書に対しても自動的に検索索引を作成することができる。以下、自動索引作成を実現するための技術を紹介する。

(1) 形状特徴併用検索方式

紙文書に対しては、図3に示すように、検索索引作成時

表1. 文書ファイル形式の例

文書の種別	ファイル形式
特定アプリケーション文書	Word, Excel, PowerPoint ^(注) , PDFなど
イメージ文書	BMP, TIFF, JPEGなど
Web文書	HTML, XMLなど

(注) “Word”“Excel”“PowerPoint”は、Microsoft Corp.の商標である。

に文字認識処理を行うとともに文字パターンの“形状特徴”を付与した形態で検索索引を構成する。検索時には、文字認識誤りが発生した場合でも、形状特徴を用いて照合をとる形状特徴併用検索方式によって正確な全文検索を実現する。これにより、単純な文字認識を行った際に問題となる文字認識誤りの修正作業を行うことなく、紙文書をスキャンするだけで文書を自動登録・検索することが可能となる⁽¹⁾。

(2) カラー文書読み取り技術

社内外で発行されるカタログなどのカラー文書をそのまま保存したいという要求がある。カラー文書に対する自動索引付けを実現するために、輝度・色相において文字線と考えられる局所変動を抽出するカラー文書文字列抽出方式を開発し、反転文字や背景色が濃い文字の読み取りを実現した(図4)。

(3) 印刷・手書き混在文書読み取り技術

過去に蓄積された紙文書の電子化要求にこたえるため、印刷文字列と手書き文字列の自動判定機能と、個々の文字列特性に合わせた文字認識処理を開発した。印刷・手書きが混在した文書の自動索引付けを可能とする(図5)。

2.2.2 文書の閲覧・検索技術

電子化された文書の閲覧・検索では、文書の視認性を配

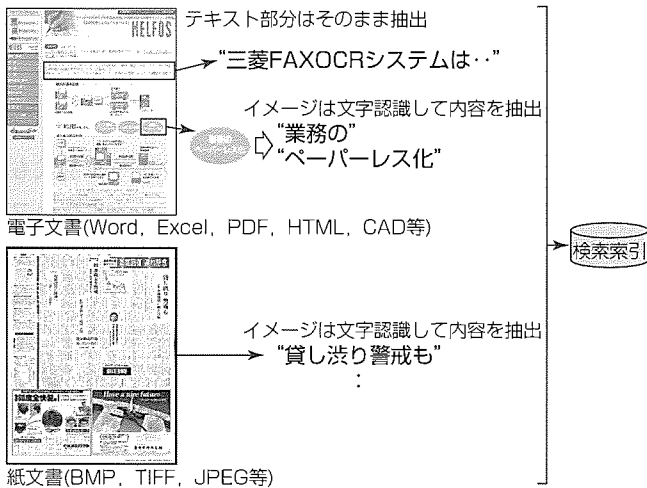


図2. 検索索引自動作成の概要

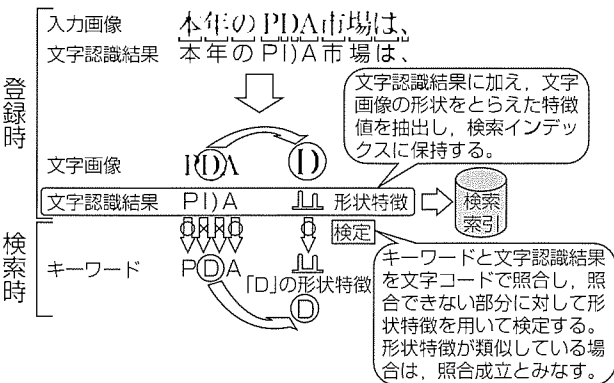


図3. 形状特徴併用検索方式の概要

慮した画面インターフェースが必要である。当社では、一覧性の確保、原文に忠実な文書閲覧画面、視認性の高い検索結果表示を目指し、下記に示す機能を提供している。

(1) サムネイル表示

文書画像を圧縮表示し一覧性を向上する(図6)。

(2) キーワード照合位置表示

文書はレイアウト情報も含めて忠実に表示し、同一画面上にキーワードと照合した位置を表示する(図7)。

(3) 検索結果ダイレクトアクセス

検索結果はページ単位に表示され、参照したいページにダイレクトアクセスが可能である(図8)。

(4) フルWebによる文書管理機能

フルWeb化された上記文書閲覧・検索機能により、既存のWeb環境で文書の閲覧・検索が可能である(クライアントに特別な閲覧ソフトウェアのインストールが不要)。また、文書登録機能もWebインターフェースで提供する。

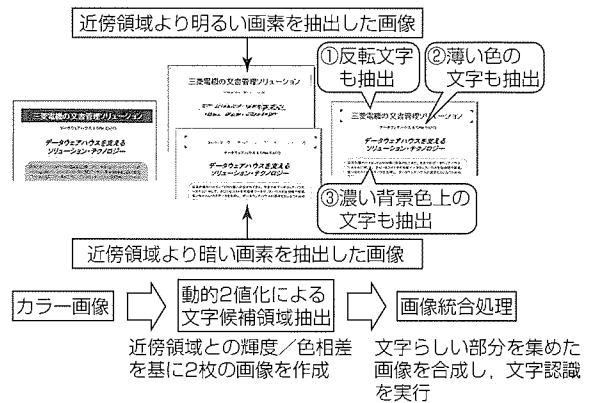


図4. カラー文書に対する文字列抽出方式

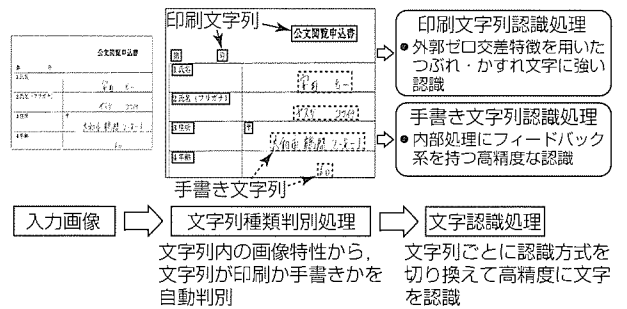


図5. 印刷・手書き混在文書読み取り方式

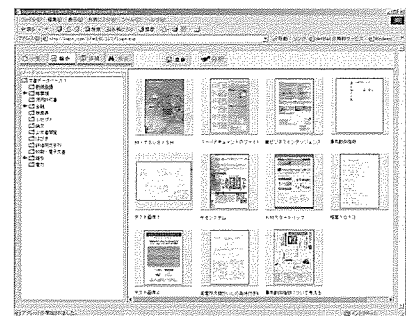


図6. サムネイル文書画像表示の例

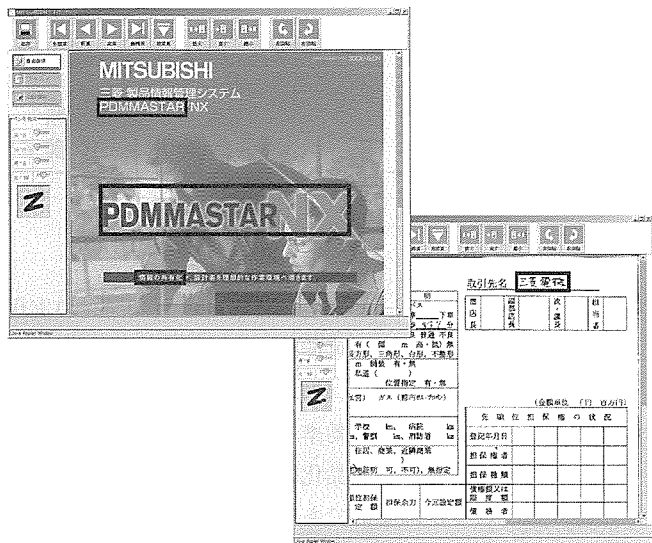


図7. イメージ文書に対するキーワード照合位置表示の例
(黒枠がキーワードヒット箇所。三菱電機, PDMMASTAR, 情報の共有化がキーワード検索された例)

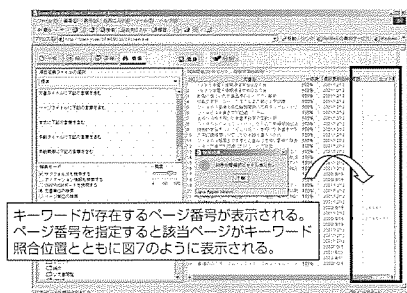


図8. 検索結果表示画面の例

3. 情報化オフィスを実現する文書管理製品

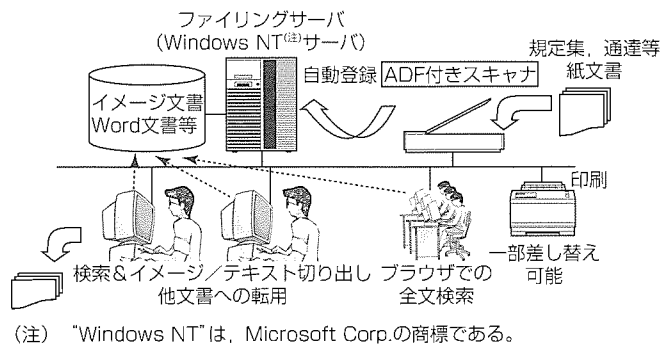
この章では、2章で示した文書インタフェース技術の実用化事例として、当社の文書管理システム及びサービスの製品を紹介する。

3.1 図面文書管理システム“FINALFILING”⁽²⁾

三菱電機インフォメーションシステムズ^(株)では、上記技術を搭載した図面文書管理システム“FINALFILING”^(図9)を製品提供している。この機能を搭載することにより、紙文書をスキャンするだけで文書内に記述された文字列に対して全文検索することが可能となる。

3.2 文書図面管理サービス⁽³⁾

三菱電機アプリケーションサービス^(株)では、データセンターに上記FINALFILINGシステムを設置し、文書図面管理のサービス提供を行っている^(図10)。お客様から預かった紙文書、ワープロ文書、表計算、CADデータなどを電子化し、データセンターのサーバに保管する。登録した文書・図面はインターネット経由で検索が可能であり、ブラウザソフトのみで距離が離れた事業所間や協力会社との情報共有が容易に実現できる。規格書、図面、カタログ管理などに適用されている。



(注) “Windows NT”は、Microsoft Corp.の商標である。

図9. 三菱図面文書管理システム“FINALFILING”の概要

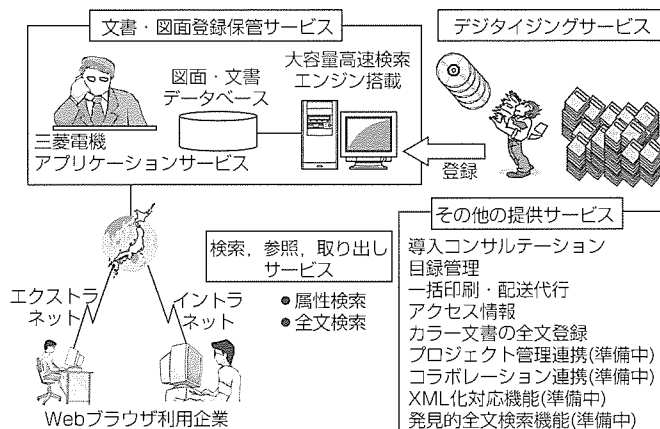


図10. 文書図面管理サービスの概要

4. むすび

企業活動において重要な情報交換媒体である文書を用いたインタフェース技術について解説した。オフィスに散在する紙文書や各種電子文書をシームレスに自動登録・閲覧・検索でき情報共有を図る文書インタフェース技術は、企業内の情報の壁をなくし人と人、人と組織間のコミュニケーションを円滑に行う上でなくてはならない技術である。今後は、電子化・共有化を図った上で個々の文書情報を有効活用して、企業活動の原点である新たな価値創出に資する文書情報活用技術を開発していく予定である。

参考文献

- (1) Kameshiro, T., et al.: A Document Image Retrieval Method Tolerating Recognition and Segmentation Errors of OCR using Shape-Feature and Multiple Candidates, Proceeding of the Fifth International Conference on Document Analysis and Recognition, Bangalore India, 681~684 (1999)
- (2) <http://www.mdis.co.jp/service/eoa/finalfiling/finalfiling.html>
- (3) <http://www.dsplaza.com/service/bunsho/index.html>

インタフェースデザイン評価技術

若松正晴*
沢田久美子*

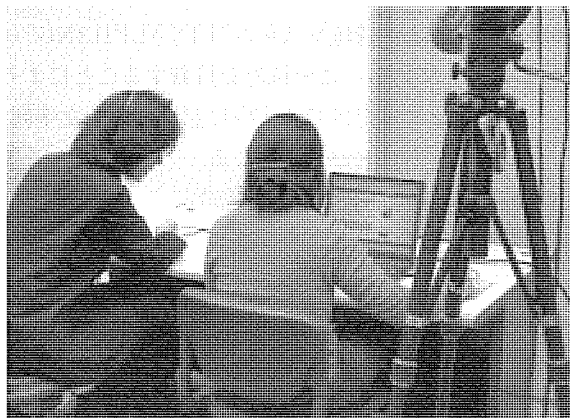
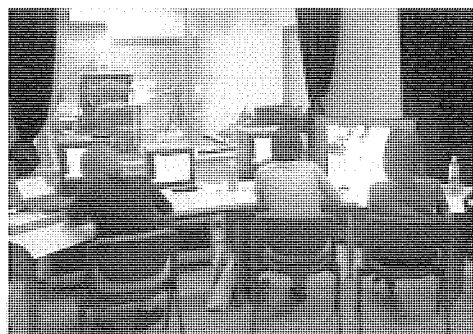
要 旨

三菱電機のデザイン研究所では、インタフェースデザインの評価改善をユーザビリティワークショップという手法を導入して実施している。この背景には、第三者、すなわちユーザーの視点で使いやすさを評価検証し、だれにでも使える製品を開発し提供していくことが、これからのメーカーにとって重要な課題になってきているという認識がある。

ユーザビリティワークショップは、評価スタッフのみでなく、デザイン、設計、企画、営業、品質管理など対象製品の開発にかかわる様々なメンバーの参画を前提としており、多面的な評価の実施と開発関連部門間の問題意識の共

有化によって、ユーザーとの親和性の高い製品開発の実現を目指している。評価は、可能な限り開発の上流段階から実施し、問題点の早期発見、早期解決を行うことで、工程の後戻りを少なくし、開発効率を高めることが期待できる。一製品に対する評価は、製品化の各プロセスに対応して複数回繰り返すことが理想的であるが、少なくとも一回でも実施すれば、ユーザビリティの品質を着実に向上させることができる。

本稿では、この評価手法の概要を解説し、事例として、車載機器(カーナビ)とWebサイトそれぞれの評価方法、評価に基づく成果の一端を紹介する。



ユーザビリティワークショップの実験シーン

左上：ドライビングシミュレータで運転操作をしながらタスクを実行する車載機器評価実験の被験者。左下：マジックミラー越しの車載機器評価実験観察状況。右上：Webサイト評価実験を観察する開発関係者。右下：ノートパソコンの画面でタスクを実行するWebサイト評価実験の被験者。

1. ま え が き

1.1 情報を介したインタフェース

人と道具の関係は、人類誕生以来、技術の進展に伴って、次第に、間接的に、複雑になってきた。特に近年の情報通信関連製品においては、ユーザーの最終ゴール(実行したい目的)と具体的な操作対象(インタフェース)との間に複雑な情報処理過程が存在し、直感的にこの過程を理解することは困難になってきた。これら情報を介したインタフェースにおいては、製品の機能を実現する際の人と道具との対話(相互作用)をユーザーにとって自然で分かりやすい仕組みにしていけることが、機能の完成度を高めることと同様に、又はそれ以上に重要であるという認識が次第に高まってきた。

1.2 ユニバーサルデザインの視点

一方、2015年には我が国の人口の4分の1が65歳以上の高齢者になるという予測にもあるように、急速な高齢化が進む中、今後の製品はマーケットボリュームの拡大する高齢者層をも視野に入れ多様なユーザーを想定した開発を目指す必要があるということも認識されつつある。高齢ユーザーは情報を介したインタフェースに対して十分に適応できないという場面が多く見られる。これは、高齢者以外の未経験ユーザーも含め、いわゆるデジタルデバイド(情報格差)の現象として、今後解決すべき重要課題となってきた。今後、ユーザーの多様性を十分理解しつつ、ユーザーの視点で、“使えるインタフェース”を目指していくことが必要である。

1.3 インタフェースデザインの評価

ユーザーに適合したインタフェースを開発するためには、開発の初期段階からそのデザイン案をユーザーの視点で評価し、改善提案をデザインにフィードバックするというサイクルを繰り返すことが有効である(図1)。前に述べたようなインタフェースへの認識の高まりと相前後して、当社では、デザイン開発の中で、ユーザーの視点でのインタフェースデザイン評価プロセスを確立し導入してきた。ユーザビリティワークショップと呼ぶこの手法の概要と成果の一端を紹介する。

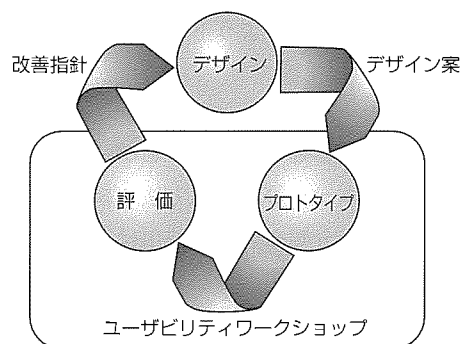


図1. 評価プロセス

2. インタフェースデザイン評価手法の概要⁽¹⁾

2.1 第三者視点での評価

D.ノーマンがいう“開発者のデザインモデルとユーザーのメンタルモデルとの食い違い”に気付きこれを解決することが、インタフェースデザインにとっては重要である。開発者は自らはユーザーにとって良かれと思いデザインを創造するのであるが、そのデザインが必ずしもユーザーにとって使いやすいとは限らないことがある。デザイン開発の中で客観的な第三者視点での評価プロセスを組み入れることの重要性はここにある。当社では、第三者視点での評価を実現するために、被験者を使うユーザーテストを基本的な方法として採用している。

2.2 評価方法

評価は、デザイン案のプロトタイプ(デザインスケッチ、仕様書、デザインモデル、パソコンシミュレーション、試作機など)を対象に、モデルユーザー(想定ユーザーの属性に近い被験者)に模擬的な対話(操作課題)、いわゆるタスクを実行してもらい、その過程での行動と発話(操作中の思考内容)を観察記録し、これを分析してインタフェースの問題点を抽出して、改善指針を導き出す、というプロトコル解析をベースとした実験的手法を用いる。評価指標は、モデルユーザーの行動と発話のデータ(プロトコルデータ)、タスク達成の所要時間や誤操作率、モデルユーザー自身の主観評価(インタフェースに対する印象)などである。経験上、インタフェース上の主要な問題点を発見するためには、被験者数5～6名で十分目標を達成できることが多い。ただし、定量的分析を行う場合は少なくとも数十名の被験者実験が必要である。

2.3 車載機器(カーナビ)・Webサイトの評価方法

車載機器(カーナビ)のインタフェースは、車の運転というメインタスクに付随したサブタスクとして位置付けられ、安全性が優先されるという制約条件がある。これを評価する場合、被験者にはドライビングシミュレータの運転操作と車載機器プロトタイプの操作を含む二重課題を実行してもらおう実験を行っている。車載機器操作と運転操作との両立は、被験者のステアリング操作の履歴を参考に、その妥当性を推定することができる。

一方、Webサイトのインタフェースは、使用するパソコンによってはページの見え方が異なってくるため、多様なハードウェアとOS・ブラウジングソフトへの対応を要求されるという特徴がある。また、企業のホームページでは、そのユーザビリティの良し悪しがアクセス数に大きく影響し、ブランドイメージそのものをも左右しかねない。Webサイトの評価では、プロトタイプをテストサイトとして立ち上げ、この中で被験者にタスクを実行してもらおうという実験を行っている。

3. ユーザビリティ評価の事例

3.1 車載機器(カーナビ)のユーザビリティ評価

現在のカーナビ市場では、従来のボタン操作やリモコン操作方式に加え、音声操作のできるタイプが増えつつある。

当社では2001年にV7000を発売している。デザイン研究所では、2000～2001年にかけて、これら音声操作のできる各社市販のカーナビ機種及び研究用シミュレーションモデルを対象に、ユーザビリティ評価を実施した(図2)。評価実験では、カーナビの代表的な操作タスクを設定し、ドライビングシミュレータを用いて、運転中、停車中の想定環境下における検証を行った。ここでは、これらの実験結果から得た、今後のカーナビのインタフェースを向上させる上での課題について述べる。

3.1.1 音声操作を用いるメリット(評価された点)

(1) 従来のカーナビでは、目的地設定などを行う際、画面遷移に従って順に絞り込んでいく必要があった。これに対し、音声操作では、あるステップを飛ばしてのショートカット操作やダイレクトに名称を入力することが可能になった。

(2) 画面を注視せずに操作ができるので、運転中の操作がやりやすい。

(3) 音声入力とともに音声によるガイダンスがあり、操作に困ったときの手助けとなる。

3.1.2 音声操作のインタフェースの課題

(1) 発話できるコマンド(音声認識辞書)がある程度限定されるため、コマンドの種類やその製品独自の文法を覚えなくてはならない。特に高齢者ユーザーは自分なりの表現をする傾向が強く、より自然に近い、慣れた言葉遣いで発話できるとよいと言える。また、“えーと”や“うーん”など操作に関係のない言葉の発話やコマンドの言い直しなどはあらゆるユーザーに起こることであり、これらに対するより許容性の高いものが望まれる。

(2) ユーザーは設計者の考える操作フローに従った思考ができるとは限らない。実験では、何を言えばいいかわからない、いつ発話すればいいかわからないといった場面が多かった。ショートカット操作やダイレクト操作ができる一



図2. カーナビ評価実験風景

方で、音声認識エラーも伴うため、いかに分かりやすいインストラクションを与えるかが重要である。

(3) 音声ガイダンスがあると便利な反面、タイミングがズレたり必要な情報が得られないと、逆に不安になったり不信感を抱いたりする。聞き取りやすい早さや大きさ、状況に応じたタイミングを検討する必要がある。また誤認識された場合、どう間違っただのか、次にどうすればいいのかといった適切なりカバー法を提示することも重要である。

(4) 一つの製品の中に音声操作とボタン操作2通りの操作方法が存在する場合、ユーザーは、一方の操作手順に習って他方も同様にできると考える。操作手順の一貫性やシームレスな使い分けを可能にすることが必要と考える。

(5) 機械に話しかけることへの抵抗感はやはりまだ大きい。“機械に強制されたくない”という感情、また、同乗者を含めた車内環境や、様々な運転環境下での使用に対する配慮が課題と言える。

3.2 Webサイトのユーザビリティ評価⁽²⁾

2001年4月に当社の公式Webサイトの大幅なリニューアルを行った。デザイン研究所では、Webサイトを統括・運営する本社宣伝部から依頼を受け、旧Webサイト及びリニューアルプロトタイプのユーザビリティ評価を繰り返し実施し、評価結果を踏まえ、より使いやすいWebサイト構築のためのインタフェース提案、及びWebサイトデザインのためのガイドラインを作成した。その後、再リニューアルのタイミングや、各事業部の個別サイト等に対しても、順次評価を実施し改善提案を行っている。ここでは、評価結果が反映されたりリニューアルサイトのインタフェースのポイント、及びガイドラインの概要について紹介する。

3.2.1 リニューアルサイトにおけるユーザビリティの主な特徴

(1) 旧サイト(図3)はトップページの情報量が多すぎ検索しづらいものだったが、画面の情報量を整理し、多様なユーザーの使用環境下でも影響されにくい表示サイズに統一し、スクロールは最小限にし、一覧性を確保した(図4)。

(2) ユーザーはトップページで迷うとほとんどすぐに閲覧をやめてしまう。入口で迷わないように利用目的に応じた入口(個人、法人別)を設け、最も利用頻度の高いと思われる製品情報検索をしやすいようトップページからの検索性を確保した。

(3) サイト共通のナビゲーションバー(ヘッダ)及びフッタを設け、トップに戻る等の共通のナビゲーションルールやメニュー・サブメニュー等の表示方法を統一する。

(4) 自社独自の製品分類や名称の使用は避け、ユーザーに分かりやすい一般的な用語、分類を用いる。

(5) ユーザーが当社のサイトであることが容易に認知できるように、ブランドロゴの扱いとページデザインイメージの統一を図る。



図3. 旧Webサイト (M's Diner)

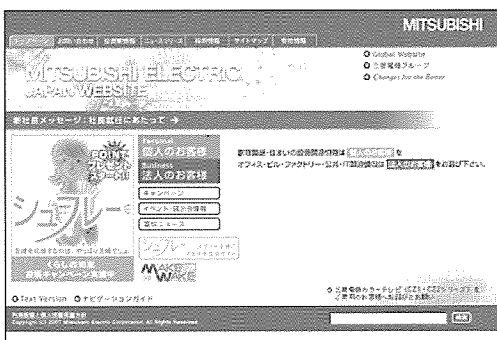


図4. 三菱電機新Webサイト

3.2.2 Webサイトデザインのためのガイドライン構築

(1) Webデザインガイドラインの考え方(指針)

始めに“分かりやすい・見てもらえる・使いやすい”デザインをガイドラインの考え方の基本とし、Webサイトのユーザーインターフェース指針を示した。

“分かりやすい”とは、サイトの全体像が分かり、見やすく、表現が分かりやすい等を示す。

“見てもらえる”とは、期待する情報があるかどうか分かりやすい、素早く確実に表示する、だれにでも使える等を表す。

“使いやすい”とは、文字通りユーザビリティを示し、操作が簡単で、確実なナビゲーションとフィードバックがあること等である。

(2) Webデザインのためのルール

ここでは、実際のWebサイトデザインに必要な基本的ルールを記述した。項目のみ以下に紹介する。

- (a) ページレイアウトデザインに関するルール
- (b) ナビゲーションに関するルール
- (c) コンテンツデザインに関するルール
- (d) ユニバーサルデザイン対応

なお、上記ルールに基づき、実運用のガイドライン、サイトの登録・管理や運用に関する項目、具体的なデザインテンプレート等は、本社宣伝部のWebサイト事務局で作

成し、イントラネット上で公開している。

3.2.3 Webサイトユーザビリティの重要性

年に2回行われている日経BP社のWebブランド調査(2001年7月)結果によると、当社Webサイトが主要800サイト中ブランド指数の上昇度で2位にランクされた。また実アクセス数も伸びており、リニューアルの効果が現れている。

Webサイトは、とかく顔の見えにくい企業にとって、直接顧客との接点となり、情報発信のできる手段である。企業の方針や最新情報などがリアルタイムで発信され、また同時に企業にとっても顧客情報を収集したり商売をする上での有効手段ともなり得る。

ブロードバンド化の進展など環境の変化ともあいまってインターネット人口も増加している。当然ユーザー層も拡大し、あらゆる人がサイトを訪れることになる。ユーザーの満足を得るには、サイトのコンテンツの充実度は当然だが、使いやすいことすなわちユーザビリティを高めることは不可欠な条件と言える。また、常に新しい使い勝手を求めてインタフェース向上を図らなくては、ユーザーに見てもらえるサイトを維持し続けるのは難しい。そのため、今後も継続的にユーザビリティ評価を実施し、ユーザーの立場に立ったサイト構築を進めることが必要と考える。

4. む す び

4.1 評価手法の確立

インタフェースデザインの評価手法を確立し、様々な製品の開発に適用してきた。これらの事例では、主にインタフェースの物理的側面(感覚特性、生理特性、身体寸法などとの適合性)や認知的側面(思考特性、認知特性などとの適合性)の検証を行うことができた。この成果は、評価対象の製品やシステムのインタフェースデザイン改善に反映させることは当然ながら、さらに一般化して、Webサイトの場合、前に紹介したようにデザインガイドラインとして整備していくことも行っている。

4.2 今後の課題

上記4.1節に述べた物理的側面と認知的側面の更に上位にあるインタフェースの感性的側面(嗜好(しこう)特性、受容特性などとの適合性)の検証を方法論として試行し、手法の確立を図っていきたいと考える。

参 考 文 献

- (1) 三菱電機(株)デザイン研究所編：こんなデザインが使いやすいを生む、工業調査会(2001)
- (2) Jakob Nielsen著、篠原稔和監修、グエル訳：ウェブ・ユーザビリティ、エムディエヌコーポレーション(2000)

ヒューマンセンシング技術の開発と作業計測への応用

白松直樹* 大須賀美恵子***
坂口貴司*
平澤宏祐**

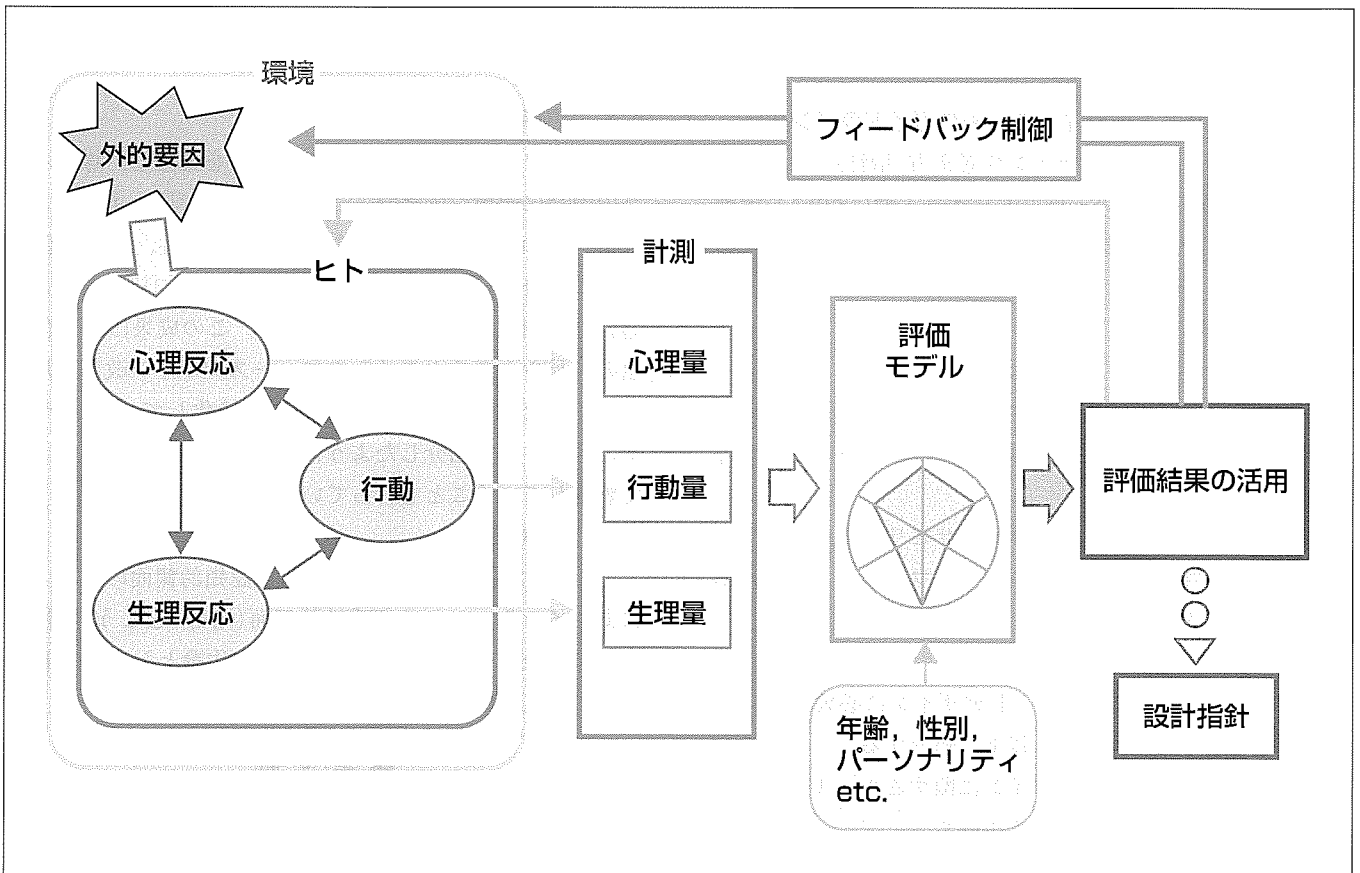
要 旨

機器・システムを操作している際の人間の状態を正しく把握することは、ヒューマンインタフェースの良い設計を行うためには極めて重要である。ヒューマンセンシング技術とは、人間の状態を定量的に表すための手法を総称したもので、行動量(動作, 発話, 視線, 表情など), 生理量(心電図, 心拍, 脈波, 血中成分など), 心理量(快・不快など)の計測・解析技術から構成される。

本稿では、上記のうち、行動量と生理量の計測・解析技術を工作機械を使用する作業行動場面における人間の状態把握に適用した事例について述べる。

行動量に関する事例としては、作業者の動作及び注視点

計測を行うための三次元位置計測システムの概要について述べる。このシステムは、無線式赤外線LED発光マーカを計測対象者の身体部位に装着し発光させ複数カメラのステレオ視によって三次元位置計測を行うもので、高精度かつ非拘束なりアルタイム計測を実現した。生理量に関する事例としては、工作機械を使用する際の作業者の生理データ計測とそれを基にした自律神経系指標について説明し、自律神経系指標の振る舞いから熟練者の作業行動解析が可能となることを述べる。なお、この研究は、産業経済省産業技術基盤研究開発プロジェクト「人間行動適合型生活環境創出システム技術」において実施した。



ヒューマンセンシング技術

人間は、何らかの外的要因によって心理的・生理的に反応し、これに追従して特定の行動を引き起こす。ヒューマンインタフェースの良い設計を行うためには、心理量、行動量、生理量を計測・解析して人間の状態を把握しなければならない。

出典：大須賀美恵子：ヒューマンセンシング技術のウェルネス産業への展開, BME, Vol.13, No.7, 20 (1999)

1. ま え が き

人間の状態を正しくとらえ定量化することは、機器・システムのヒューマンインタフェース向上に必要な技術である。人間の状態を表すための行動量(動作, 発話, 視線, 表情など), 生理量(心電図, 心拍, 脈波, 血中成分など), 心理量(快・不快など)を計測し解析する技術を総称してヒューマンセンシング技術と呼ぶ。

本稿では、産業経済省産業技術基盤研究開発プロジェクト「人間行動適合型生活環境創出システム技術」におけるヒューマンセンシング技術の適用について紹介する。このプロジェクトは、ものづくり作業行動場面でのユーザビリティ評価と、評価した結果のフィードバックによる作業支援システムの構築を目指している。ここでは、NC (Numerical Control) 工作機械を模擬するシミュレータを使った作業行動場面における作業者の動作及び注視点計測と、自律神経系指標評価の概要について述べる。

2. 三次元位置計測システムによる動作・注視点計測

NC工作機械を操作する際の動作と注視点の計測を目的に、三次元位置計測システムを開発した。当初は磁気センサによる計測システムを構築したが⁽¹⁾, 計測を行うために発生させる磁界環境の計測精度への影響や有線計測による拘束感などの問題点を持っていたので、画像計測による高精度かつ非拘束なりアルタイム三次元位置計測システムを開発した⁽²⁾。

2.1 システム構成

このシステムは、無線式赤外線LED発光マークを計測対象者の頭・肩・手などの身体部位に装着し発光させ(制御信号の受信器は腰に装着), 複数カメラのステレオ視によって三次元位置計測を行うシステムである。図1にこのシステムの計測部を示す。3個のマークはヘルメットに装着されている。位置計測用CCDカメラは4台である。通常はマークをすべて点灯させ、直前フレーム内の最寄りのマークID情報を利用してマークIDを特定する。マークIDが不明の場合は、マークを順次点灯させてマークIDを認識する方式に切り換える。PC(Personal Computer)では、入力した画像データから各マークの重心計算を行い、カメラローカル座標を各カメラごとに算出する。これを基にステレオ視を用いて各マークの三次元位置を計算する。3個以上のカメラで認識できている場合には、全ペアのカメラでステレオ視の結果を平均することによって計測精度を高めている。

このシステムを用いてマークの三次元位置を実測した結果、高さ180cmの位置に設置した一辺60cmの正三角形板の頂点に固定したマークの位置3点に対して0.8~1.5cm, NCパネル面上の5点において1.4~3.4cmの計測精度であ

った。

2.2 頭の三次元位置と注視点の計測

このシステムを利用して、頭(頭の回転位置中心)の三次元位置及びNC工作機械シミュレータ上の注視点を算出した。頭の三次元位置は、3個のマークの位置から重心及び頭の三次元姿勢を求め、事前に操作者ごとに計測した3マーク重心から頭の回転中心までの距離をオフセットすることによって求めた。注視点は、頭の回転位置中心を原点とした頭部方向ベクトルとNC工作機械シミュレータ面との交点とした。顔の向きと視線方向は必ずしも一致しないが、ビデオ視察実験の結果、ほぼ(約90%)一致することを確認した。

図2に、このシステムによって計測した頭部三次元位置と注視点位置の軌跡を示す。計測・計算速度は50~55Hzと十分な処理速度が得られた。注視点推定の精度はNCパネル面上で5~10cmであり、現在、眼球方向計測専用の小型カメラの導入で精度向上を図っている。

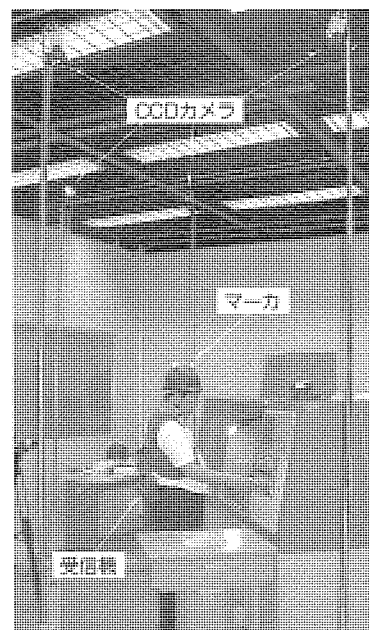


図1. 三次元位置計測システムの計測部

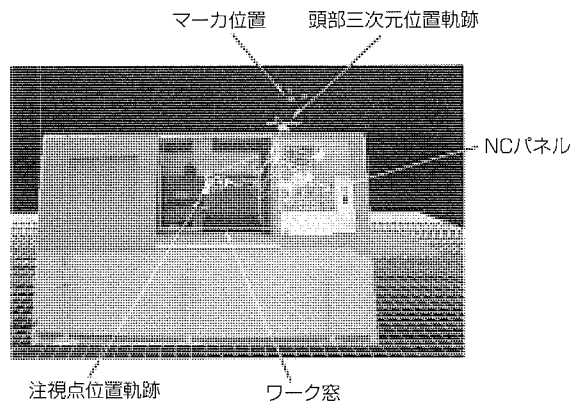


図2. 頭部三次元位置と注視点位置の軌跡

3. 自律神経系指標による作業行動解析

NC工作機械による作業行動場面における行動解析を行うことを目的に、熟練技能者と技能未習得者の生理データを計測し自律神経系指標を算出した。注意集中の度合いが高まっているときに技能が発揮されると考え、自律神経系指標から注意集中の度合いを抽出することを検討した⁽³⁾⁽⁴⁾。

3.1 自律神経系指標

生理データとして、①胸部誘導による心電図、②ベストに取り付けたピックアップによる呼吸、③近赤外I.F.Dとフォトランジスタが一体化した反射型のフォトセンサによる左耳垂部の光電式容積脈波を計測した。30秒間を1フレームとしてこれらの生理データから以下の自律神経系指標を求めた。

- (1) 平均心拍 HR(bpm)
- (2) 心拍変動RF成分 HR_RF
- (3) 心拍変動MF成分 HR_MF
- (4) 呼吸間隔 RESP(s)
- (5) 呼吸ピーク周波数 PF(Hz)
- (6) 呼吸重心周波数 GF(Hz)
- (7) 呼吸変動MF成分 EVOL_MF
- (8) 平均脈波伝達時間 PTT(ms)
- (9) 脈波伝達時間変動MF成分 PTT_MF

呼吸重心周波数は、呼吸ピーク周波数での振幅の1/2以上の振幅である周波数領域の面積を計算し、その面積を2分する周波数とした。(2)、(3)、(9)に示す変動成分は不整脈に対応する指標である。心拍変動RF成分は呼吸重心周

波数GFを中心とする $\pm 0.05\text{Hz}$ の範囲の平均振幅を計算したもので呼吸性不整脈に起因する成分であり、心拍変動MF成分 $0.0781\sim 0.1367\text{Hz}$ の範囲の平均振幅を計算したもので 0.1Hz 近辺のMayer Wave由来の不整脈に起因する成分である。PTT_MFも同様の考えに基づく。以上の指標を5秒間隔で算出してその時間変化を求めた。

3.2 指標の時間変化と作業内容との対比

技能者と技能未習得者に対し、直方体の素材の輪郭を削り指定された箇所に穴あけ加工を行う課題を課し、その際の生理データを計測した。5秒間隔で算出した各自律神経系指標時間変動を図3に示す。図は技能未習得者の一例である。図に挿入した平行線は、それぞれ各指標分布の75%点及び25%点を表している。

次に、これまでの知見による生理メカニズム⁽⁵⁾を基に、表1に示すような自律神経系指標の組合せによる生理反応パターンの分類を考え、各パターンに対応する時点を抽出した。表中の成立条件における“高”“中”“低”は、それぞれ、指標の値が分布の75%点以上、25%以上75%未満、25%未満の場合に対応する。

ここでは詳細については述べないが、一連の作業における個々の作業内容と生理反応パターンを対応付けると、熟練技能者と技能未習得者では結果が異なった。熟練技能者では、指を動かすのを止め問題紙を見て考え込んでいるときにパターンAが、実際の入力動作時にパターンBが多く検出された。これに対して技能未習得者では、A、Bのパターンが現れた場合の作業行動の分離はできなかった。

図4に熟練技能者、技能未習得者の呼吸重心周波数GF

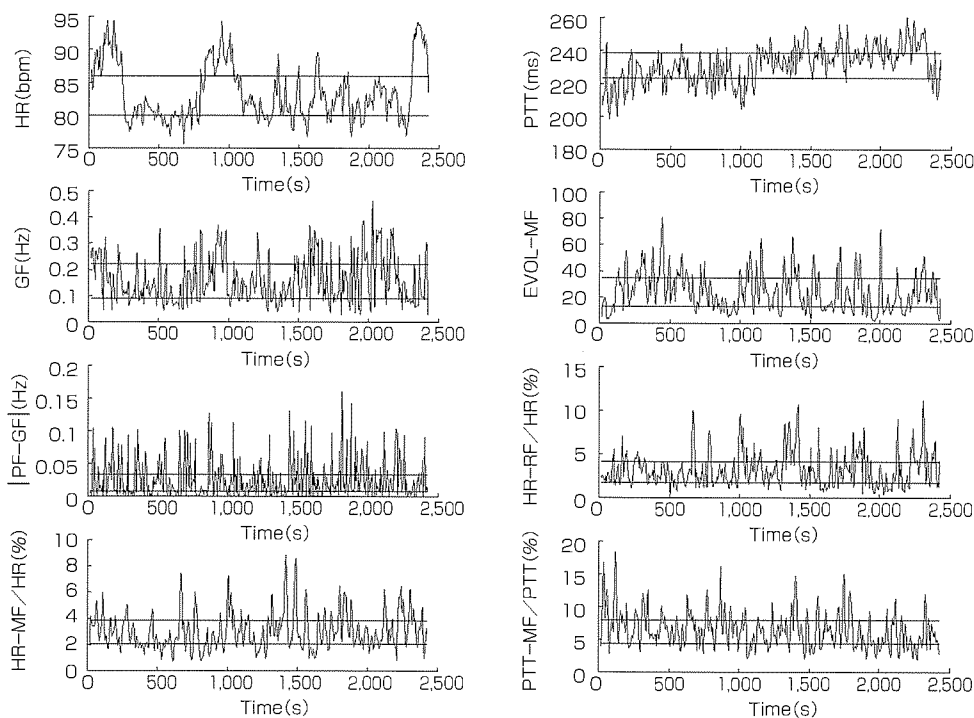


図3. 自律神経系指標の時間変動

表1. 生理反応パターンの分類

生理反応パターン	成立条件
A 全般的な緊張(精神作業負担)	高HR, 短PTT, 低HR _{RF} /HR, 高GF
B 交感系賦活	高又は中HR, 短PTT
C 圧受容体反射低下(外的環境への注意集中)	低HR _{RF} /HR, 低HR _{MF} /HR, 低又は中GF
D 単調ストレス(倦怠(けんたい)又は発話の影響)	高PTT _{MF} /PTT, 高呼吸MF, 高 PF-GF

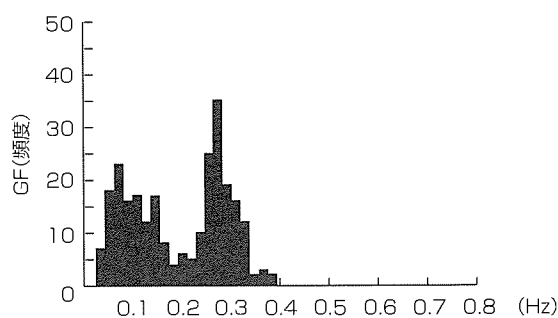
の分布をヒストグラムで示す。この図は一例であるが、技能者では二峰性を示し、技能未習得者ではピークが一つだけ現れる傾向を示した。呼吸関連指標以外の他の指標に関しては、熟練技能者と技能未習得者の間でこのような特徴的な差異は現れなかった。熟練技能者においては、呼吸が速い時と遅い時の変化のメリハリがあり、技能未習得者においては、そのメリハリがない。このことから、技能者では作業中に注意集中の度合いを臨機応変に変化させ、技能を集中的に発揮している可能性がある。これらの結果により、生理反応パターンと作業内容をより詳細に対応させる可能性が示唆された。

我々は、これまで、作業者のストレスを検出し、作業内容とストレスの関係を解明するための手法を検討してきた⁽⁶⁾。今後は、ストレス検出時点を注意集中の度合いが高まっている時点に置き換え、作業行動における技能の発見に取り組む。

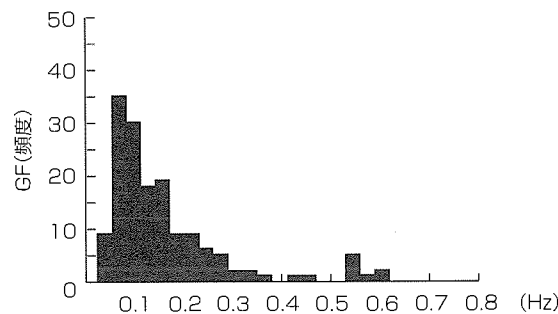
4. む す び

ヒューマンセンシング技術の適用事例として、NC工作機械を模擬するシミュレータを使った作業行動場面における作業者の動作及び注視点計測と、自律神経系指標評価について述べた。今後は、これらの技術を実作業現場に適用するための開発を進め、技能を十分に発揮するための生産環境や設備機器の改善提案を行っていく所存である。

この研究は産業経済省・産業総合技術研究所の産業技術基盤研究開発プロジェクト「人間行動適合型生活環境創出システム技術」として、新エネルギー・産業技術総合開発機構(NEDO)からの委託を受けて、(社)人間生活工学研究センター(HQL)を通して実施したものである。協力してプロジェクトを遂行している関係各位に感謝する。



(a) 熟練技能者



(b) 技能未習得者

図4. 呼吸重心周波数の分布

参 考 文 献

- (1) 坂口貴司, ほか: 磁気センサを用いた作業者の動作計測システム, 情報処理学会第61回全国大会抄録集(2000)
- (2) 坂口貴司, ほか: 非拘束マーカ式3次元位置計測システム, 情報処理学会第63回全国大会抄録集(2001)
- (3) 平澤宏祐, ほか: 自律神経系指標を主に用いた作業行動解析, 第15回生体・生理工学シンポジウム論文集(2000)
- (4) 平澤宏祐, ほか: 自律神経系指標を主に用いた作業行動解析—第二報—, 第16回生体・生理工学シンポジウム論文集(2001)
- (5) 大須賀美恵子, ほか: 心臓血管系モデルを用いた自律神経指標の解釈, BME, 11, No.1 (1997)
- (6) 達野陽子, ほか: 自律神経系生理指標を用いた問題発見型作業評価手法の開発と協調作業への適用, ヒューマンインタフェース学会研究報告集, 1, No.1 (1998)

産業用ヒューマンインタフェース 評価システム

坂上聡子*
寺下尚孝*

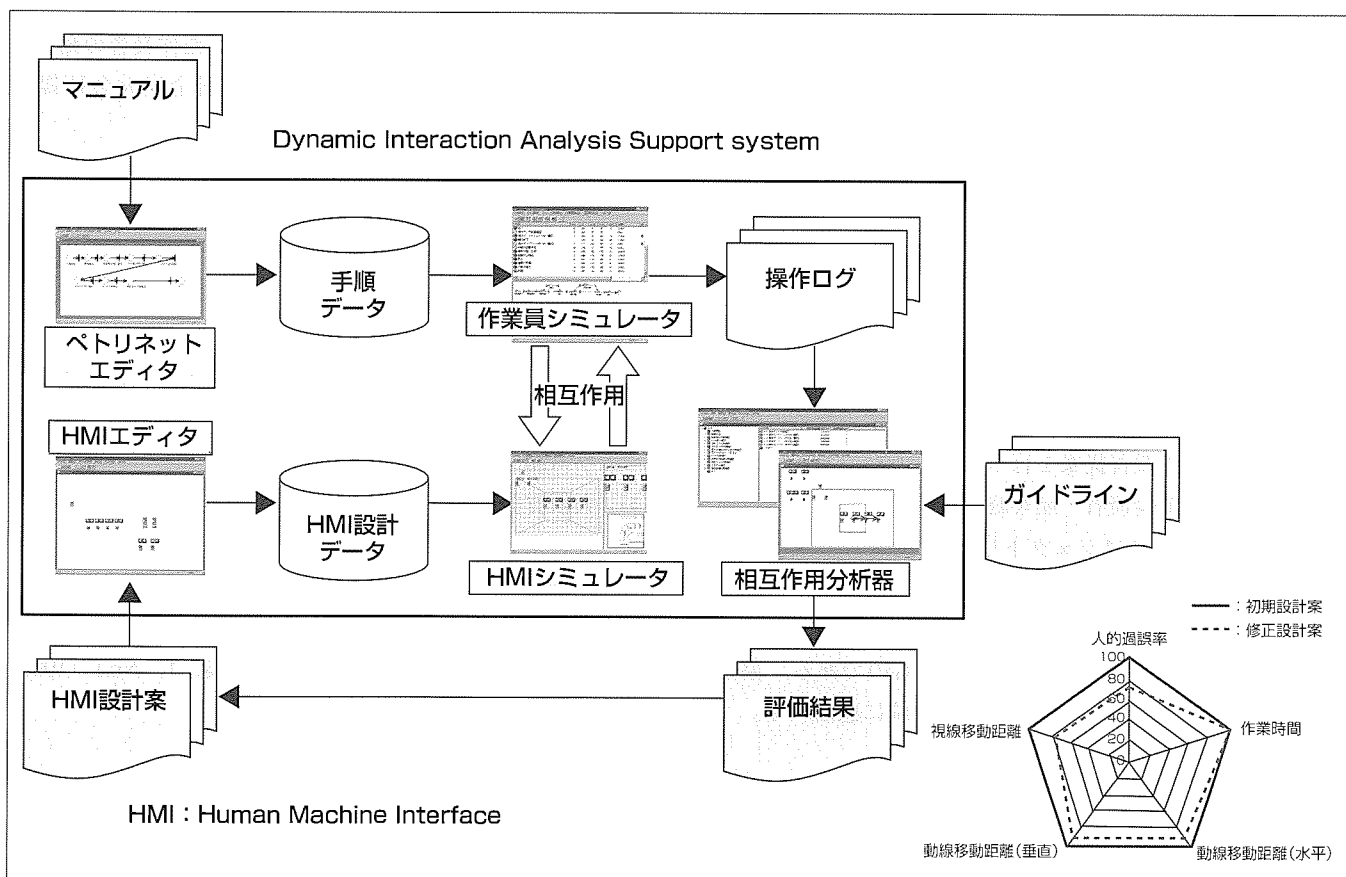
要旨

高い信頼性と安全性を持つ大規模システムは、近年の計算機技術、知的情報処理技術などの進歩により、高度に自動化されてきた。しかし、それを操作する人間の身体・認知特性を考慮しない自動化は、かえって信頼性を低下させることになる。したがって、ハードウェアの信頼性だけでなく、人間を含めたシステム全体の信頼性の向上に取り組む必要がある。

一般に、ヒューマンインタフェースを設計する場合、製品のプロトタイプを製作し、評価を行い、その結果をフィードバックすることを繰り返し、設計案を洗練化するというIterative Design手法が重要であると指摘されている。従来、産業用ヒューマンインタフェースは、実物大のモックアップを製作して、実際の作業員による被験者実験を実施し、各種評価指標を導出する手法や、人的過誤率のデー

タベースを用いて作業を誤る確率を机上評価する手法による検証が行われてきた。しかし、それらの手法で繰り返し評価を実施するためには、多大な時間的・金銭的コストが必要となる。著者らは、作業員のモデルに基づく計算機シミュレーションを行い、作業員と設備との相互作用を観察・記録し、その結果を分析して、定量的にヒューマンインタフェース評価を行うシステムを開発し、実際の産業用プラントの設計検証を実施してきた。このシステムを用いることにより、短期間で総合的なヒューマンインタフェース評価を実施できるため、効率的に信頼性の高い設計を実現することができる。

本稿では、作業員のモデル化の方法、ヒューマンインタフェース評価方法と、実際の産業用プラントの制御盤インタフェースを評価した例について述べる。



ヒューマンインタフェース評価システム(DIAS)

DIASは、二つのエディタ、二つのシミュレータと相互作用分析器からなるシステムである。評価者は、マニュアルと設計案を基に、ペトリネットエディタとHMIエディタを用いて手順データとHMI設計データを作成する。それらのデータを用いて、作業員の行動を模擬する作業員シミュレータと制御盤の機器の配置や状態を模擬するHMIシミュレータとが相互作用を行うことにより、制御室内で作業員が作業を行う様子を再現する。相互作用分析器は、相互作用を記録した操作ログを読み込んで分析することにより、ヒューマンインタフェースの良さ悪しを評価する。

1. ま え が き

人間にとって操作を誤りにくく使いやすいインタフェースを設計するためには、製品のプロトタイプを製作し、評価を行い、その結果をフィードバックするというプロセスを何度も繰り返し、設計案を洗練化するというIterative Design手法が重要である。また、原子力プラントを始めとする大規模システムにおいては、監視・操作の対象となるヒューマンインタフェースに対して高い信頼性が要求とされるが、設計の妥当性を効率的に検証する手法が確立されていない。従来、このような産業用ヒューマンインタフェースは、実物大のモックアップを製作して実際の作業員による被験者実験を実施し各種評価指標を導出する手法や、人間信頼性評価手法の一つであるTHERP(Technique for Human Error Rate Prediction)⁽¹⁾を用いて作業を誤る確率を机上評価する手法による検証が行われてきた。しかし、それらの手法で繰り返し評価を実施するためには多大な時間と費用が必要であるため、現実的には困難である。

我々は、計算機シミュレーションを用いて、産業用ヒューマンインタフェースの評価を行うシステム(Dynamic Interaction Analysis Support system : DIAS)⁽²⁾を構築した。

ヒューマンインタフェースの良し悪しは、誤りにくさ、タスク実行時間の短さ、作業員の身体的負担など様々な独立した指標によって評価される。DIASは、複数の評価指標を効率的に導出することができ、複数の設計案で同一タスクを実行するパフォーマンスを相対比較(ベンチマークテスト)したり、各指標間のトレードオフを把握したりすることができる。

本稿では、DIASの機能と評価方法について述べる。

2. DIAS

DIASは、作業員が手順書に従ってタスクを実行する際の作業員と設備の相互作用を再現したシミュレーション結果に基づいてヒューマンエラー率や作業員の肉体的作業負荷などに関する定量評価を行うシステムである。DIASの構成を図1に示す。

2.1 分散シミュレーションシステム

分散シミュレーションシステムは、作業員シミュレータとヒューマンマシンインタフェース(HMI)シミュレータという二つのシミュレータで構成される。

2.1.1 作業員シミュレータ

作業員シミュレータ⁽³⁾は、人間の認知・記憶特性⁽⁴⁾と視野領域(焦点視野領域、周辺視野領域)、歩行速度、操作速度などの身体的特性を考慮しており、手順データに基づいてタスクを実行する作業員を模擬する。手順データは、ペトリネットモデルで記述されており、Petri-Net Editor

を用いて作成・編集を行うことができる。ペトリネットモデルを用いることにより、作業手順や作業員の行動の状態推移を容易に理解することができる。

作業員シミュレータの特性を以下に示す。

(1) 視野角

人間工学の知見⁽⁵⁾⁽⁶⁾により、作業員の視野角を図2に示す。シミュレーションでは、作業中において操作対象機器が最良な視点である図中の点Fになるように移動を行う。また、操作対象機器の検索時間は、作業員の視野と機器の関係によって変えている。

(2) 視野の到達する距離

視野角に入り、かつ作業員から8m以内の対象物を認識する。

(3) 移動速度

一般的な歩行速度から、作業員は毎秒1mで移動を行う。

(4) 作業経路

作業員は、盤面又は盤などの障害物を回避し、最短距離を移動する。ただし、身体の正中線が壁面から30cm以内には接近しない。

(5) 作業の間合い

操作中の作業員と操作対象機器との距離を50cmとする。

(6) 目の高さ

身長170cmの作業員を想定し、目の高さは160cmとする。

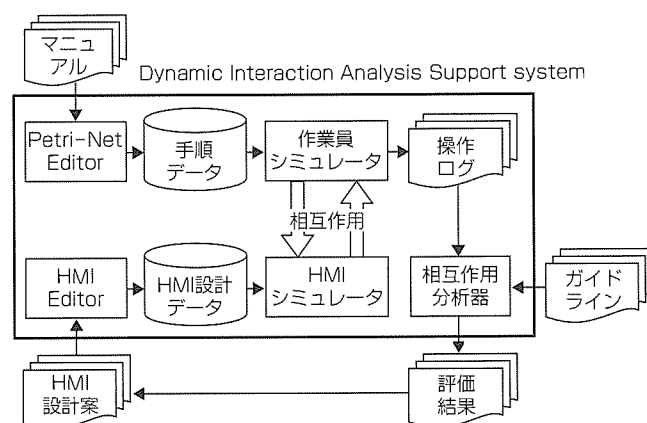


図1. DIASシステムの構成

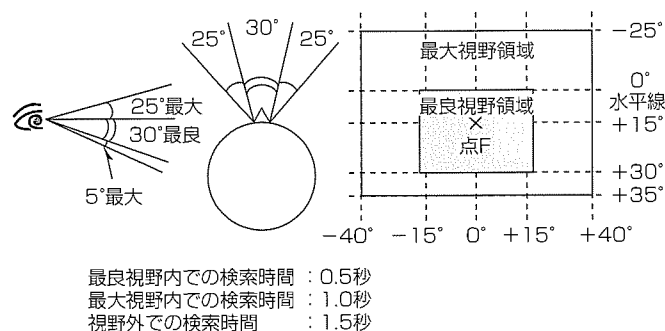


図2. 作業員の視野領域

(7) 認知特性

作業員の認知記憶量に制限を設けない。したがって、作業中に手順を忘れるというエラーは発生しない。

2.1.2 HMIシミュレータ

HMIシミュレータは、作業員シミュレータに対し、HMI設計データに基づいて、各種制御盤の室内配置と、盤面上の機器配置、機器の状態(指示計の値、スイッチのON/OFF)などの情報を提供する。HMI設計データはHMI Editorによって編集することができ、評価者はアイコンで表された装置をクリック、ドラッグすることで、機器の情報を入力したり、機器の配置を変更したりすることができる。

2.2 相互作用分析器

分散シミュレーションシステムは、シミュレータ間の相互作用を時系列に記録して、操作ログとして出力する。操作ログには、実行したタスクの内容、作業員の位置、視点、機器の状態などが記録されている。評価者は、相互作用分析器を用いて、その結果を分析する。

2.2.1 評価方法

図3に結果の分析画面の例を示す。DIASは手順データをペトリネットで表現しており、相互作用分析器ではその階層性をツリー構造で表しているため、評価者は手順の階層構造を理解しやすい。したがって、どの階層における作業員の移動距離が大きいかというような同一レベルの階層同士の比較や、どの機器を操作するときの実行時間が長いかなというような階層ごとの詳細な評価が容易である。

相互作用分析器は、幾つかの定量的指標を計算する。指標については2.2.2項を参照のこと。また、評価者の判断が必要な指標を評価する場合、相互作用分析器は、何らかの方法で評価支援を行う。例えば、THERPによってヒューマンエラー率(Human Error Probability: HEP)を推定する場合、相互作用分析器は、図3に示すように、機器の配置や作業員の視野や機器を操作した順序などを表示して

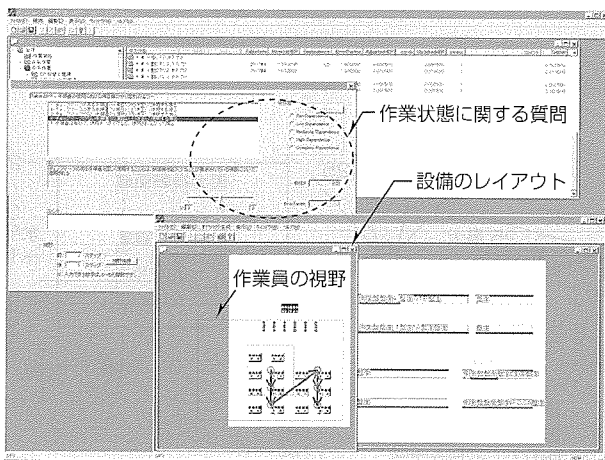


図3. シミュレーション結果の分析画面

詳細な作業状況を示すことにより、評価者が適切なHEPを導出する支援を行う。これらの評価支援には評価が主観的になるのを防ぐ効果もある。

評価者は、得られた評価指標を基に、複数の設計案を相対的に評価したり、どの部分のインタフェースを修正すべきであるか、修正するとどの程度改善効果があるかを定量的に分析したりすることができる。このように、DIASによる評価は、複数の指標について同時に評価することができるため、指標間のトレードオフを発見することができ、設計案や手順の総合的な改善が可能である。

2.2.2 評価指標

相互作用分析器が出力する評価指標は、以下のとおりである。

(1) 動線移動距離(水平)

作業実施に伴う作業員の移動距離(歩行距離)の総計

(2) 動線移動距離(垂直)

作業員が立ったり座ったりした距離の総計

(3) 視線移動距離

作業員が作業対象を探すために盤面上で視線を移動した距離の総計

(4) 作業時間

作業開始から終了までの時間の総計

(5) 斜め視線移動距離

作業員は作業対象に正対して作業すると考えられるが、直立姿勢で目の高さより高い位置、又はしゃがんだ状態で目の高さより低い位置に作業対象を斜めから操作・注視する。このときの見上げた距離、見下ろした距離の総計

3. 評価例

3.1 初期設計案

DIASにより、実際の産業用プラントの制御盤の部分更新工事において、ヒューマンインタフェース評価を行った⁽⁷⁾。初期設計案(DIASの表示画面)を図4に示す。作業員の視線移動軌跡が矢印で示されている。

相互作用分析器によって分析を行った結果、初期設計案においてヒューマンエラー率などの指標値が大きい(悪い)原因として、次の2点が挙げられることが判明した。

(1) 操作機器の位置が非常に低く、無理な体勢で作業することにより、機器のネームプレートを見誤るなど、正確な作業ができない危険性がある。

(2) 機器の配置と作業手順の整合がとれていないために、作業対象操作器を誤る危険性がある。

3.2 修正設計案

3.1節で指摘した設計案の問題点を考慮し、修正設計案を作成した。修正設計案を図5に示す。

修正設計案に対しても、DIASを適用し、インタフェー

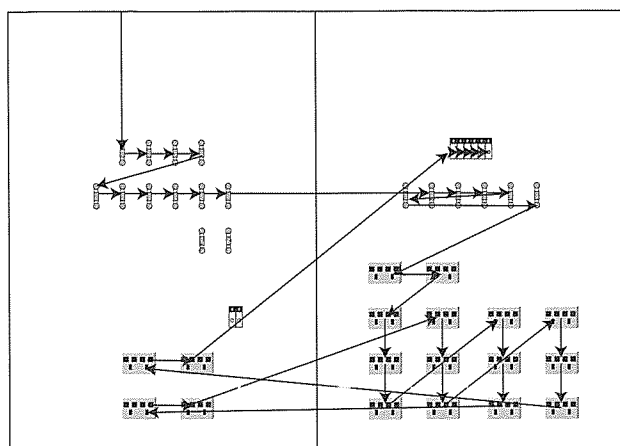


図4. 初期設計案

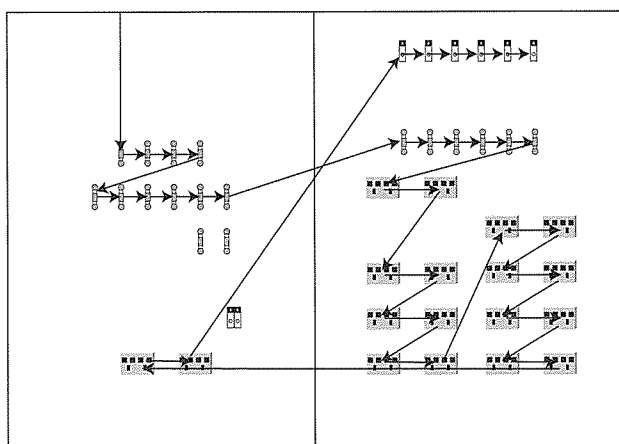


図5. 修正設計案

ス評価を実施した。修正設計案は、次のような設計方針に基づいて設計されている。

- (1) 操作器は上から下へ、左から右へ操作する。
- (2) 機器を関連性によって幾つかのグループに分類し、グループごとに分離して配置する。

3.3 設計案の比較

初期設計案と修正設計案の評価を比較した結果を図6に示す。図は初期設計案の評価結果を100としたときの相対比較評価で、数値が小さくなるほど改善率が高いことを意味する。

図により、修正設計案は、作業時間を除いた指標に関して、初期設計案よりも約20~40%改善されている。また、修正設計案は、どの指標においても、初期設計案よりも悪化していない。したがって、修正設計案における設計変更の有効性が検証された。

4. むすび

本稿では、計算機シミュレーションに基づく産業用ヒュ

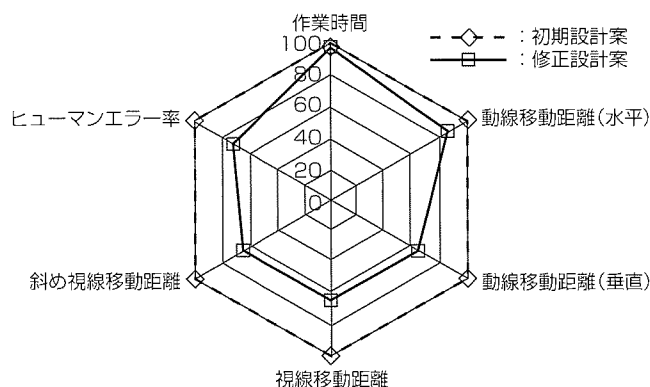


図6. 設計案間の評価結果比較

ーマンインタフェース評価システム(DIAS)を提案し、作業員のモデル化とDIASの評価方法について述べた。DIASは、計算機シミュレーションを適用し作業員と設備の相互作用を模擬することにより、ヒューマンエラー率、肉体的作業負荷、認知負荷などの指標を容易に定量的に評価することができる。したがって、被験者実験のように多くのコストを必要とせず、産業用ヒューマンインタフェース設計を容易に評価を実施できるため、繰り返し適用することで、効率的で高品質なヒューマンインタフェースの設計が可能になる。

参考文献

- (1) Swain, A. D., et al.: Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications, NUREG/CR-1278 (1983)
- (2) 中川隆志, ほか: 計算機シミュレーションを用いた保守設備インタフェースの評価, 計測自動制御学会論文集, **34**, No.12 (1998)
- (3) Nakagawa, T., et al.: Simulation Based Interface Evaluation of Man-Machine Interface Design in Power Plants, Cognitive Systems Engineering in Process Control(CSEPC 96), 332~338 (1996)
- (4) Reason, J.: Human Error, Cambridge University Press (1990)
- (5) U. S Department of Defense: Human Engineering Design Criteria for Military Systems, Equipment and Facilities, Military Standard 1472 C. (1981)
- (6) Sanders, A. F.: Some Aspects of the Selective Process in the Functional Visual Field, *47(6)*, 101~117 (1970)
- (7) Matsuo, S., et al.: Simulation-Based Human Interface Evaluation for Maintenance Facilities, Proc. of Human-Computer Interaction(HCI Int'l '99), **1**, 948~952 (1999)

現在、エアコンの買換え需要は総購入の半分以上を占めています。この買換え動機として、使い続けたことによる能力低下及びエアコン内部の汚れによる要因が非常に高くなっております。

以上のようなユーザーニーズや社会環境により、初期性能の改善はもちろん、長寿命化という観点を考慮して、ユーザーニーズの高い空気清浄機能とエアコン内部を汚さないことを両立した全く新しい空気清浄“イオンプラズマシールド”を開発し、吹き出す空気の清潔さとライフエンドまで初期性能を維持可能な省エネルギー性を実現しました。

さらに、人の快適要素である室温・湿度に加え床温度を検知する“快測センサー”を搭載し、快適性と省エネルギー性の両立を実現しました。

特長

1. 業界初！エアコン内部を汚さないイオンプラズマシールドにより、ライフエンドまでの省エネルギー性を改善

“イオンプラズマシールド”とは、強力な帯電力を持つマイクロ繊維によって細かい空気の汚れをキャッチする“大型イオン空清フィルター”と、プラズマ電極部と活性炭フィルターによってにおいや有害ガスを脱臭・分解する“イオンプラズマユニット”の組合せからなり、従来にない空気清浄性能を発揮できるようになりました。空気清浄フィルターで吸い込み口を広く覆うことは、目詰りによる能力の低下、電気代の悪化を招くため、従来難しいとされてきました。しかし、空気は通しながら汚れはイオンの力によって吸着し除去させる新フィルター“大型イオン空清フィルター”の開発と、大口径ファンの採用と風路設計により、WXシリーズでは、エアフィルターに加え、熱交換器の97%をカバーする“イオンプラズマシールド”で微細なほこりや煙も除去し、熱交換器・送

風ファン・風路への汚れの付着を抑制し、清潔に保ちます。

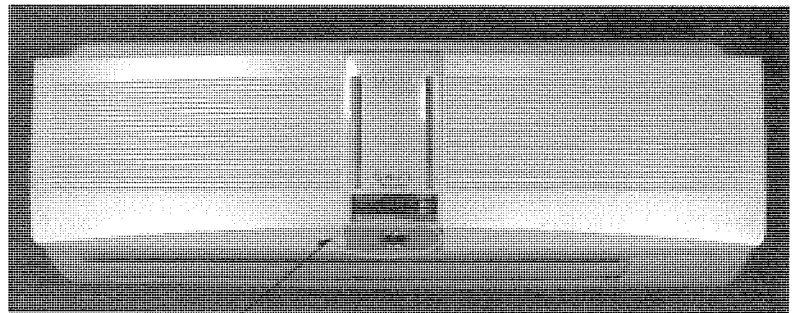
このエアコン内部を汚さない機能により、フィルターの汚れによる電気代の悪化を約15%抑制し、さらに、送風ファンで約25%、熱交換器で約5%とトータルでは約45%の省エネルギー性を改善することができます。

2. 快適性要素を測る独自のセンサ群搭載により、理想の快適空間を創造する“快測センサー霧ヶ峰”誕生

従来のエアコンは“足元が暖まらない”というエアコン暖房への不満を招いていました。そこで“快測センサー霧ヶ峰”WXシリーズでは、床面全体の温度を測ることのできる床温度センサを室内機に搭載することで、床面が冷えていれば強力に足元を暖め、暖まってきたら自動的に省エネルギー運転を行います。室温・湿度・床温度を測って運転制御をするため、無駄な暖め過ぎ・冷やし過ぎを抑え、約20%の省エネルギー効果を発揮します。

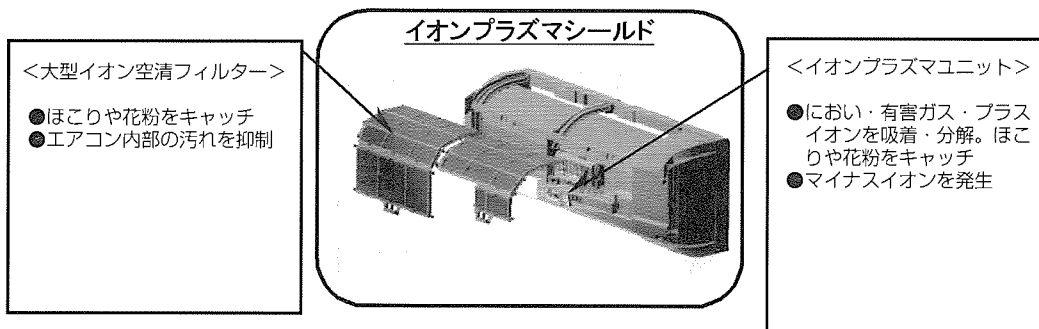
3. 主要部品の省エネルギー性向上による初期性能を改善

当社独自技術である圧縮機の連結コア型集中巻きモータと駆動回路の効率向上、熱交換器の切り起こし形状の最適化及び伝熱面積約5%の拡大など、主要部品の効率向上を図りました。これにより、省エネルギー法の規制値である冷暖平均COPの達成率は115%を実現しました。



床温度センサ

WXシリーズ外観と床温度センサ



イオンプラズマシールドの外観

住 所：〒100-8310 東京都千代田区丸の内2-2-3 (三菱電機ビル)

会社名：三菱電機株式会社 家電・デジタルメディア事業部 お問い合わせ先：販売グループ TEL 03-3218-9156



特許と新案***

三菱電機は全ての特許及び新案を有償開放しております

有償開放についてのお問合せは
三菱電機株式会社 知的財産渉外部
電話(03)3218-9192(ダイヤルイン)

対話処理装置 (特許 第3212618号, 国際公開WO98/13771号)

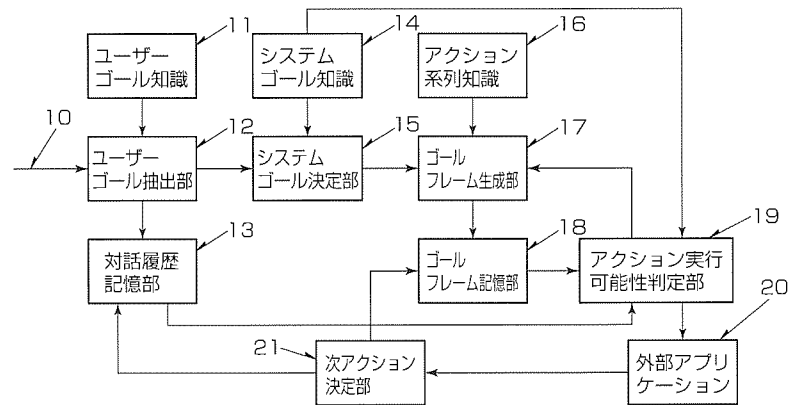
発明者 渡邊圭輔, 永井明人

この発明は、自然言語によるマンマシンインタフェースに用いられる対話処理装置に関するものである。

従来の対話処理装置は、対話の主導権が常に装置にあり、ユーザーはあらかじめ規定された対話手順によってなされる装置からの質問にしか答えられず、非常に利便性が低かった。

この発明は、装置とユーザーの双方が主導権を持てる対話の実現を目的としている。この発明をホテル検索・予約アプリケーションに適用した場合について図に基づいて説明する。まず、ユーザーが主導権を持ち“鎌倉ホテル明日予約してください”を入力(10)した場合、ユーザーゴール抽出部(12)がユーザーゴール“予約”を抽出し、システムゴール決定部(15)は“予約”をシステムゴールとする。次に、ゴールフレーム生成部(17)は、“予約”を達成するための外部アプリケーション(20)へのアクション(動作)“空室チェック”“予約実行”をゴールフレームとして生成する。アクション実行可能性判定部(19)は“空室チェック”に必須(ひつ)のパラメータ“客室タイプ”の値が対話

履歴記憶部(13)には存在しないため実行不可能と判定し、システムゴール知識(14)を参照して“客室タイプ取得”を新たなシステムゴールとする。これにより、装置が対話の主導権を持ち、アプリケーションに“客室タイプはいかがいたしますか”を出力させる。ユーザーはこの質問に答えることで当初の希望を達成することができる。以上のように、システムゴールをユーザーの入力と装置の内部状態から生成することにより、主導権の制御を適切に行うことができる。



表を含む帳票処理装置 (特許 第3179280号, 特開平7-282193号)

発明者 平野 敬, 岡田康裕

この発明は、帳票に記入された文字を読み取る文字認識装置において、帳票画像上の表から文字認識の対象とする領域(フィールド)の位置を抽出するフィールド抽出方式に関するものである。

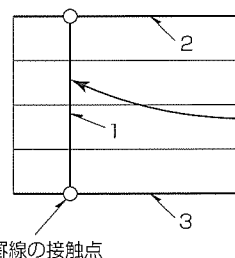
従来のフィールド抽出方式は、スキャナを用いて入力した品質の良い帳票画像から固定位置の領域を抽出するものであり、FAX送信等で傾きや拡大縮小、局所的な変形が生じた低品質な帳票画像には対応できない課題があった。

この発明では、図の実施例に示すように、帳票画像から抽出した罫線(けいせん)と表モデルが持つ罫線との対応付けを行うことで、特定の罫線に囲まれた帳票画像上の領域をフィールドとして抽出する。この罫線同士の対応付けは、罫線の位置や長さの情報に加えて、縦罫線(1)の両端が横罫線(2)と(3)に接触しているという罫線間の論理的な接続関係の情報から行う。

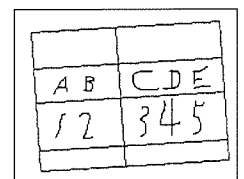
この発明によれば、傾きや拡大縮小、局所的な変形で罫線の

位置や長さ変動した場合でも罫線間の論理的な接続関係に基づいて正しく罫線に対応付けることができ、結果としてFAX送信された低品質な帳票画像からも正確なフィールド位置の抽出が可能となる。

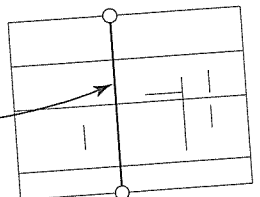
表モデル



帳票画像



帳票画像から抽出した罫線



罫線の対応付け

罫線の接触点



特許と新案 * * *

三菱電機は全ての特許及び新案を有償開放しております

有償開放についてのお問合せは
三菱電機株式会社 知的財産渉外部
電話(03)3218-9192(ダイヤルイン)

位置決め制御装置および位置決め制御方法 (特許 第2861277号, 特開平4-38504号)

発明者 上田隆美, 水野滋基, 山本 哲

この発明は、位置決め制御装置に関するものである。

従来の位置決め制御装置は、まず位置決め機構の機械共振近傍の周波数成分を抑えた駆動信号による速度制御を行う。次に、減速基準速度プロフィールに追従させる速度追従制御を行う。さらに、目標位置近傍では、位置追従制御を行う。従来の装置は、各制御の切換え時に、信号の不連続性によって過渡応答が生じ、位置決め時間が長くなるなどの欠点があった。

この発明は、上記問題点を解決するためになされた。図

1に示すように、フィードフォワード補償部(13)は、フィードバック補償部(12)の逆特性を持った第1フィルタ手段(3)と、制御対象をモデル化した第2フィルタ手段(4)とを直列に接続した。指令信号発生手段は制御対象の特性に合わせて調整した指令信号(6)をフィードフォワード補償部(13)に与えるものであって、当該調整が、図2に示すように、制御対象の共振周波数に対して生成した指令信号のスペクトル分布の谷部が一致するように、波形形状パラメータを操作した。

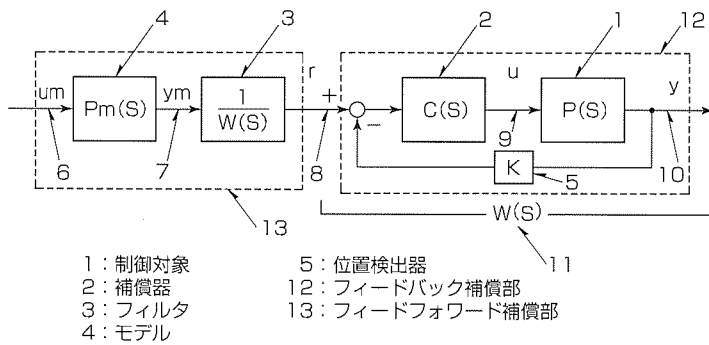


図1

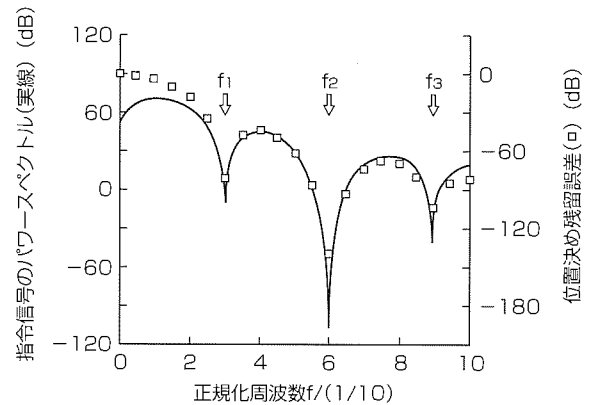


図2

〈本号記載の商標について〉

本号に記載されている会社名、製品名はそれぞれの会社の商標又は登録商標である。

〈次号予定〉三菱電機技報 Vol.76 No.9 「インターネット時代の社会インフラシステム」特集

三菱電機技報編集委員	三菱電機技報 76巻8号	2002年8月22日 印刷
委員長 井手 清	(無断転載・複製を禁ず)	2002年8月25日 発行
委員 高橋 大 畑谷正雄 堤 清英	編集人 井手 清	
乗原幸志 村松 洋 松本 修	発行人 福本 紀久男	
浜 敬三 石野 慎将 中川博雅	発行所 三菱電機エンジニアリング株式会社 e-ソリューション&サービス事業部	
中島克人 部谷文伸	〒105-0011 東京都港区芝公園二丁目4番1号	
黒畑幸雄 山木比呂志	秀和芝パークビルA館9階 電話 (03)3437局2692	
事務局 松本敬之	印刷所 株式会社 三菱電機ドキュメンテクス	
本号取りまとめ委員 依田 文夫	発売元 株式会社 オーム社	
	〒101-0054 東京都千代田区神田錦町三丁目1番地	
	電話 (03)3233局0641	
	定 価 1部735円(本体700円)送料別	
URL http://www.MitsubishiElectric.co.jp/giho/	三菱電機技報に関するお問い合わせ先	cep.giho@ml.hq.melco.co.jp

〈訂正〉

第76巻7号(2002年7月)64ページ記載の「本号取りまとめ委員」を右記のとおり訂正いたします。(正)塚田光政 ← (誤)依田文夫

スポットライト

三菱FAXOCRシステム “MELFOS”

三菱FAXOCRシステム“MELFOS”は、ファクシミリで送られた帳票の手書き文字を認識し、コンピュータのデータに変換する製品です。様々な業務をスピーディに処理し、夜間や休日に関係なく24時間自動受付を実現します。お客様の受発注業務、集計業務、報告業務、各種管理業務などの大幅な効率化・省力化が可能です。

三菱文字認識エンジンの特長

(1) ノイズ除去技術

文字を欠損させずにFAX特有の縦線ノイズ・ゴマシオノイズを除去します。このためノイズに強くなり、FAXで受信した帳票に文字を書いてFAX返信しても、その帳票に書かれた文字は認識されます。

(2) 特徴点抽出による高度なフィールド認識技術

ノイズを除去した帳票イメージから特徴点(帳票の交点等)を抽出し、その並びを比較することで、帳票の種類と認識フィールドの位置を特定します。これにより、FAX特有のゆがみ・かすれに強くなります。

(3) 文字の変形や、切れ・つぶれに強い文字認識

文字の変形に強い輪郭解析法と切れ・つぶれに強い特徴マッチング法によるハイブリッドな文字認識です。

(4) 各種罫線に対応

従来の罫線(けいせん)だけでなく、升目を離れたはしご枠罫線、点線区切り、半線区切り、イメージ枠にも対応(予定)でき、従来型の、より記入しやすい帳票をデザインできます。

三菱FAXOCR MELFOS製品の特長

(1) 高度な文字認識

上記の文字認識技術をソフトウェアで実現し、業界最高水準の手書き文字認識率を誇ります。手入力に比べ、データエントリー時間は激減し、入力ミスも少なくなります。

(2) 低ランニングコスト

OCR帳票は市販OAソフトウェアで作成できます。メールやWWWによる帳票配信が可能なので、帳票作成や印刷・配布コストを低減できます。記入した帳票は今までどおりFAXで送るだけで、特別な教育は不要ですので教育コストがかかりません。

(3) 業務のペーパーレス化/効率化

受信イメージは変換した文字データとともに電子ファイルで処理・保存します。紙へは出力しませんので、紙媒体保管場所は不要です。ISO14001を推進するオフィスに最適です。FAX内容の問い合わせに対しても簡単に検索・

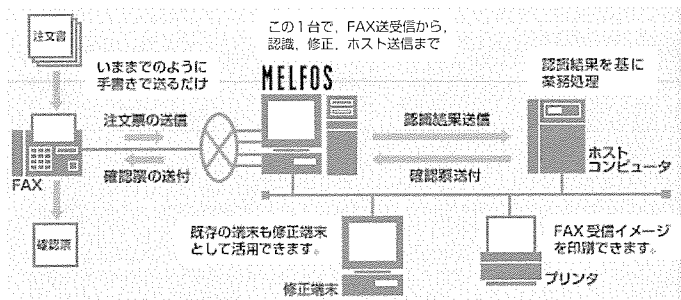
回答できますので、業務の効率化を図ることができます。

(4) 柔軟なシステム構成、大規模システムにも対応

回線数、修正端末数、スキヤナ接続等、お客様の業務に合わせた最適なシステムを構築できます。マルチサーバによって300回線以上のシステムも構築できます。

(5) 短納期

ホスト接続、帳票定義など、使い勝手の良いシステム構築・運用支援ツールを標準で備えており、短期間にシステム構築・導入が可能です。



システムの構成

帳票例

三菱FAXOCR MELFOSの認識仕様

	項目	機能	
OCR帳票	用紙サイズ	B4縦横/A4縦横/B5横/A5横	
	紙質	普通紙	
	読み取り字種	JIS OCR手書き文字(ANKS)、マーク、手書き漢字(JIS第一水準、3,238文字) ※ANKS[Alphabet[英大文字]]Number[数字] Kana[カナ]Symbol[記号]	
	読み取り文字枠サイズ	最小 横5×縦6mm 最大 横12×縦12mm	
ユーティリティ	帳票フォーマット定義	最大999種類	
	知識辞書定義	登録辞書数	最大255辞書/システム
		登録単語数	最大10万語/辞書
		登録単語長	最大30文字/語
認識文字種定義	ANKS	15種類	
	漢字	15種類	