

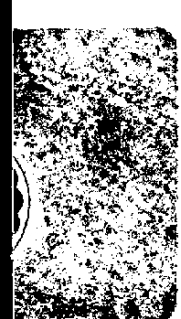
データベース構築促進及び技術開発に関する報告書

メディア変換型感性データベースの構築方法の研究

平成7年3月

財団法人 データベース振興センター

財団法人 イメージ情報科学研究所





この事業は、競輪の補助金を受けて実施したものです。

序

データベースは、わが国の情報化の進展上、重要な役割を果たすものと期待されている。今後、データベースの普及により、わが国において健全な高度情報化社会の形成が期待される。さらに海外に対して提供可能なデータベースの整備は、国際的な情報化への貢献および自由な情報流通の確保の観点からも必要である。しかしながら、現在わが国で流通しているデータベースの中でわが国独自のものは1/3にすぎないのが現状であり、わが国データベースサービスひいてはバランスある情報産業の健全な発展を図るためには、わが国独自のデータベースの構築およびデータベース関連技術の研究開発を強力に促進し、データベースの拡充を図る必要がある。

このような要請に応えるため、(財)データベース振興センターでは日本自転車振興会から機械工業振興資金の交付を受けて、データベースの構築および技術開発について民間企業、団体等に対して委託事業を実施している。委託事業の内容は、社会的、経済的、国際的に重要で、また地域および産業の発展の促進に寄与すると考えられているデータベースの構築とデータベース作成の効率化、流通の促進、利用の円滑化・容易化などに関係したソフトウェア技術・ハードウェア技術である。

本事業の推進に当って、当財団に学識経験者の方々と構成されるデータベース構築・技術開発促進委員会(委員長 前山梨学院大学教授 蓼沼良一氏)を設置している。

この「メディア変換型感性データベースの構築方法の研究」は平成6年度のデータベースの構築促進および技術開発促進事業として、当財団が(財)イメージ情報科学研究所に対して委託実施した課題の一つである。この成果が、データベースに興味をお持ちの方々や諸分野の皆様方のお役に立てば幸いである。

なお、平成6年度データベースの構築促進および技術開発促進事業で実施した課題は、次表のとおりである。

平成7年3月

財団法人 データベース振興センター

平成6年度 データベース構築・技術開発促進委託課題一覧

分野	課題名	委託先
社 会	1. 報道ニュースのマルチメディアデータベース構築 2. 携帯型電子新聞プロトタイプの開発 3. SPORTS POWER INDEXのパイロット版データベースの構築 4. 画像付き地図資料データベースのプロトタイプ の作成 5. パソコンを利用した副作用症例データベースの 構築調査	日本電子計算(株)名古屋支店 (株)日本経済新聞社/(株)日経データ (株)ビデオ・リサーチ (財)地図情報センター 協立医師協同組合
中小企業振興 地域活性化	6. 関西イベント&プロジェクトデータベース構築 7. ハイビジョン大型映像ソフト流通促進用データ ベース構築 8. 分散協調型データベースのための高度運用シス テムの開発と学習支援環境への応用	(株)京都新聞社 (財)大阪科学技術センター (株)エマーズ
海 外	9. 電子デバイス情報の電子化ドキュメントのプロ トタイプ作成 10. 海外ユーザ向けの国産データベース検索支援用 端末ソフトウェアの開発	電子デバイス情報サービス(株) カテナ(株)
技 術	11. 連想検索写真データベースのプロトタイプ作成 12. レーザー研究データベースの構築 13. 砂漠化防止技術データベース 14. メディア変換型感性データベースの構築方法の 研究 15. フォールトトレランスな多データベースサーバ システムに関する調査研究 16. 創出キーワードの自動付与に関する調査研究	シャープ(株) (財)応用光学研究所 (株)新産業創造センター (財)イメージ情報科学研究所 (株)シネジャーナルプロダクション (株)エレクトロニック・ライブラリー

目 次

序

調査研究の目的と計画

第1章 ネットワーク・マルチメディア・メディア変換

1. 1	インターネットとマルチメディアデータベース	3
1. 1. 1	WWWとMosaic	3
1. 1. 2	企業・大学ベースから個人・家庭ベースへの広がり	4
1. 2	マルチメディアとメディア変換	4
1. 2. 1	マルチモーダル知覚	4
1. 2. 2	感性情報と物理情報	6
1. 2. 3	メディア変換による音響と画像の相互利用	7
1. 2. 4	データベースへの応用展開の必要性和将来性	8

第2章 産業・研究分野でのニーズと従来の関連研究

2. 1	印象語による相互変換	10
2. 2	知覚情報による相互変換	10
2. 2. 1	画像から音への変換	11
2. 2. 2	音から画像への変換	11
2. 3	信号による相互変換	12
2. 3. 1	画像から音への変換	12
2. 3. 2	音から画像への変換	12
2. 4	画像と音の統合	13
2. 5	具体的なニーズ事例・研究事例の紹介	13
2. 5. 1	メディア変換型感性データベース構築の試みと問題点	14
2. 5. 2	印刷とカラーマネージメントシステム	18
2. 5. 3	「音」の表現方法に関する実験と考察 ～効果音検索システムに向けて～	22
2. 5. 4	繊維検反作業とその自動化	27
2. 5. 5	超音波探傷結果の画像化例	30

第3章 メディア変換の要素技術の現状

3. 1	画像から音響への直接変換	33
3. 1. 1	リングオペレータの利用による濃淡画像からの音の生成	33
3. 1. 2	二値図形の音響化	34
3. 2	物理モデルからの映像と音の同時生成	40
3. 2. 1	基本概念	40
3. 2. 2	処理の流れ	40
3. 2. 3	画像と音の生成	41

第4章 メディア変換型感性データベースのプロトタイプの概要

4. 1	イメージにもとづくイラスト画の検索システム	43
4. 1. 1	システムの特徴と位置づけ	43
4. 1. 2	作者イメージと物理的特徴	46
4. 1. 3	検索方法	47
4. 1. 4	実験	48
4. 2	感性情報にもとづくイラスト画の自動生成システム	51
4. 2. 1	システムの概観	51
4. 2. 2	イメージの推定と解析	52
4. 2. 3	データベースの構築	53
4. 2. 4	イラスト画の自動生成	54
4. 2. 5	生成結果	57
4. 2. 6	心理評価実験	57
4. 3	音楽からの感性情報の抽出	59
4. 3. 1	音楽プリミティブ	59
4. 3. 2	リズムパターンの認識	60
4. 3. 3	調性の認識	61
4. 3. 4	メロディの抽出	61
4. 3. 5	コード進行の認識	62
4. 3. 6	音楽プリミティブと形容詞のリンク	63
4. 3. 7	実験および検討	64
4. 4	感性パラメータを利用した作編曲支援システム	66
4. 4. 1	システムの構成	66

4. 4. 2	分析ステージの処理	67
4. 4. 3	生成ステージの処理	69
4. 4. 4	実験結果	71

第5章 将来に向けての課題と展望

5. 1	システム提案	75
5. 2	これからの課題	77
5. 2. 1	要素技術における課題	77
5. 2. 2	システムとしての課題	77

調査研究の目的と計画

1. 目的

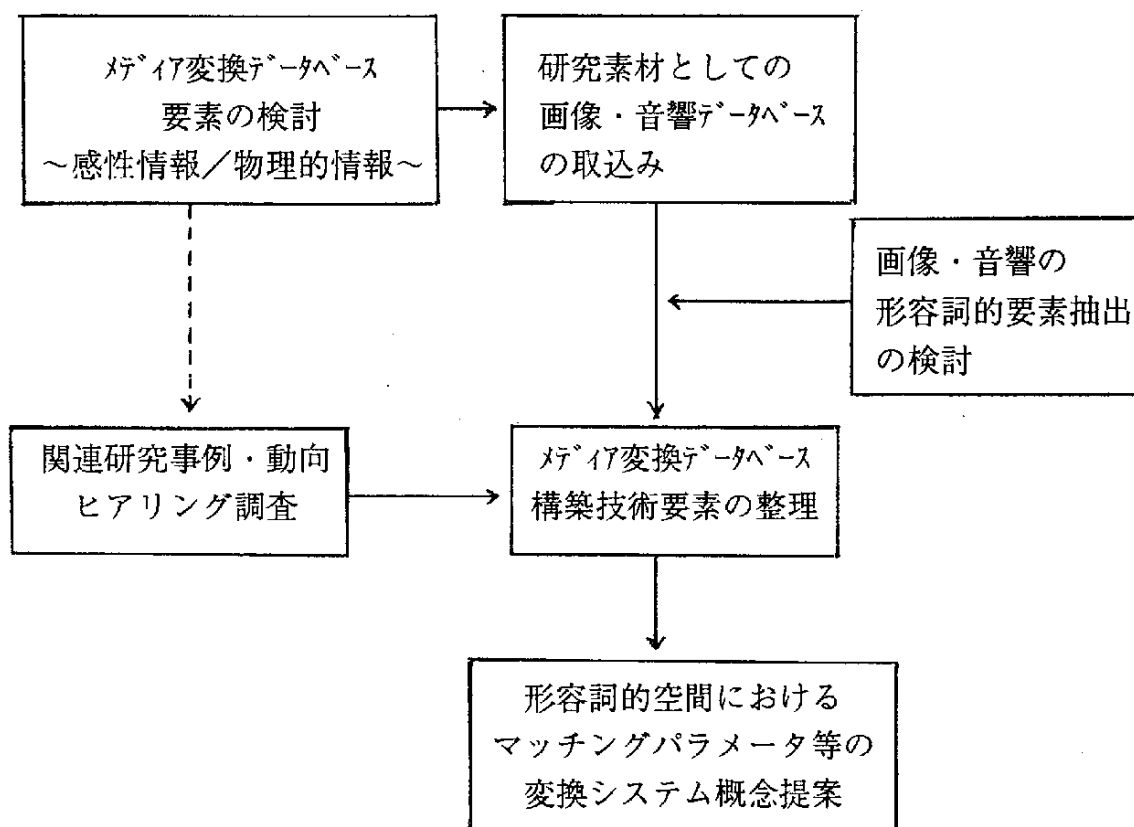
近年の豊かな生活を追求する時代背景の中で、コミュニケーションや商品デザインなどにおいて個性を重視する試みが盛んになるとともに、文化や芸術のような創造活動が作家や芸術家だけでなく一般市民の間でも高まりつつある。

他方、デジタル化技術の進歩によるマルチメディア技術の進展、人工知能や感性情報の研究にみられるようにコンピュータを利用して人間の個性や創造活動を支援する環境が整いつつある。

しかしながら、個性や創造活動に密接に関係する感性に係わるイメージ情報の研究は緒についたところであり、データベースについても、画像は画像、音楽は音楽などと個別的に人間の感性にマッチした画像や音楽などの検索技術は研究されているが、元来人間の感性を表現するためには、画像や音楽などの視覚情報と聴覚情報を統合的に扱うことが重要であり、視覚、聴覚情報をリンクさせて検索する技術が必要と言える。

そこで、本研究では、画像で音楽を、音楽で画像を検索することができる感性を中心にした新しいデータベースの可能性を調査研究し、より人間中心の情報処理技術の進歩と普及促進に寄与する。

2. 調査研究フロー



3. 委員会メンバー（順不同・敬称略）

委員長	井口 征士氏	大阪大学 基礎工学部 教授
委員	田中 央氏	神戸芸術工科大学 工業デザイン学科 教授
〃	橋本 周司氏	早稲田大学 理工学部 教授
〃	今井 正和氏	奈良先端科学技術大学院大学 情報科学研究科 助教授
〃	岩宮 眞一郎氏	九州芸術工科大学 音響設計学科 助教授
〃	佐藤 宏介氏	奈良先端科学技術大学院大学 情報科学研究科 助教授
〃	三浦 佳世氏	樟蔭女子短期大学 人間関係科 助教授
〃	東 和弘氏	凸版印刷（株） 関西商印事業部販売促進本部 システム開発グループ
〃	北川 克一氏	東レ（株） エレクトロニクス開発センター 所長
〃	田原 敏孝氏	関西テレビ（株） 編成制作局 専門職局長
〃	田村 洋一氏	住友金属工業（株） システムエンジニアリング事業本部 専門部長
〃	山形 毅章氏	西日本旅客鉄道（株） 技術開発推進部 レールウェイテクノセンター 所長
〃	和気 早苗氏	日本電気（株） 関西C&C研究所 研究員
協力研究員	才脇 直樹氏	大阪大学 基礎工学部 助手
〃	藤田 泰成氏	大阪大学 基礎工学部
〃	高垣 農氏	石川島システムテクノロジー（株） 第2プロジェクトグループ

第1章 ネットワーク・マルチメディア・メディア変換

1. 1 インターネットとマルチメディアデータベース

1. 1. 1 WWWとMosaic

「インターネット」と「マルチメディア」、昨今においては、まさにこの二つのキーワードが世界を席卷している。アメリカのゴア副大統領の提言による情報スーパーハイウェイ構想と、それに伴うNII(National Information Infrastructure)も、インターネットをその良きモデルとして位置づけていると思われる。さて、そのインターネットの世界規模のデータベースとしての機能と可能性に改めて世間の注目を引きつけ、ネットワーク上での情報共有の流れを加速させるのに大きな役割をはたしていると言われるのが、Mosaicである。

Mosaicは、1993年秋にNCSA(National Center for Supercomputing Applications)から発表されたマルチメディアアプリケーションソフトウェアであり、主としてインターネット上のWWW(World Wide Web)というサービスのブラウザ(検索参照ツール)としての機能を持つ。

ここでWWWとは、1989年にCERN(European Particle Physics Laboratory)で開発されたもので、さまざまなサーバーに分散して保存されているデータを、ハイパーテキストのリンク構造でもって関連づけて参照するための仕組みである。すなわち、インターネット上の世界各地のサーバーが持っている情報に、文字どおり網の目のようにリンクを張り巡らせて、それを手繰ってサーバー間を行き来することにより、より詳細な情報や関連知識を容易に得ることが可能である。むろん、ユーザーにとっては、ただ現在参照しているドキュメントの中からより詳しく知りたいキーワードなどを選択すると、その情報が表示されるだけのことであって、自分が世界中のサーバーを渡り歩いているというようなことは特別意識されないし、その必要もない。

さてもともとWWW自身、単なるテキスト以外に音声や画像なども扱える機能を有していたのだが、テキストコマンドベースの運用ではいまひとつその機能を活かしきれていなかった。そこへ、テキスト・音声・画像を統合して、あたかもモニターやスピーカが埋め込まれた雑誌のページのように表示し、GUIを採用して、画像をクリックすればその部分が動画として再生されたり、文章中の特定の単語をクリックすれば関連情報が表示される、などの機能を有したMosaicが登場したのである。これによって、WWWの利用者が飛躍的に増大したことは言うまでもなく、現在では一種のブームとなっている。これを利用してネットワーク上に雑誌や放送局、ショールームを実現するような試みもあって、まさにインターネットとマルチメディアの代名詞となっている感がある。

1. 1. 2 企業・大学ベースから個人・家庭ベースへの広がり

従来では、インターネットの利用者はほとんどが企業や大学であったが、アメリカにおいてはすでに数年前から、我が国においても昨年あたりから、インターネットブームとあいまって、一般の各家庭レベル、すなわち自分の机の上のパソコンからでも、インターネットに接続できるサービスが普及の兆しを見せている。そして、それ以前から存在している各種パソコン通信サービスにおいても、さまざまなデータベースサービスが提供されている。

このように一般の多くの人々が、ネットワークを通じて容易に大規模なデータベースサービスを受けられる状況が急速に整いつつある今日においては、よりヒューマンフレンドリーな観点に基づくデータベースシステムの構築や、人間の感覚や感性になじみやすい検索手法の確立が望まれており、それらに関する研究の重要性が増大している。

1. 2 マルチメディアとメディア変換

1. 2. 1 マルチモーダル知覚

人間は生まれながらにして、複数の異なった属性（マルチモーダル）の感覚を駆使して外界の情報を取り入れる能力を持っている。しかもその能力は、単に複数チャネルの情報を同時に個別処理するというだけでなく、相補し統合されてより高次の感覚を産み出しており、それこそが人間の高度な認識判断の元になっていると思われる。まずは、なかでも視覚と聴覚の相互作用の重要性について、2つの例を挙げて検討する。

(1) マガーク効果 (McGurk and MacDonald, 1976)

イギリスの心理学者マガークらの発表によると、視覚と聴覚に互いに矛盾した情報を与えてやった場合、脳は2つの情報の中間的な第3の情報を認識するという。

たとえば被験者に対して、「ば(ba)」という音声に同期させて、「が(ga)」と発音している人の顔の映像のみを見せた場合、その被験者には「ば」でも「が」でもなく「だ(da)」という音が認識される。ATRの東倉らの調査では、この現象は、日本人よりもアメリカ人に、また母国語認識よりも外国語認識の際に強く現れることがわかった。つまり、相手の表情や口元により注意を払うほど、効果が顕著に現れるということであろう。

(2) 雑音中での音声認識能力

町の雑踏の中や人が多く集まっている室内など雑音・騒音が多い場所で、誰かと会話をする際には、我々は知らず知らずに相手の口元を注視していることが多い。

我々は雑音中の音声のみでも、あるいは唇の動きからだけでも、ある程度もとの内容を認識することができるが、この二つの知覚を同時に利用した場合、その認識率は個々の場合の単純な足し算以上の上昇を示すという結果がある（図1-1）。

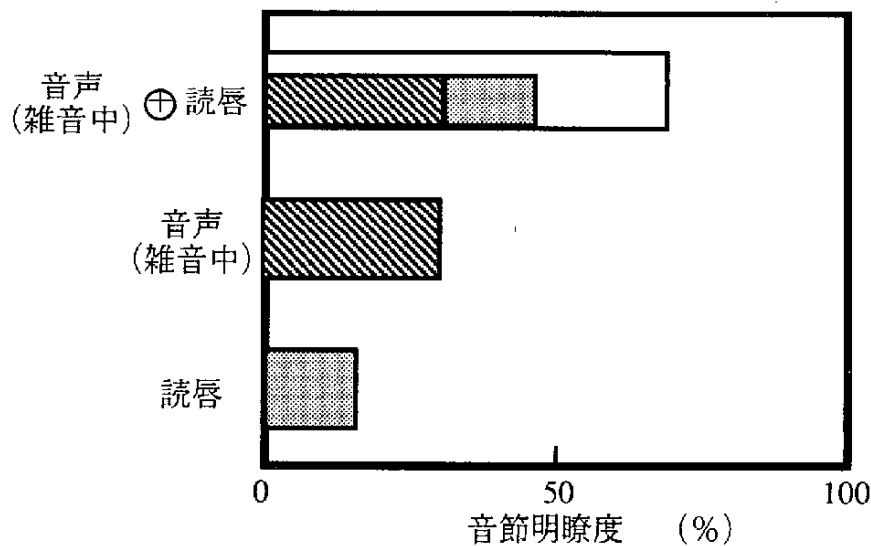


図1-1 雑音中の音声知覚

ここで、聴覚（音）と視覚（画像）の属性の違いについて少し考えてみる。例えば、音は一過性のため記憶に留めたり他と比較するのが難しいが、画像はフリーズできるため定量的な比較が可能である。一方、聴覚は特に注意を払っていなくても、刺激に反応することができるが、眼の方は閉じていたり視野に入っていなかった場合には、まったく情報入力が行われない。また、画像のダイナミックレンジは音のレンジに比べて狭く、微妙な変化を検知するには耳の方が感度が高い。このように、どちらか一方だけではそれぞれ欠点もあるが、2つを協調的に利用すれば互いに補い合い、全体としての認識能を格段に高めることができる。上の例は、まさにその証明であると言えよう。

さて以上の2例は、視覚と聴覚の協調性の効果と重要性を示すものであったが、その他にも、物をつかむ際の視覚と触覚と運動感覚の協調フィードバック機構などは言うに及ばず、話をする場合の聴覚のフィードバック機構についても次のような興味深い話がある。

(3) 変換聴覚フィードバック

マイクに向かって自分が話した内容がそのままヘッドホンから聞こえるようにした装置を使い、被験者に対して、自分の声をリアルタイムに聞ける状態の場合と、一定の遅延時間をもって声が聞こえてくる状況の場合で、それぞれ話をさせて結果を聞き比べてみると、後者の場合には、どもったりしどろもどろの話振りになる傾向が強く見られた。

このように人間にとっては、マルチモーダル知覚による情報処理や認識は、我々自身がふだん意識している以上に自然なものであり、さまざまな情報処理システムにおいてより良いヒューマンインタフェースをめざすためには、このような人間の知覚システムによりなじみやすい体系や手法を開発し採用する必要があると言える。

また、このようなマルチモーダル知覚は、いわば人間の感覚レンジにおける横のレンジを広げる役割を担っているが、縦のレンジを広げるものとして錯視・錯覚が挙げられる。すなわち、本来存在しない対象を認識したり、同じ対象を異なるものとして認識したりというような作用であるが、これらは我々のミスや混乱を誘う半面、活用の仕方によっては、人間にとって自然な感覚や印象を与える有効な手段となりえている。このような作用により、人間の知覚システムの個々のレンジが本来それほど広くないにもかかわらず、非常に高度な処理をこなすことができるのだとすれば、将来的には実用的研究対象として考慮にいれる必要があると思われる。

1. 2. 2 感性情報と物理情報

さて、前項では生体的な感覚に視点をおいたが、我々の日常においては、さらに心理的・感情的な要因も含んだ「感性」というものが、大きなウェートを持ってくる。しかしながら、論理的・客観的な物理情報に比べて、主観的感覚によるところの大きい感性情報の利用技術は、その有用性にもかかわらず今だ一般的とは言えない。そこで本項では、感性情報の性質・内容について簡単に述べる。

物理情報とは、画像であれば色・濃度・距離など、音響であれば音の大きさ・高さなど、文字どおり物理量によって客観的かつ定量的に表現することができるものである。それに対して感性情報とは、絵や音楽から受ける印象などの、感情や情動に関わる情報である。感性情報のより詳しい定義としてはいろいろな説があるが、一つの例として以下のような分類が提案されている。

(1) パターン感性情報

パターン情報の中で、物の質感、楽器の音色、音楽の表現など、物理量ではあるが同定が困難であるもの。

(2) シンボル感性情報

例えば、「暗い」＝「画像における明度の低さ」＝「音楽のマイナー」など、パターン感性情報の中で、特定の形容詞（シンボル）と密接に結び付いているもの。

(3) パラメータ感性情報

あらかじめ定義された概念（形容詞）空間において、1つの座標として表現できるパターン感性情報で、複数の感性パラメータ軸で記述されるもの。

(4) イメージ感性

人間の心理システムに深く依存し、イメージとして働く主観や直感のこと。これは認知科学や心理学の領域であり、工学的情報としては扱いにくい。

現段階での感性情報処理の研究は、これらのうち(1)～(3)に分類される感性情報の扱いが中心であり、(4)についてはまだあまり進んでいるとは言えない。しかし、ふつう我々が「感性」と言う場合には、(4)のようなものを指していることが多く、これからはさらにこの領域に踏み込んだ研究が必要とされている。

1. 2. 3 メディア変換による音響と画像の相互利用

本節第1項では、マルチモーダル知覚の積極的活用がより優れたヒューマンインターフェースの実現に不可欠であることを、いくつか例を挙げて検証した。そのための技術としては、いわゆるマルチメディア技術が挙げられるが、従来のマルチメディア技術の分野においては、画像処理と音響処理をただ単に並立同期に処理する技術の開発に重点が置かれており、両者の枠を越えてより融合的・相互的に利用するインターメディア的とも言うべき観点に欠けている。第2項で述べたような感性情報を活用し、よりヒューマンフレンドリーなシステムを構築するためには、我々人間が、ふだん何げなく視覚と聴覚を融合的・相互的に利用して、高度な知覚能力を実現しているのと同じように、機械やコンピュータにおいても画像と音響の融合的・相互的利用手法を用いて、自然な感覚で人と機械がコミュニケーションを行うことが可能であることが望ましい。

本研究ではこのような状況を踏まえて、一つのキーワードとして、音の可視化や画像の可聴化といった画像情報と音響情報の相互変換(メディア変換)を提示し、その可能性を調査・検討した。その際、ひとつ注意を要するのは、このようなメディア変換技術というものは、包括的な概念であって決して単一の方法論から成り立っているわけではなく、対象に応じてそれぞれ異なる手法の開発が必要とされるということである。

例えば、フクロウが、聴覚から得た音の周波数・時間・音圧を並列に処理して聴空間地図を作成し、これを視覚から得られた視空間地図とマッチングすることにより周囲の環境を把握している、という研究が最近大きな注目を集めた。これは、メディア変換の観点からすれば、異なるメディアから得られた情報を空間という共通の中間表現に置き換えてマッチングを行った事例に相当するものであり、もし現在の工学技術で人工的にこれを実現するとすれば、三次元画像計測装置とソナーを組み合わせれば可能であろう。

一方、人間の場合には、この例のように単に数学的・物理的に視覚と聴覚のマッチングをはかるだけではなく、心理的な要素がからんだメディア間の変換作用が見受けられる。例えば、我々は絵画を觀賞する際に、その絵から受ける印象によって

いろいろな音楽を頭の中に思い浮かべる事がある。また逆に、ある音楽から何かの場面や映像を連想する事もありえる。このような精神活動に深く根ざしたところの作用は、過去にその人が受けた教育や体験・記憶にもとづく先入観や、その時点での心理状態などにより、著しく影響を受けるものであり、同じく脳内にて行われるメディア間の対応付けであるとはいえ、先の例とは全く異なる原理にもとづくものであると考えねばならない。

最後にこのメディア変換というキーワードに対する理解の一助のために、その概念図を示しておく。

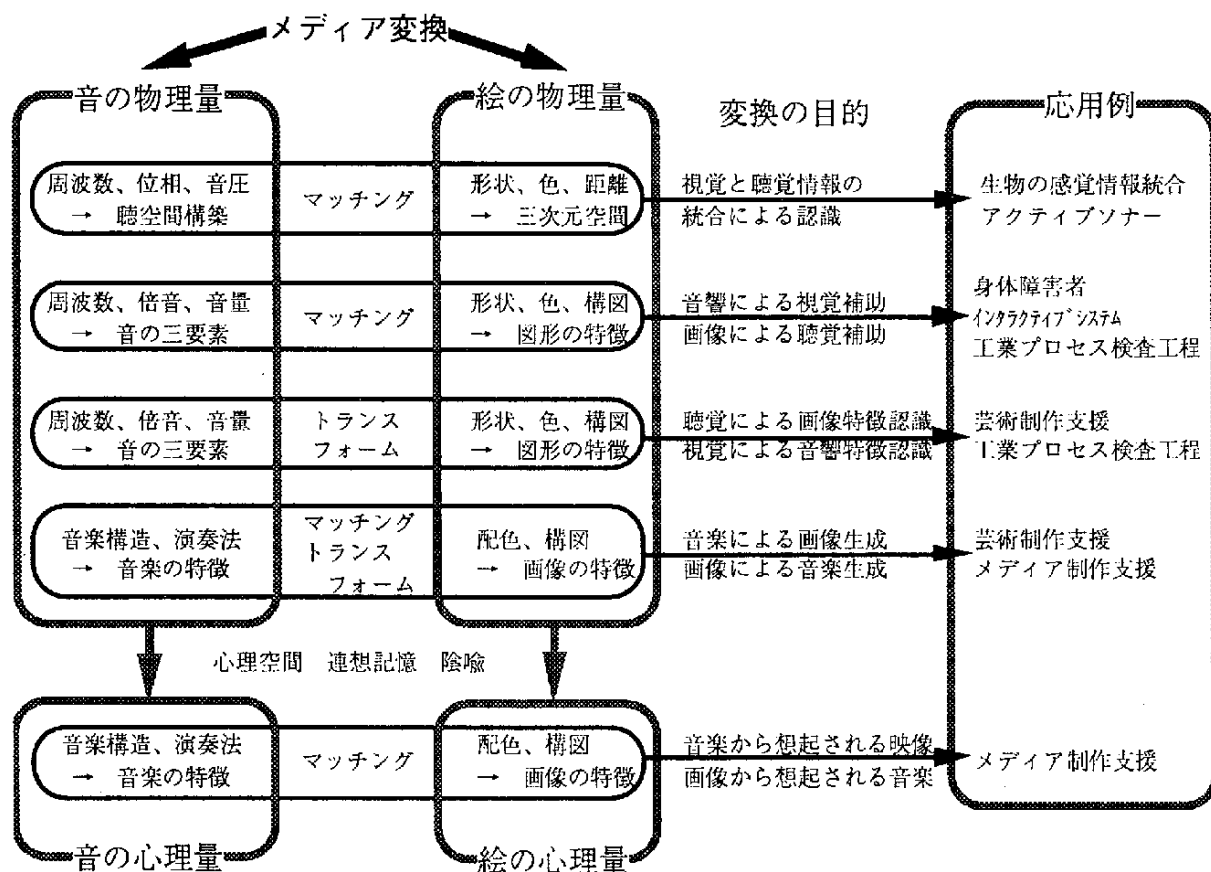


図 1-2 メディア変換の概念

1. 2. 4 データベースへの応用展開の必要性と将来性

本章第1節で概観したように、現在では、ネットワークの普及とマルチメディア技術の進展により、一般のより多くの人々にとっても、日常的にさまざまなデータベースを利用できる機会と手段が、急速に整備されつつある。そのような状況下においては、健常者から身体障害者まで、あるいは子供から高齢者までが、様々なデータを様々な形態でもって、自然な感覚にもとづき容易に検索し利用できるデータベースを提供する必要性が高まっている。

また各種分野において、人間の感性にもとづきあるいは利用したデザイン・コン

の元となる種々の素材をレイヤー化して、統合的・効率的に利用するための環境としての、データベースの存在意義も増大している。本研究では、そのようなニーズに対するアプローチとして、感性をベースに置き、メディア変換技術の導入により、画像と音響を統合的に取り扱うデータベースの構築方法について検討する。

第2章以降では、このような観点から、メディア変換技術やそれを用いた感性データベースについて、社会における具体的ニーズや従来の関連研究、要素技術について検討し、最後に、メディア変換型感性データベースのプロトタイプのあり方を探る。

参考文献

インプレス社：「WWW大航海記」，INTERNET magazine，No.2 (1994)

東倉洋一：「人間情報・マルチメディア・感性」，イメージスクウェア'94，基調講演資料

井口征士他：感性情報処理，電子情報通信学会編，オーム社 (1994)

第2章 産業・研究分野でのニーズと従来の関連研究

第1章では、現在のマルチメディア社会の流れを前提として、人間のマルチモーダル知覚を活用することの必要性から、感性をベースとしたメディア変換技術の導入を提唱した。本章ではそれを受けて、まず画像と音の相互変換という観点から、従来の関連研究例や社会におけるニーズを、変換手法別にまとめて概観し、後半では、具体的にいくつかのニーズ・研究事例を紹介する。

2. 1 印象語による相互変換

人間は、五感を通してさまざまな印象を受ける。特に、絵画や音楽に対して抱く感情は複雑であり、一言の形容詞で表しきれないことも多い。このような心理量を統計分析することにより、さまざまな研究がなされている。

例えば、CMやイメージビデオなど映像と音の統合したメディアについて、映像と音の組み合わせによる印象の変化を調べたり、絵画の色彩情報と形容詞の関係から絵画の時間性を分析したりする研究が行われている。また、感性情報処理の分野では、イラストに対する印象を元にイメージの合成を試みたり、音楽から得られる印象を計算機に感じさせる研究などがある。

このように、印象語を用いて、画像・形容詞間と音・形容詞間の変換をある一定の規則にもとづいて行うことができれば、次のような応用が考えられる。例えば、無音で絵画を見ているときの印象と曲を聞きながら見ているときの印象の違いを、関数で表現することが可能であれば、その関数を逆に用いることによって、音に応じて絵画の方を変化させることができる。すると、音楽のリズムや音色によって絵画が変化し、難聴者に「音楽を見せる」ことができる。

また、デザインや映像・音楽作品の制作などの分野において、素材の検索に大いに活用でき、作業の高度化・効率化を支援できるものと思われる。

2. 2 知覚情報による相互変換

知覚情報とは、形容詞で表される印象語と異なり、個人の感覚に左右されない物理量のことである。例えば、画像では色の3要素（明度・色相・色彩）で示されたり、山や川などのようにオブジェクト名で示されるものである。また、映像においては、物体の衝突などといったような事象も含まれる。一方、音の世界では、音の3要素（音量・音高・音色）によって特徴づけられるドの音や大きい音といった情報や、ピアノの音・鳥のさえずり・ガラスの割れる音などのように具体的な名称で示される情報が考えられる。

2. 2. 1 画像から音への変換

画像からの色彩情報や四角形・三角形といったオブジェクト情報を、個人の感性もしくはニューラルネットで学習させることによって、音のリズムやメロディへ対応付けることができれば、抽象的な音の生成が可能となる。同様な考えで、人間のジェスチャーを画像的に認識し、音の生成や音楽演奏の制御に用いようとする研究もある。また、文字認識技術と組み合わせ文章を音声で出力する、あるいはGUIにおけるアイコンを音楽化するなどの、視覚障害者を支援するシステムなども開発されている。

その他にも、各種シミュレーションやゲームにおいては、映像における動作をサンプリング音に対応させて効果音として出力し、臨場感の形成におおいに役立てている。なお、このようなサンプリング音は、実在する音に限らず架空の音であってもいいわけで、天体シミュレーションにおいて、全体にエコーのかかった星の流れる音などを出力して、「耳で聞くプラネタリウム」を実現することもできる。

2. 2. 2 音から画像への変換

人間は音によって見えないものを想像することができる。例えば、野球場の外で歓声を聞くことで試合の状況を想像したり、遊園地でジェットコースターを待っている間に聞こえてくる悲鳴によって期待と不安を抱いたりする。これらの過程をコンピュータで表現できれば、ボクシングの会場に設置されたマイクからの打撃音・歓声・ブーイング・拍手・ゴングなどの入力によって、その状況を示すCGを作りだすことができるはずである。これはいささか非実用的な例であったが、このような技術を用いれば、監視カメラの設置が非常に困難もしくは不可能であるような場所の状況を、あたかも直接監視しているかのように把握することが可能となるため、さまざまな業界でのニーズが期待できる。

ところで、音刺激が与えられたときに色覚が随伴する現象を色聴というが、この関係を分析して、歴史的建造物を舞台として音楽にシンクロナイズさせて照明を制御し、音楽と光と色を統合したスペクタクルの演出なども試みられている。また、音楽からビートを抽出し、そのビートに合わせてCGダンサーが踊るシステムも構築されている。

あるいは、音声認識技術を用いて、話の内容に対応する手話映像を自動生成するシステムや、表情の自動合成システムなどへの利用が考えられる。

2. 3 信号による相互変換

画像と音を電気信号に変換した情報は、今まで紹介した他の例に比べて最も的確な情報である。それゆえ、正確な出力を必要とするシステムに多く用いられる。画像における信号情報としては、各画素の色情報、物体の形や配置、運動の周期性などが考えられる。また音に関しては、音量・音高・音色の数値情報などが挙げられる。

2. 3. 1 画像から音への変換

映像における周期性を利用することにより、エンジンやプラント内の視覚的に理解しづらい微細情報を、音信号として表現することができる。工場における目視検査やプラントの故障診断、緊急患者の容体チェックなど、常に監視が必要とされる場面においては、この技術を用いて音による情報を与えることにより、見落としや見間違いを防止したり、他の作業を平行して行えるようになる。

また、天気予報で見られる雲の映像で、画素の濃度やパターン認識による台風・前線・気圧分布などの判別を元に、ある地点での雨や雷や波の音といった自然音を直接出力することもできる。同様な考えでもって、血液の流れを可視化して、それを信号に対応させて可聴化した例なども報告されている。

あるいは、実際に信号波形そのもので変換する例として、画像の二次元データを一次元に変換して、音のレンジに対応づけて出力するようにした研究もある。

2. 3. 2 音から画像への変換

自分の聞いている音が好みの音でないときには、それを好みの音にするための分析が必要となるが、グラフィックイコライザなどを用いれば、その音の特性を図形化して目で確認することができるので、好みの音作りが可能となる。

警察では、音響スペクトログラフによって音声の信号を図形化して、特定の話者の識別に利用している。また、工場における非破壊検査の技術として、超音波探傷試験における試験結果の記録性や再現性を改善するために、三次元画像表示することによって解釈が容易な記録表示を実現した例もある。

また、医療の現場においては、昔ながらの打診法により、身体をぼんと叩いた音の大小・高低・減衰率から体内組織の空気占有量を判別して、各器官の形状や組織の状態を診断したり、聴診器で心音や呼吸音を聞いて正常かどうかを判断しているが、これら各医師が知識と経験にもとづいて判断している過程を整理して、コンピュータを用いて画像化して見せることなども考えられる。

その他、ある建物内に設置されたマイクに拾われる足音を分析することによって、防犯カメラの死角部分をカバーして、不法侵入者の発見に役立てることもできる。

2. 4 画像と音の統合

画像と音を同時に示すことにより、臨場感が増し理解が深まるということについては、これまでも何度か述べてきた。

テレビや映画の世界では、臨場感を増すためだけでなく、例えば、アニメの映像の不連続さを効果音の補助によって連続的に感じさせたり、ドラマのシチュエーションの切り替え時に、映像よりもBGMの切り替えを先行させることにより、不自然さが出ないようにする、などの工夫に用いられている。

人工現実の分野では、相互に関連した情報をすべて画像で表示するのではなく、あるものは画像、またあるものは音、といった具合に相補的に用いたり、コンピュータが多量の計算を行っていてユーザーを待たせなければならない場合に、簡単な音楽を流すなどの配慮がなされている。

工業応用の例では、騒音の中でも話声を聞き分ける視聴覚センサフュージョンシステムがある。これは、視覚系で口を動かしている人を見つけ出してその人に最も近い聴覚系を機能させ、口の動きと連動して変化する信号だけを取り出すという仕組みになっており、カクテルパーティ効果をロボットで実現したものである。また、発電設備の現場異常検知システムとして、異常音の発生源を三次元音源同定により特定し、カメラをそこに向けて現場の画像を取り込んで処理するようなものも開発されており、画像と音の連係が異常検知に役立てられている。

2. 5 具体的なニーズ事例・研究事例の紹介

次頁以降で、いろいろな大学や企業で実際に行われている、メディア変換もしくはそれに類すると思われる事柄についてのニーズや研究の内容を、いくつか取り上げて具体的に紹介する。

2. 5. 1 メディア変換型感性データベース構築の試みと問題点

(1) 感性情報の直接変換と間接変換

異なるモダリティー間でデータ変換を行う場合、それぞれの情報を直接的に変換する方法と、イメージや概念を仲介させて変換を行う方法とが考えられる。ここでは、視覚と聴覚の間でのモダリティー変換、それも感性情報に関わる例を取り挙げ、メディア変換型感性データベースの可能性と問題点とを考えたい。

視覚情報を聴覚情報に直接的に変換する例として、神山純一の楽曲「オーロラを聴く」(1994)や野村仁の写真「バード・スコア」(1987)を挙げることができる。前者は時間軸上でのオーロラの輝度変化(視覚情報)を12音階に振り分け、シンセサイザーと木管楽器によって演奏を行ったものであり、後者は5線譜を写し込んだ35ミリフィルムによって湖面に浮かぶ水鳥の群れを撮影し、それを楽譜に見立ててチェロ等による演奏を試みたものである。こうした環境音楽的な試みに対し、人工的な視覚対象を音へと変換することを企画した作品としては、子供にブロックを組立てさせ、出来上がったブロックの高さや広がり、色によって自動演奏を行う「オートブロック」(多摩美大, 1992)のアイデアを挙げることができる。

しかし、こうした直接変換では、視覚情報と聴覚情報における印象の一致は偶然性にゆだねられ、原曲に操作を加えて作品の完成度を高めることはできても、データベースの基礎として取り挙げるには問題となる。

一方、異なるモダリティー間に印象や概念を仲介させ、両者を間接的に変換する方法を用いると、汎用型データベースの構築が可能となる。その際、仲介させるものが必ずしも言語である必要はないが、言語を用いることにより、検索性や多目的性を高めることが期待できる。次に、そうした変換をめざして行われた研究例を紹介する。

(2) 視覚・聴覚間の感性情報変換の試み：時間印象を仲介として

ここでは形容語によるイメージ表現を仲介に、音と映像の変換を目指した研究(三浦, 1994, 1995)を取り挙げる。

18世紀の美学者レッシングは、音楽など主として聴覚情報に基づく「時間芸術」と、絵画など主として視覚情報に基づく「空間芸術」とを区別した。しかし、古来より天体の運行や川の流れなどの「視覚情報」の中に「時間印象」を見い出し、時にはそのイメージを円環や渦巻あるいは直線に抽象化させて時間印象を視覚的に表現してきたのが人間である。一方、「空間表現」、例えば、静止した絵画や写真の中に、「疾走感」や「瞬間」、「永遠」などの「時間印象」を読みとることも我々にとって難しいことではない。とすれば、「時間印象」を介在させることにより、視覚芸術と聴覚芸術を対応づけることは可能なはずである。さらに、「時間印象」は「精神テンポ」に示されるように、概念であると同時に情緒的、身体的な感性情報でもある。したがって、「感性」データベースを構築するに当たって、「時間」は一つの有効な仲介イメージとなるはずである。こうした問題意識の下、まず、絵

画・写真における時間印象の分析から試みた。

32枚の現代芸術（絵画、写真）に対し、60の形容語リストによって印象を評定させ、クラスター分析、因子分析を行った結果、2次元静止画像から感受しうる時間印象には、「穏やかに持続する時間」、「動きや変化、速度感」、「停止した時間」の3種類のあることが示された。一方、作品も印象の類似度によって3群に分けることができ、各群の作品は、穏やかな、ゆっくりした；動きのある、落ち着かない；神秘的な、時が止まった；印象をもっていることが示された。この結果は形容語による分類結果と一致するものである。また、時間関連の20の形容語に限って因子分析を行い、第4因子まで求めると、上記の3因子に加えて、時間軸（過去・現在・未来）が分離され、これら4つの時間関連印象に対応する空間的もしくは視覚的要因は表2-1のようになることも見い出された。さらに、これらの視覚要因を変数にした追加実験で、視覚要因の変化は時間印象の変化をもたらすことも明らかとなった。

表2-1 時間印象と表現における特徴

過去・現在・未来	：色彩（色相，明度，彩度，色数）
（時間軸）	
持続する時間	：構図（水平・垂直軸による安定感， 大きな面での色の塗り分け）
停止した時間	：余白（比較的単純な構図， 物語の投映の余地）
動き・変化・速度感	：高空間周波数（画面の細分割，分散配置， 激しいタッチ）

一方、表現時においては、これらの空間（視覚）要因に加えて、シンボリックあるいは概念的な表現に重点が置かれることも示された。ただし、用いられた象徴表現が常に鑑賞者の理解を得られるとは限らず、メディア変換型感性データベースの構築においては、多様な評価者による理解度の調査を必要とすることが示唆された。

次に行うべき課題は、音楽作品での時間関連印象の分類と、その印象に寄与する聴覚要因の分析、および映像と音との相乗効果の検討であるが、空間表現の結果が「速度」に関連する次元において展開されたことから、音楽作品での検討はより容易になるものと思われる。広告における音楽と映像の相互効果を検討した稲葉

（1991）の結果では、音楽と映像の交互作用は「躍動感」において有意にもたらされることが示された。今回の実験において「速度感」溢れると判断された作品に、アップテンポの音楽をかぶせることにより、相乗的な効果が得られることが期待で

きる。

なお、フラッター（聴覚的ちらつき感）がCFF（視覚的ちらつき感）に促進的に働くというモダリティー間の相互関連性を指摘する研究がある。「速度感」を規定する主な空間要因が「空間周波数」であったことを考えると、「時間周波数」との関連についての基礎的研究が展開できるものと思われる。

（3）異なるモダリティー間の相互関連に関する基礎的研究

モダリティー間での相互影響あるいは共通感覚に関して、心理学では古くから、「色聴」（ある特定の音に特定の色相が呼応する現象）などの「共感覚」や、「共鳴」（例えば、低音は色を暗く、柔らかい印象に移行させる等の現象）、「通様相現象」（異なるモダリティー間における印象の共通性、例えば「黄色い声」）等の基礎研究が行われてきた。

こうした異なるモダリティー間での感覚レベルでの共通性を、感性情報のデータ変換に利用することは効果的であると考えられるのであるが、実際にそうしたデータを作品の制作やモダリティー間での情報変換に利用した例は、寡聞にして知らない。制作者の感性が結果的に基礎的なデータと対応したケースとしては、色付きの線画に文字のガイドを付加し楽譜として提示したJ. ケージの「アリア」（1958）や、音のピッチとボリュームを色相と面積に対応させ、色環状の譜面を構成した武満徹と杉浦康平の「弦楽器のためのコロナ」（1962）などを挙げることができるが、いずれも演奏における偶然性や不確定性をねらった実験的あるいは概念的な作品であり、その点を考えると、感覚レベルでの直接変換は必ずしも感性情報の変換に有効な手段とならないかもしれない。

（4）感性情報におけるデータベースの特殊性

しかしながら、「感性情報」の変換型データベース構築において、確定的な一対一対応型が望ましいかどうかとも検討されるべきであろう。「イメージ」を伝える場合には、「事実」を伝える場合と違って、より多くの接近可能性があり得る。完成度の高い「変換型」データベースでなくても、自由度の高い「対応型」データベースで十分かもしれない。匿名性を基本とした従来のデータベースとは異なり、個性やオリジナリティを前面に出した複数のデータベースが望まれるかもしれない。一方、変換型データベースを使って自らオリジナリティの問われる仕事をする場合には、個性の強すぎるデータベースは望まれないであろう。ユーザーの好みや個性にに応じて、拡張や学習が可能な成長型のシステムが要求されるかもしれない。一方、異なるモダリティー間での感性情報の「相乗効果」を考える場合、時には、ミスマッチが効を奏する場合もある。どのようなシステムがより豊かな可能性を引き出し得るかについての十分な検討が行われるべきであろう。システムの構築において問われるべきは、「技術的に何が可能か」ではなく、「誰のための」、「何のための」システムかであろう。

(5) 情報変換における感性次元での効果

ユーゴスラビアで開かれた視覚会議のオープニングセレモニーは印象的なものであった。生の合唱（教会音楽）に東方教会のイコンのスライドを重ねただけの素朴な手法であったが、視聴者に大きな感動を与え、単一モダリティーによるプレゼンテーションでは得難い感性に強く訴える作品となっていた。

視覚情報と聴覚情報の相乗効果は、認知・理解の次元よりも感性次元に影響するという知見がある（例えば、Sewellら、1980）。モダリティー変換型データベースの構築に当たって、感性情報が注目されるのはこの点にも依る。

かつて音楽評論家、吉田秀和は、映像を見ながら聞くのは、聞くのに専心する場合よりも、深い聞き方を可能にすることを指摘した。芥川也寸志は、「コンサートで目を閉じて音楽を聞くのは、作曲家の意図を誤解して聞き取ってしまう危険性がある」とまで語った。しかし、演奏をそのまま映像として提示することが、音楽の感受あるいは理解に最適な方法であるとは限らず、他方、音楽を対応する映像に変換し、両者を重ねて常に最適の形で提示することが望ましいとも限らない。結果的に視聴者の想像力を抑制したり、一定の印象へ強制的に導かないとも限らないからである。

マルチメディアの発達は、異なるモダリティーによる情報の変換とそれに基づく効果的な情報提示を可能にする。しかし、そうしたメリットの中にあっても、ハイデガーの指摘した「文化的得失」、すなわち、技術によって我々が失うものにも目を向けていく必要がある。

引用文献

- 三浦佳世（1994）「時間」表現の空間構造—絵画における視覚要因の検討を通して、平成6年度文部省科研費重点領域研究「感性情報」全体会議予稿集,80-84
- 三浦佳世（1995）「時間」表現の空間構造—絵画における視覚要因の検討を通して、平成6年度文部省科研費重点領域研究「感性情報」研究成果報告書
- 稲葉哲郎（1991）広告反応におよぼす音楽と映像の効果，第55回日本心理学会発表論文集，873
- Sewell,E.H.Jr. & Moore,R.J.（1980）Cartoon embellishments in informative presentations.Educational Communication and Technology Journal,28,39-46

2. 5. 2 印刷とカラーマネージメントシステム

(1) はじめに

印刷における色の再現は他のハードコピー技術に比べはるかに高い品質を提供してきた。これは常にクライアントの厳しい要求に対して、その要求に答えられるような商品（印刷物）を作り出さないと生き抜いて行けなかったという事情によるものではある。

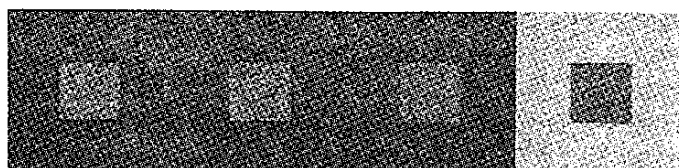
ところで、この色の再現という問題は実はあいまいなものでもある。色とは多分に心理的なもので、周波数や波の強度といった物理量だけでは記述できず、色相や明るさといった心理的感覚量で変わってくるものなのである。

ここでは、以下に人間の知覚のあいまいさ感覚、その感覚に印刷が答えてきた例、そして最近のデジタルイメージからの印刷について紹介する。

(2) 人間の感覚について

イ. コントラストと明るさ感覚

通常、光の物理的強度が増加するにつれ、心理的な知覚も同じように増加すると考えられるが、対象の明るさは周辺視野（背景）に依存している（図2-1）。対象物と背景の物理的なコントラストは、環境全体が明るくなれば、対象物の強度も背景の強度も同じ割合で増加するため、その環境が明るくても暗くても変わらない。しかし環境が明るくなった場合、対象物の明るさは必ずしも増加するとは限らない。中央部（対象）と周辺部のいろいろなコントラスト比に関して、照明の増加に伴って中央部の明るさが変化する様子を図2-2に示す。このコントラスト比が高い場合、照明光の強度が増加するにつれ次第に暗く見えてくるといふ効果が生じるのである。



Cornsweet
(1970) より。

図2-1 コントラストと明るさ感覚の実験

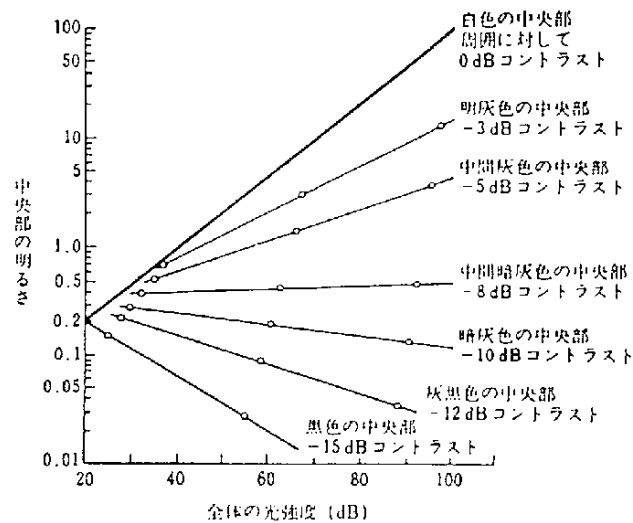
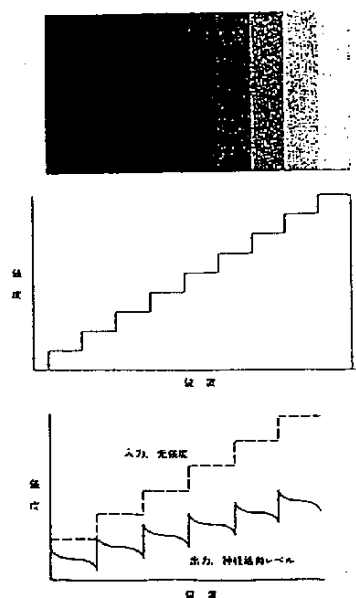


図 2-2 コントラスト比における明るさの感覚

ロ. マッハ現象

マッハ現象とは、輪郭強調特性の現象の事である。これは薄暗く照明された領域から、かなり強く照明された領域までの濃淡を持つ視覚イメージは、実際よりもかなり鮮明な光強度の変化として知覚されるというものである。これを図 2-3 に示す。各々の帯は一定の強度のはずにも関わらず一様な明るさには知覚されないのである。各灰色の帯は、その帯の左から右へかけて明るさが変化しているように見える。



成績で示された入力強度分布から予測される出力パターン (実験)。
Corasweet (1970) より。

図 2-3 マッハ現象

ハ. 印刷へのノウハウ

上で紹介したように人間の感覚は、物理的には同じ物が違って見えたり、逆に違うはずの物が同じに見えたりするし、また全く同じ物を違った場所で違った環境下で見るとまた違って見えたりするという、非常に定量化しにくいやっかいなものなのである。

印刷物を作る上では、物理的に同じであるという事は問題ではなく、いかに同じに見えるかというほうが重要なテーマとなってくる。このために上のような効果をノウハウとして取り入れながら商品作りをしているのである。

(3) 印刷物の校正

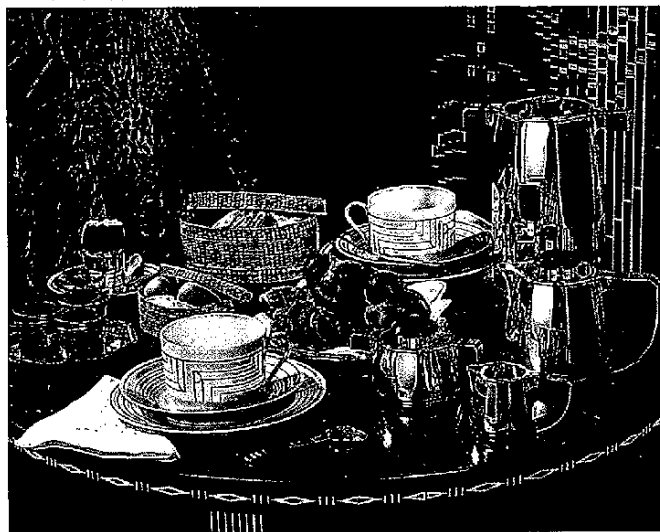
印刷物は基本的に受注されるものであり、必ずそこにはクライアントが控え、気に入った印刷物にするために、さまざまな注文を出してくる物である。この注文には、印刷物として表現するにはかなり難しいものもある。

例えば写真について「もっとおいしそうに」とか「もっと子供らしく」「ムードを出す」などといった、どのようにそれを表現するのかが分からない雰囲気的な注文も実際には多い。このような注文はなるべくやめて頂きたいのが本音であるが、今までのノウハウとしてそのような場合での表現方法もあるにはある。

上記の印刷物の校正での直しの例を図2-4に示す。ただし、上のような注文は入っていない。校正の表現方法での標準とは言えないからである。

校正紙では、明るい部分が暗く仕上がってしまい、全体に重く、シャドー部分の微妙なトーンもつぶれています。
校正の指示は「ハイライト部分を出す」が適当です。あるいは「カーテンのトーンを出す」と指示すると、全体の明暗が出て、メリハリのある仕上がりになります。

▼全体に暗い例



▼修整後・ハイライトと明暗がはっきりと出た

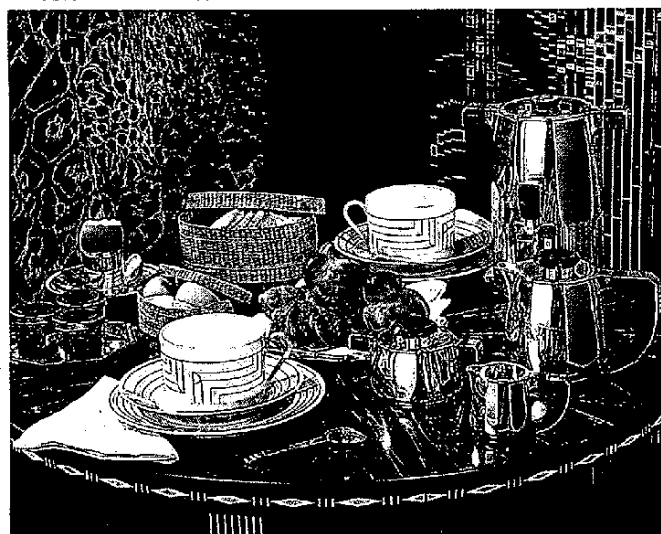


図2-4 印刷物の校正での直し

(4) ハイビジョンの印刷応用

イ. ハイビジョン印刷

印刷分野ではハイビジョン映像を原稿とするハイビジョン印刷がすでに実用化されている。ハイビジョン印刷には「動画から静止画への変換を行うため決定的瞬間を捉えることができる」「写真処理を必要としないため処理時間の短縮が可能」といった利点がある。画質の点で制限はあるものの、スポーツ・イベント等の速報には期待がもたれている。実際に「日刊スポーツ」において、神宮球場でのハイビジョンナイター中継から決定的瞬間を選び出し、翌日の1面を飾ったこともある（ハイビジョン中継自体がまだテスト段階のため恒常的には行われてはいない）。

ロ. ハイビジョン静止画ソフト

ハイビジョンと印刷の結びつきでは、ハイビジョンモニタで絵画をはじめとする芸術作品を鑑賞する「ハイビジョン静止画ソフト」も注目されている。製版スキャナを利用した高度な画像入力技術と、文字と画像を美しく組み合わせるページネーションのノウハウを生かし多くのハイビジョン静止画ソフトが作られている。美術館相互のソフト交換による展示内容の充実や、展示しきれないコレクションの公開手段として、利用が増え始めている。ハイビジョン印刷が入力デバイスとしてハイビジョンを利用しているのに対し、ハイビジョン静止画ソフトではハイビジョンを出力デバイスとして利用する。

(5) まとめ

いわゆるイメージ情報というものを、印刷ではどのように表現してきたか、そして印刷ノウハウを使った新しい形での表現方法を紹介した。

あらゆる条件下で、あらゆるメディアで、あらゆる表現方法で人間の感性に訴えかける手法は今後も増えてくるであろうが、どのような場合においても一定の評価を得られるためにはまだまだ解決しなければならない問題は多い。例えば「色の見えの一致」においてはカラーマネジメントシステムの確立、色彩科学においてはCRTとカラーハードコピーを同時に扱えるカラースペースの問題。好ましい色再現に関する再現手法の確立。また画像と音の融合なども求められている問題の一つである。今後もイメージ情報での問題解決が望まれる。

2. 5. 3 「音」の表現方法に関する実験と考察

～効果音検索システムに向けて～

(1) はじめに

近年マルチメディア化の動きが顕著であるが、我々はインタラクションに音を利用する立場から、AUI (AuditoryUserInterface) の研究を行なっている。ここでは、音声に並んで効果音の利用が重要であるが、多種多様な効果音のデータベース化および検索方法は未整理のままである。今回、効果音検索システムの開発に先がけ、人間の「音の表現方法」に関する実験を行ったので、実験結果と考察を報告する。また、最後に音実験の結果をもとに、メディア変換技術の可能性について考察する。

(2) 実験

イ. 目的

実験の目的は“人間は頭の中でイメージした音をどのように表現するのか”を明らかにし、その特徴を検索システムに反映することである。

ロ. 方法

様々な音を相手に説明するという状況において、人間はどのような表現を用いるのかを観察する。

1) 音の分類とサンプル音の選出

サンプル音の選出にあたり、音をその音源および目的から、音楽、音声、環境音、という3つに分類した。さらにそれぞれをイメージ音と符号化音に分けた。イメージ音とは“何事かの結果生じる音素”であり、符号化音とは人間が目的を持って作成した“意味をもつ音および音列”であるとする。これによって音は計6つのカテゴリーに分類することができる。また1つのカテゴリー内で説明できる音(単位音)に対して、複数の単位音によって構成される音(複合音)を考えることができる。

これらのカテゴリーおよび音の発生源、発生状況を考慮し、17種類のサンプル音を選出した。音の分類表とサンプル音は実験結果と共に図2—5に示す。

2) 実験の手順

実験の手順を説明する。

- a. 被験者Aはサンプル音を1種類聴取する。
- b. サンプル聴取後、被験者Aは被験者Bに聞いた音を説明する。説明方法に制限はない。
- c. 被験者Bは説明を聞き、できる限り具体的に音を想像する。被験者Aに対して質問をしてもよい。
- d. 被験者Bは音の想像後、サンプル音を聴取する。

e. 被験者Bに想像した音とサンプル音の類似度および相違点を述べてもらう。
実験は男女14名（7ペア）に対して行なった。PC9801DAを用い、16bit,32kHzステレオサンプリングの音データをペアによってランダム順に提示した。また、実験の様子はビデオテープに録画した。

（3）実験の結果と考察

イ. 音の表現方法（被験者Aの観察）

被験者Aによる音の説明はほとんど全てが「言語」で行われた。言語以外に“絵”，“ジェスチャー”がみられたが、その使用頻度は非常に少なく、言語説明の補助として用いられる場合がほとんどであった。以下では説明に用いられた言葉の特徴を述べる。

1）音の説明に用いられる語彙

用いられた言葉は次の3タイプに分類できた。

- a. 波形 の説明
 - b. 音源 の説明
 - c. 主観 の表現
- a. 波形 聞こえる音そのものについての説明。“物理量の説明”および“擬音語”を含む。
- b. 音源 知識により推理される音源や周囲の状況に関しての説明。“発音源、発音状況の説明”が中心であるが知識に基づいた説明という点から“音の属性<カテゴリー>”も含む。“発音源、発音状況の説明”はさらに以下のように分けられる。
- ・ What…何の音か。
 - ・ How…何がどうした時に鳴るか。
 - ・ When…いつ（時間）発せられるか。
 - ・ Where…どこで発せられるか。
 - ・ Which…比較を用いての説明。
 - ・ Event…どんな状況（行事）で発せられるか。
- c. 主観 音を聞いて感ずる主観的な感想を表現するもの。“形容詞”を中心に“透明な音”といった本来音には適用されない修飾も用いられる。“音を聞いて何かを感ずる”ことの要因は以下の3つにみることができる。
- ・ 波形に起因するもの
 - ・ 音源に関する知識に起因するもの
 - ・ 上記以外のもの（波形・音源の双方に関係）
- “波形・音源・主観”の概念的な関係を図2-6に示す。

音楽		単音		環境音		複合音
イメージ音	符号化音	イメージ音	符号化音	イメージ音	符号化音	
楽音	音楽	音声	言語	日常音	マッセン音	
1. ピアノ音 9:9:0	2. ギター曲 1:15:5	5. 笑い声 3:15:1	7. 話し 6:15:0	8. 川 3:18:0 9. 電車 3:17:1 10. 歩く 4:17:0 11. まな板 8:13:0 12. 割れる 5:15:1 13. 風鈴 6:13:2	14. 電話の ベル 9:12:0	15. 花火大会 6:15:0 16. 秋の野 1:19:1 17. 学校 3:17:1

(注) 数字は、
波形：音源：主観

図 2—5 波形・音源・主観の関係

2) 音の種類、性質、特性と語彙の関係

それぞれのサンプル音に対し最初から3つめまでの説明に前項で示した3タイプのいずれの言葉が用いられたかを集計した。図2—5に波形：音源：主観の比を示す。また、図2—7にサンプル音全体について、3つめまでの説明、最初の説明、に用いられた言葉の比（パーセント比）を示す。全体としては圧倒的に“音源の説明”が多く用いられる。結果から得られる見解を以下に示す。

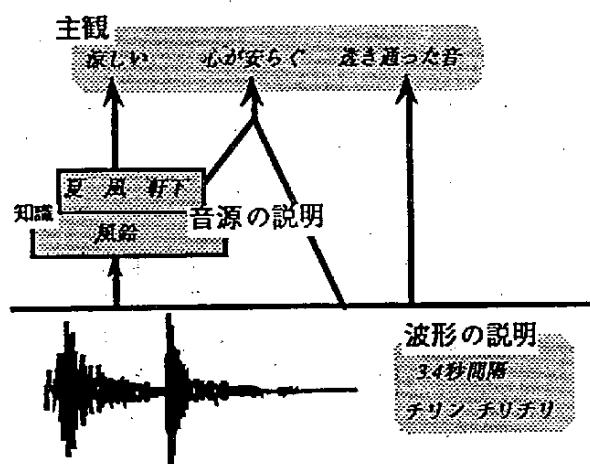


図 2—6 説明の3つめまでに用いられた言葉

	波形	:	音源	:	主観
説明 3つめまで	2 3	:	6 9	:	8
説明 1つめ	7	:	8 6	:	8

図2—7 音の説明に用いられた言葉（パーセント比）

【生活の中に存在する音】音源を説明することで、受け手は発音状況を視覚的にイメージし、音を思い出すようだ。但し、これには音源に関する共通の知識が必要である。擬音語も頻繁に用いられる。

【日常には存在しない音】音源を説明することは困難なため、主観的に感想を述べる。

【その他の単発音】楽音、音楽については、独自の表記方法（音階、調性etc）が存在するためそれが用いられる。言語については、声色の説明と内容の説明の双方が行なわれる。

【複合音】全体についての説明と複合音を構成する単位音それぞれについての説明の双方が行われる。

ロ. 2者間の対話（被験者A・被験者B）

2者間の音伝達において最も重要な要素は“共通の音記憶を探す”ことである。「～の音に比べて～」といったように共通の記憶を基準とすることで、個人による差を埋めようとする。共通の音記憶は、音の想像に（検索においても）非常に大きな役割を持つ。

（4）まとめ

音の説明は、“波形の説明”“音源の説明”“主観の表現”によって行なわれる。“音源の説明”が圧倒的に多く、これが困難な音に関しては“波形の説明”や“主観表現”が用いられる。擬音語は音の種類を問わず、比較的多く用いられる。また、音の伝達には双方共通の知識・音記憶が重要である。

(5) おわりに

メディア変換を行なう場合に問題となるのは、

○各メディア間をつなぐ媒体の設定

○メディア変換を行なうレベルの選択

の2点であろう。上記の実験結果から、変換にある程度の普遍性を求める場合は、媒体としては言語の利用、変換のレベルとしては構成物／音源など知識レベルが適すると類推できる。しかし一方で、媒体として言語を用いた場合、対象が同じ言語圏に限定されるなど、マイナス面もでてくることを忘れてはならない。いずれにせよ、まずメディア変換の利用目的を明確にし、その目的に応じて媒体の設定とレベルの選択を行なう必要がある。

2. 5. 4 繊維検反作業とその自動化

(1) 検反作業

繊維産業における検反作業は、生地の種類、加工プロセスの相違により、欠点の内容が多種多様である。検査対象（工程）としては、繊維工程、織物工程、染色工程の製品があり、それぞれに、無地の場合と柄物の場合がある。

現状の検反作業は、図2-8のような検反機を用いて、目視で行われている。この検反作業の特徴は、次のようにまとめられる。

- ・ 検査幅／速度：幅1000～1500mm；速度20～60m/min
- ・ 特徴：1) 非均一なバックグラウンド；2) 非常に多種多様な欠陥

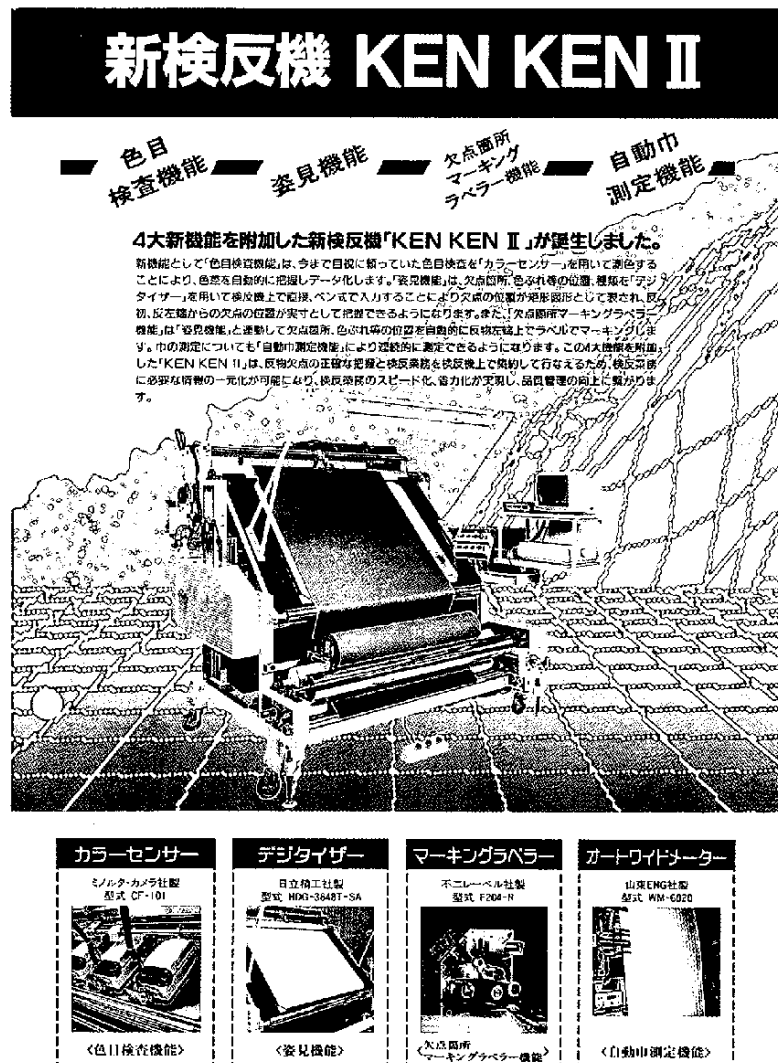


図2-8 検反機

(2) 自動化への取り組み

検反作業の自動化は、かなり以前から試みられてきた。しかし、いまだ実用化されてないのが現状である。

センサ技術としては、当初はレーザスキャンが用いられたが、その後、2次元カメラ(TVカメラ)や、CCDラインセンサが用いられるようになった。

最近(といっても約10年前から)の開発トピックスとして、次のものがある。

イ. 東レ：編物(パンスト)の画像処理

- ・ 官能検査の定量化が目的
- ・ フーリエ変換利用

ロ. 中小企業事業団：“自動縫製加工準備システム”の技術開発(図2-9)

- ・ この中で、自動検反技術開発が取り挙げられた。
- ・ 開発期間：昭和59～62
- ・ 1次元センサ：無地のきず、スラブ、汚れ、穴、破れ
- ・ 2次元カメラ：柄物を含むきず、汚れ、しわ
- ・ コスト、装置サイズなどの点で、実用化されていない。

3 自動検反技術

生地の検査工程は、生地の種類・加工プロセスの相違などにより欠陥が多様・多量であり、欠陥検出能力、検出能力の優れた熟練検査者に依存しているのが現状で、自動化が望まれている。また、生地のマーキング・自動縫製・裁断という工程においても、欠陥検出の事前検出能力が必要となるにもかかわらず、検反作業の迅速システムがいまだ確立されていないのが現状です。

このような風潮に対応するために、検査精度の安定化、検査の省力化、検査情報のシステム化等、実用化へのニーズの高い自動検反技術の開発を行っています。

1. 開発目的

アパレルの欠陥検査工程を参考にして検出精度を安定し、検査有効幅1,000mm、検査速度20m/min以上で検出できることを目標に開発を行いました。

検出検査項目

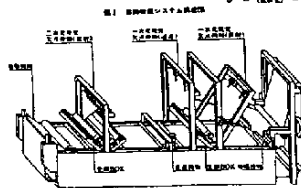
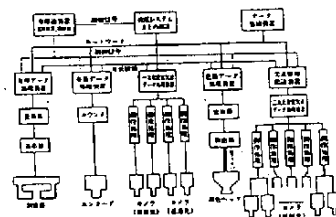
- ・ 幅間—23mm
- ・ 寸長(伸縮度)—0.3% (編物)
- ・ 内れ—グレースケール3級、2mm□
- ・ 穴きず—1mm□
- ・ スラブ—長さ1mm×長さ5mm
- ・ きず—きず—黒色柄、黒色柄、グレースケール2級、2mm□または長さ1mm×長さ5mm
- ・ 色差—ΔE0.4
- ・ しわ—長さ1mm×長さ40mm

2. 概要

装置の基本構成は、各検知センサ部、データ処理部、制御部、駆動部より成り立っています。各検知センサは、検査項目によって分類されており、また、各検知センサとデータ処理部はそれぞれ対になっており、システムとしても単体としても使用が可能です。

このうち、きず、穴きず検出欠陥を検査する機能は、加圧のきず検出を伴う一次元検出欠陥検出機能と、一定の厚さのきずきずにも対応できる二次元検出欠陥検出機能とで構成されています。

開発した装置は、従来の目視検査に比べ、多数の検査項目を検査途中で停止させることなく、連続的な検査が可能であり、検査時間も約30%の短縮を図ることができます。



中小企業事業団：“自動縫製加工準備システム”

図2-9 中小企業事業団：“自動縫製加工準備システム”

ハ. 市販製品：E V S (Elbit Vision Systems) 社 “I - T E X”

- ・ 1次元ラインセンサ
- ・ 欠陥イメージファイルでレビュー

(3) 結言

検反作業の自動化は、永年にわたり多くの研究者が取り組んできたにもかかわらず、実用化されていない。

これまでの完全自動化という目標ではなく、検査員の目視作業の軽減という観点から、画像情報を音響情報に変換して、検査員の聴覚も利用するといったアプローチが期待される。

参考文献・資料

編地 (パンスト) 等級見本

検反機カタログ

中小企業事業団：“自動縫製加工準備システム” 技術開発資料

E V S (Elbit Vision Systems) 社 “I - T E X” カタログ

2. 5. 5 超音波探傷結果の画像化例

(1) はじめに

最近の電子技術の発展はめざましく、あらゆる工業分野で自動化、映像化及びシステム化が幅広く行われている。非破壊検査の分野においてもこれらの技術を利用した方法が数多く開発されてきており、特に、超音波探傷試験においては、従来から放射線透過試験と比較して問題とされてきた試験結果の記録性、再現性などを改善し、また解釈が容易な記録表示を目的として、試験結果を画像表示する方法が検討されてきている。

超音波探傷試験における画像化の目的は、きずなどの試験体内部の状況を表示器上に忠実に再構成することであり、出来る限り解釈が容易で記録性の良いものが要求される。通常の基本表示（Aスコープ表示）の超音波探傷器の信号を断面表示

（Bスコープ表示）及び平面表示（Cスコープ表示）するものから、これをさらに3次元表示するもの、信号処理や画像処理などを用いて個別性を改良しようとするものなどがある。ここでは、代表的なものについて紹介する。

(2) 従来技術

従来の超音波探傷は、反射信号高さと、反射源までの距離のみを利用した、材料評価技術であった。その探傷結果の表示方法としては、以下の3通りのものがあった。

イ. Aスコープ表示

超音波探傷法における最も基本的な探傷法で、縦軸に超音波反射量、横軸に伝播時間（伝播距離）をとった表示法である。反射体迄の距離、超音波反射量（受信信号高さ）から反射源の大小の推定が出来る。但し、一探傷点の結果しか表示できず、被検材の連続的な変化が記録できない（図2—10）。

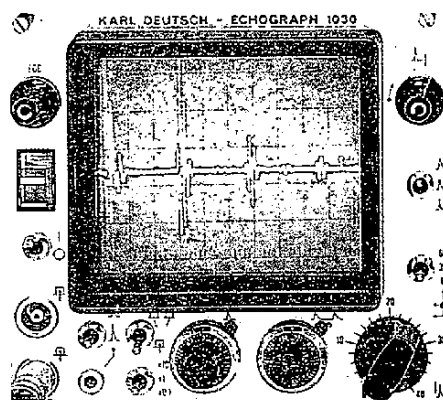


図2—10 超音波Aスコープ表示例

ロ. Bスコープ表示

超音波探触子を被検材上で走査した時に、超音波探触子位置が少しずつ移動する毎に、超音波反射源迄の距離と探傷点の位置を直角座標で表示する事で、被検材断面内の欠陥分布を直感的に表示した方法。反射源までの距離（深さ）の分布が分かる（図2—11）。

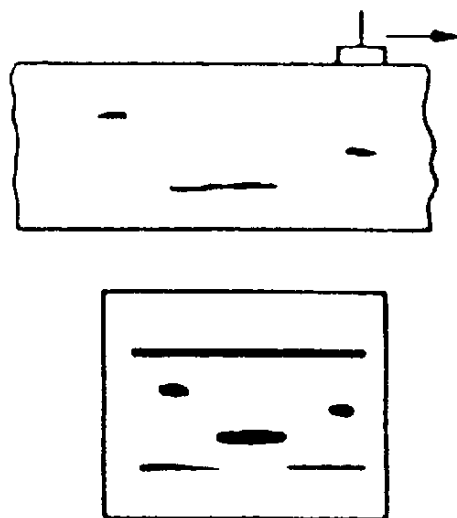


図2—11 超音波Bスコープ表示例

ハ. Cスコープ表示

超音波探触子を被検材上で走査した時に、超音波探触子位置が少しずつ移動する毎に、超音波反射源の有無と探傷点の位置を直角座標で表示する事で、被検材中の平面的な分布を直感的に表示した方法。板等では欠陥の分布が直感的に分かりやすいが、反射源までの距離、反射量が分からない（図2—12）。

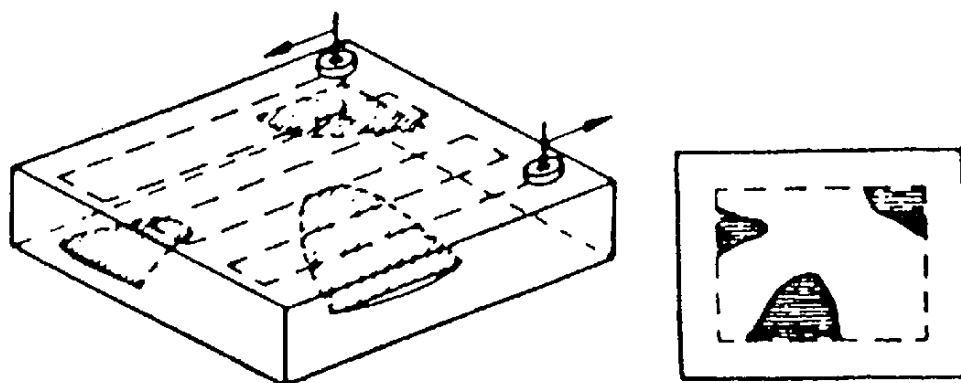


図2—12 超音波Cスコープ表示例

(3) 最新技術

デジタル信号処理技術の進歩により、各探傷点毎の反射信号波形をそのまま、CPUのメモリ内に取り込みながら、探傷する事が可能になった。その結果、受信信号波形そのものを情報として使える事が出来、従来の超音波反射量、伝播時間の情報のみならず、各種波形形状特徴（立ち上がり時間、立ち下がり時間、持続時間等）、周波数特徴（中心周波数、周波数幅、最大強度周波数）、位相特徴を利用する事が可能になり、それらの特徴量を利用したBスコープ表示、Cスコープ表示が可能となっている。また、これらの情報を組み合わせて反射源の判定も出来るようになってきている。

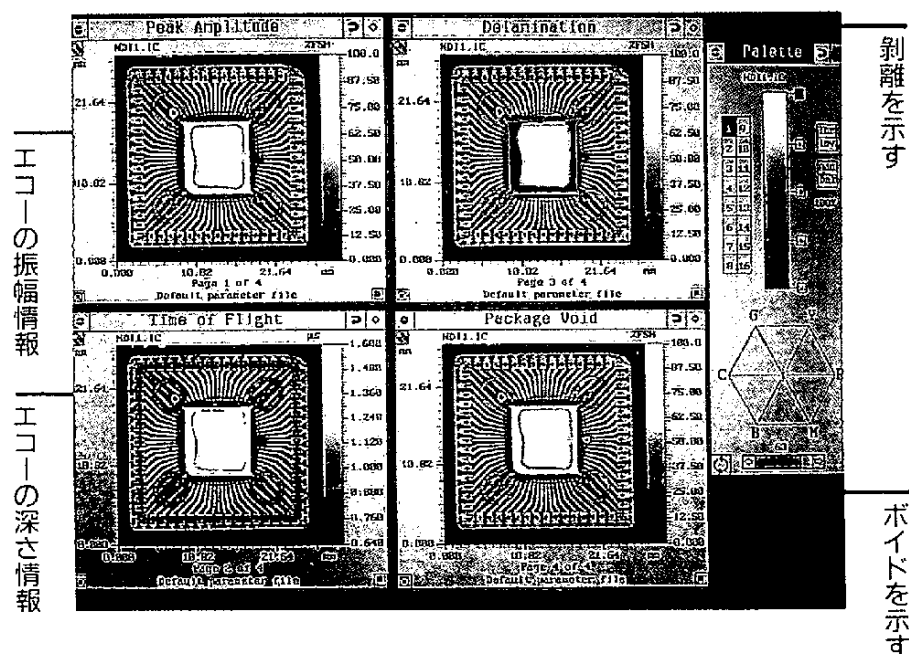


図 2—1 3 高機能超音波信号処理例

(4) 結言

超音波探傷結果の画像化装置が最近急速に発展してきている。但し、これら装置を実際の対象物への適用はこれからの課題である。このためには、実際の対象物との正しい対応、現場への適応性、使いやすさを充分考慮する必要がある。また各種の材料中のきずや異常部と超音波信号情報の相互比較や蓄積も必要になってくると思われる。

将来への展望として、超音波探傷結果を視覚化した高機能情報を用いた高性能探傷（高感度、ダメージ度判定等）が進むと思われる。さらに音響情報等を多次元的に組み合わせた新しい計測技術の発展も期待される。

第3章 メディア変換の要素技術の現状

本章では、メディア変換の要素技術の例として、画像を信号処理的に直接音に変換する手法の詳細と、物理モデルからの画像と音の同時生成法のコンセプトを解説する。

3.1 画像から音響への直接変換

3.1.1 リングオペレータの利用による濃淡画像からの音の生成

(1) リングオペレータの原理

画像内の各点は、その点を含む近傍の画像の性質によって特徴づけられる。差分型空間オペレータとモデルフィット法によるエッジ抽出は、注目点の近傍にマスクを設定し、その中の2次元配列データの形状を判定することにより実現される。マスクとしてリング上のものを設定し、そのリングに沿ったデータを1次元周期波形と考えてスペクトル分析することにより、近傍画像の特徴を求めることができる。

さて、濃淡画像においては、局所データを、ステップエッジ、ルーフエッジ、ライン、均一濃度、変化濃度の5種類に分類することができる。これらについて、上記のリングオペレータによるスペクトル分析を施すと、それぞれ異なった特徴を持った結果が得られる。したがって、これらを利用することにより、画像のエッジとラインを検出することが可能になる。

(2) 音の生成方法と傾向

このリングオペレータを閉曲線に沿って移動させることを考える。ここでは一つの例として、「8」曲線をリングオペレータの移動経路として採用した場合の処理を紹介する。

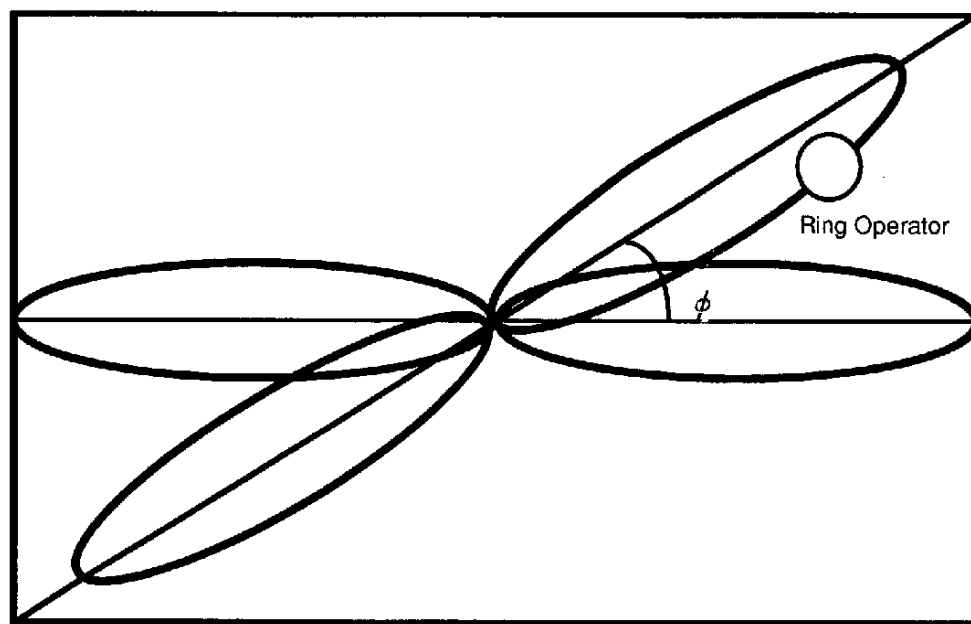


図3-1 リングオペレータ

まず、「8」曲線の中心点を画像の中心点にあわせて、反時計方向に少し回転させたものを考える。そして、リングの中心点をこの曲線上に置き、曲線に沿って反時計方向に移動しながら、リングによって曲線上の点のまわりのデータを取り込んで、それをフーリエ変換した結果を音の周波数と定める。この場合、一つのリングに関するデータが一つの周期信号を生成する。また、閉曲線の長さを S 、リングの移動距離（サンプリング周期）を L とすると、曲線を1周することにより S/L 個のリングの計算結果が得られる。

このようにして生成された音は、画像の変化により振幅が変化する。すなわち、画像の内容が変化しないところでは音の振幅が0になり、変化が大きいところでは振幅も大きくなる。周波数についても、リング上のデータの変化に応じてパターンが変化する。また、 L が大きくなると周波数の変化も激しくなり、 L が小さくなれば周波数の変化も緩やかになる。

このように、リングオペレータを用いて濃淡画像を音に変換することができる。その結果には濃淡値の変化部分が反映され、音によって動的な印象が与えられる。

3. 1. 2 二値図形の音響化

(1) 二値図形の時系列化

イ. $P1(\theta)$ 曲線表現法

$P(\theta)$ 曲線表現法は図形の時系列化手法の一つである。この方法は、まずあらかじめ求められている図形の境界（閉曲線）について、その閉曲線内に適当な一点 O とそれを原点とする座標系を設定する。そして、曲線上の任意の点と X 軸のなす角を θ とし、 O 点からその点までの距離を $P(\theta)$ で表す。 O としては、任意性を避けるために一般にその図形の重心を選択する場合が多い。このようにして得られた関数 $P(\theta)$ が、画像の時系列化の結果となる。

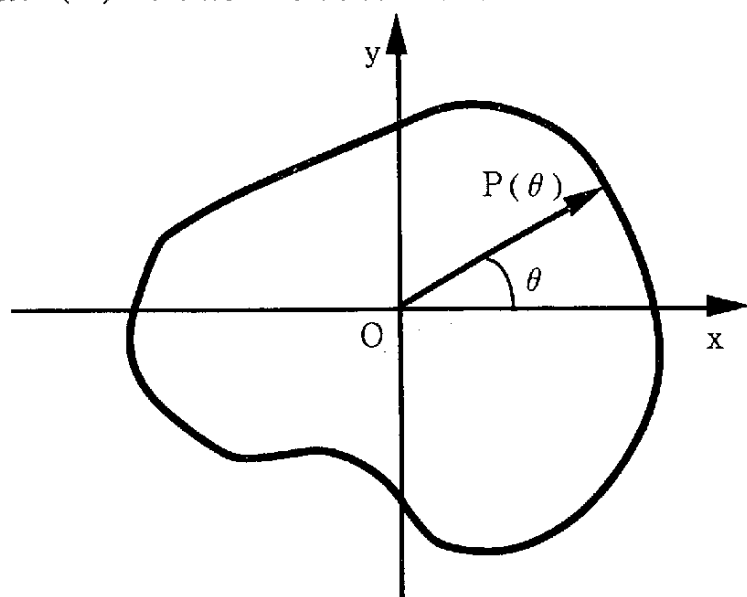


図 3 - 2 $P1(\theta)$ 曲線表現法

この方法の特徴は、以下の2つが挙げられる。

- 1) 結果としてでてくる $P(\theta)$ は、周期 2π の周期関数である。
- 2) 同じく $P(\theta)$ は点と曲線の距離を表しており、非常に直感的である。

ただし、この方法には一定の制限がある。すなわち、

- 1) 閉曲線に対してのみ適用できる。
- 2) 点 O から曲線上のすべての点が見えていること。

である。

ロ. $P2(\theta)$ 曲線表現法

この方法は $P1(\theta)$ 曲線表現法を一般化したもので、イ. において点 O の取り方を図形の内部に限定しない方法である。ただしその場合には、 P の周期は一般には 2π ではない。

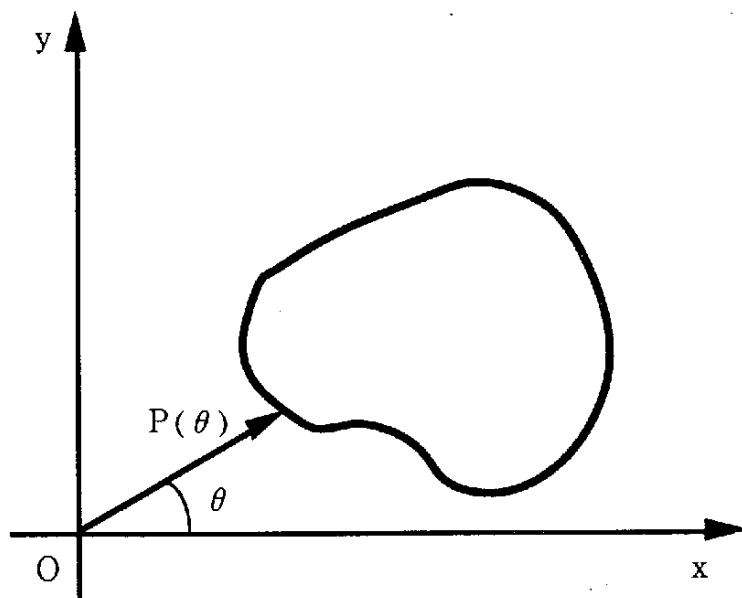


図 3-3 $P2(\theta)$ 曲線表現法

この方法では点 O を任意に選び、イ. と同様にして曲線上の任意の点と点 O の間の距離を $P(x,y)$ で表すが、この P の周期は境界線の長さによって変化するため、周期の確定が困難である。また、図形の大きさが一定の時、図形と点 O の距離が大きくなれば θ の変化は小さくなり、逆に距離が小さくなれば θ の変化は大きくなる。これを利用すれば、音信号の周波数の変調を行うことができる。

ハ. フーリエ記述子表現法

平面上の連続曲線 C (必ずしも閉曲線とはかぎらない) を、離散化して複素平面上で全曲率関数を用いて表現し (これを $W(i)$ とおく)、それをフーリエ変換したものを $M(k)$ とする。この $M(k)$ を、曲線 C の P 型フーリエ記述子と呼ぶ。

この記述子には、以下に挙げたような特徴がある。

- 1) 不変性・・・曲線の平行移動や拡大・縮小に対して不変である。したがって互いに相似な曲線の記述子は区別できない。
- 2) 安定性・・・記述子からの再生曲線は元の曲線と一致する。
- 3) 処理の中間結果 $W(i)$ が、周期 2π の周期関数である。

このような特徴により、ある画像中の同じ形状の図形に対しては同じ音が生成され、それぞれの形状について生成された一次元信号（正弦波）を連結することにより、その画像全体を表現する複合音を得られる。

(2) 図形時系列化信号の音響化

イ. 音の生成ルール

音を生成する基本は、周期的な一次元信号の生成である。しかしながら、ただ単に図形を一次元信号化しただけの音では、あまり意味があるとは言えない。そこで、まずは聴覚特性および視覚特性にもとづいたルールを設定することを考える。

a. 聴覚特性にもとづくルール

我々の耳で認識できるのは、一定範囲の周波数の音だけである。また、音の大きさなどにも制限があるので、音を生成する際には、これらを考慮しなければならない。

可聴範囲：生成された音信号の周波数は、人間の可聴周波数の範囲に入っていないといけない。理論的には、一般には20Hz～20kHzがその範囲であると言われているが、これには個人差があり、その人の年齢や性別、あるいは対象となっている音の持続時間などによっても異なる。

音の高さ：これは信号の周波数によって決まるものであるが、音の強度などとも関係がある。一般に、中音域では強度による影響は少ないが、低音域および高音域ではかなりの影響がある。したがって、生成する音の周波数は中音域になるようにすることが好ましい。

音色：音色は音の特徴の中でも非常に重要な要素の一つであり、スペクトル・位相・持続時間・強度などによって決定される。音色を表現する際には、周波数・振幅の他に、音の立ち上がりも検討すべきである。

以上のような議論にもとづき、一つの指針として、生成する音の基本周波数を50Hz～5kHzとしておく。下限については、小さすぎると人が音をあまり感じないであろうという理由から、また上限については、振幅変調による音の生成を考えた場合、5kHzを超えたものについては Low pitch が知覚されなくなるなどの理由から決定している。

b. 視覚特性にもとづくルール

音は時間的な一次元信号で、スタートポイントとエンドポイントを有し、かつ逐次感と持続感を持っている。それゆえ情報の表現も、時間的・順番的かつ持続的に行われる。それに対して、画像はコンピュータの中では二次元信号の形で表現され、時間性もなく、見るものに対し全体のイメージ情報が瞬時に与えられる。いいかえれば、情報信号的には、音は動的信号であるのに対し、画像は静的なものであると見なすことができる。したがって画像を音に変換することは、静的な二次元信号を動的な一次元信号に変換することと同じであると言える。

このような観点からすると、変換した音の流れ時間と逐次感や持続感が、元の画像とどのように対応づけられるかが、変換の際の一つの大きなポイントとなる。このような対応関係は、現在のテレビの走査方式でもって実現することが可能ではあるが、それでは画像全体のイメージを表現しにくく、局在な所に注目し過ぎるきらいがある。また、画像によっては、そのような方法では特徴が時間的に表現できない場合も多々ある。そこで、このような問題を解決するための手法を、以下で述べる。

ここで注目している点は、画像全体の複雑さである。日常において、我々が一枚の画像を見るときにまず注目するのは、その画像の中にある特徴的な要素である。普通は、大きなもの・特別な形をしたもの・色の鮮やかなものなどの刺激の著しい内容を先に認識し、それに伴い位置などの情報を認識する。そして、画像中の図形の種類や数が多いほど、全体を認識するのに要する時間は長くなる。したがって、音の基本的なパラメータを決定する際には、このような人の画像認識特性を反映させるものとする（図3-4）。

音全体の流れ時間 t_l : 上述のように画像全体の認識時間が画像の複雑さと比例することから、音の流れ時間も画像の複雑さによって決定する。具体的には、一つの図形に一つの刺激音を対応させるものとして、

各刺激音の持続時間の総和 + 刺激音間の間隔の総和と
いう式によって決定する。

刺激音の表現順序 : 刺激音の順番は刺激強度によって決定する。すなわち、刺激の強いものから順次表現する。具体的には、まず面

積の大きいものを先に、小さいものを後とする。面積が同じ場合は、形状の簡単なものを先、複雑なものを後とする。さらに、面積・形状ともに同じ場合には、各図形の重心を求め、画像の中心点から各重心までの距離の近い順番に表現することとする。

各刺激音間の間隔 t_2 : 各刺激音の表現順序が決定した後、順序が隣り合っている2つの図形の重心間距離に応じて各間隔を決定する。

各刺激音の振幅 A : これらのパラメータの決定については、各図形の面積・周波数・持続時間 t_3 周囲長などに対応付けを行う。

振幅の変化パターン : 振幅の変化パターン（中でも立ち上がりの部分）は、音色に大きな影響を与える。ここでは、J.C.RissetとM.V.Mathewsの実験による近似線を利用している。

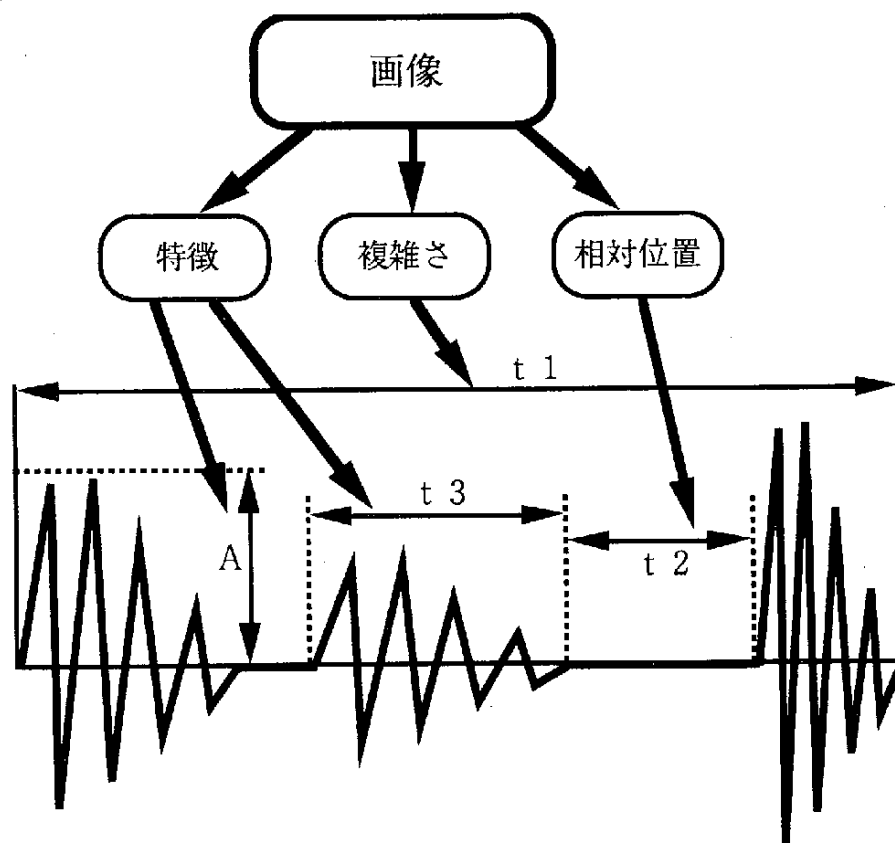


図3-4 音の基本的なパラメータの決定

ロ. P1 曲線法による多数任意図形の音表現

P1 曲線法の結果は周期 2π の周期関数となるので、そこから直流成分を取り除くことにより、そのまま一つの刺激信号の波形として出力できる。刺激信号の持続時間 t_3 は、図形の面積により決定される。

ハ. フーリエ記述子による多数任意図形の音表現

この方法では、各図形のフーリエ記述子 $M(k)$ を音の周波数とし、音の持続時間 t_3 は境界線の長さによって、また振幅 A は面積によって、それぞれ決定される。

二. 単一の図形に対する音の生成

a. 振幅変調法

変調の方法として、2つのやり方がある。1つ目の方法は、P 曲線法による点 O と境界線との距離データを用いて音の振幅を変化させ、同時に同じデータに対してフーリエ変換を行って得られた結果のエネルギーの集中部分を、出力する音信号の周波数とするものである。こうすることによって、曲線の中で特徴的な部分の情報が、生成された音の周波数に反映され、境界線の変化の様子が音の振幅に反映される。また、境界線が複雑になれば、音の周波数も多様化する。2つ目の方法は、フーリエ記述子法の間中結果 $W(i)$ にもとづいて正弦波の振幅を変化させ、フーリエ記述子 $M(k)$ を出力する音信号の周波数とするものである。この方法では、曲線の曲率の変化の様子が周波数に反映される。また、境界線が複雑になれば、周波数および振幅の変化は大きくなる。なお、フーリエ記述子法の原理により、曲線が変化しないところでは生成された音信号の振幅は一定になる。

b. 周波数変調法

P2 曲線法の結果にもとづき、正弦波の振幅を $P2(\theta)$ により変化させる。この際、 θ の変化幅が 2π より小さいので、これを周波数の調整量として変調を行う。具体的には、曲線に囲まれた部分の面積を求めて、それを周波数域に対応づけ、その結果に θ を加えたものを音信号の周波数とする。これにより、図形の面積が音の周波数に反映され、図形的位置が音の振幅に反映される。

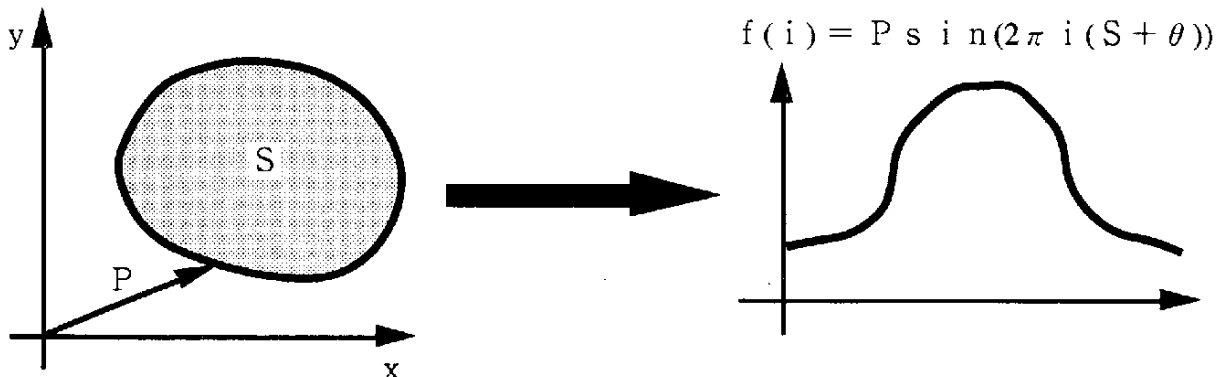


図3-5 周波数変調法

3. 2 物理モデルからの映像と音の同時生成

3. 2. 1 基本概念

CGでシミュレーションが行われ始めた頃は、単純な映像が画面上で動くというだけで画期的なものであった。その後、CGの技術は、三次元画像に陰影処理までどこまで動かせるようになるなど急速な発達を遂げ、現在では、実写と見紛うほどリアルかつ高速な動画像を作ることが可能となっている。一方、そのようなシミュレーションにおける映像技術の進歩に比べて、音の方の扱いは、単なるシステムの付加機能にとどまっている場合が多い。すなわち、映像は、人間の操作に対応してさまざまに変化するが、音の方は、あらかじめ用意されたサンプリング音が、映像に合わせたタイミングで再生されるに過ぎないのである。

このようなシステムを、より臨場感があり、我々にとって自然な感覚にそったものにするには、視覚に対して映像の変化が強い刺激を与えているのと同じように、聴覚に対しても音の変化によって強い刺激を与えてやる必要がある。

そこで、中間表現（具体的には物理モデル）を操作することにより、画像と音を同時に生成・変化させるシステムを考案する（図3-6）。

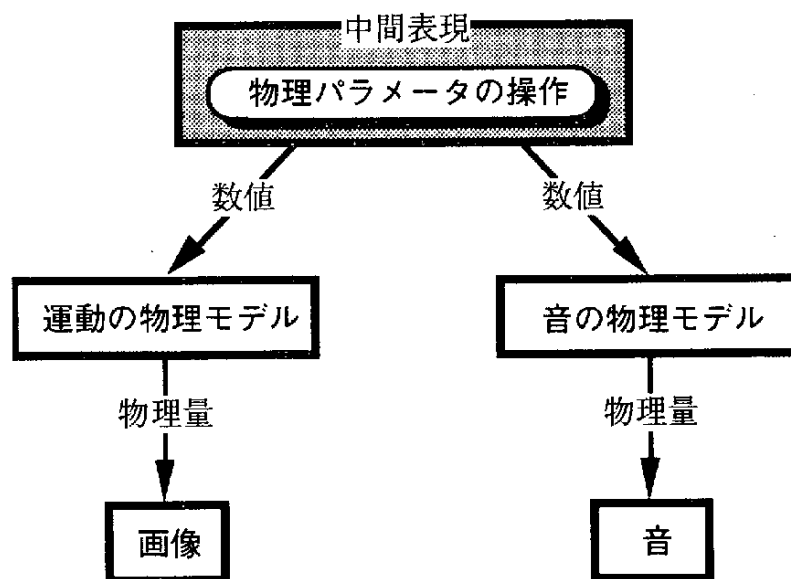


図3-6 画像と音の同時生成

3. 2. 2 処理の流れ

このシステムは、中間表現である物理モデルがコアとして構築される。したがって、すべての入力はず、まず最初に物理モデルに入り、そしてその中で計算された結果が画像生成部と音生成部に送られて、最終的に画像と音が出力される、という経路をたどる。ここで入力は、システム内に設定された仮想空間の状態を変化さ

せるものであり、このとき各状態に割り当てられる値を環境パラメータと呼ぶことにする。入力形態としては、主にマウスやキーボードからの入力が考えられるが、他にもマイクやMIDIあるいはライン入力も可能とする。

さて、処理全体の流れの中で重要なことは、同一のパラメータによって画像と音が同時に変化するという点である。したがって、物理モデルから画像と音がそれぞれ生成される過程においては、2つの処理が並列に行われなければならない。このような処理の流れを図3-7に示す。

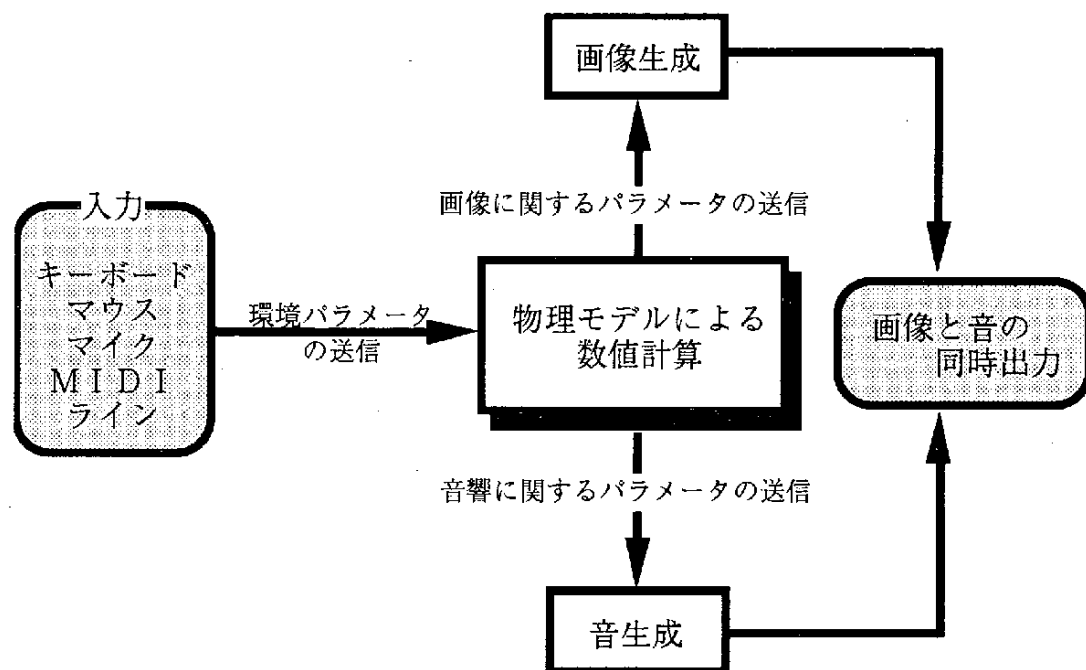


図3-7 処理の流れ

3. 2. 3 画像と音の生成

(1) 画像生成部

入力として与えられる環境パラメータは、すべて画像内の物体の運動に関与するものであるから、それにもとづいて運動方程式を解けば、対象物体の運動の様子を容易に求めることができる（図3-8）。

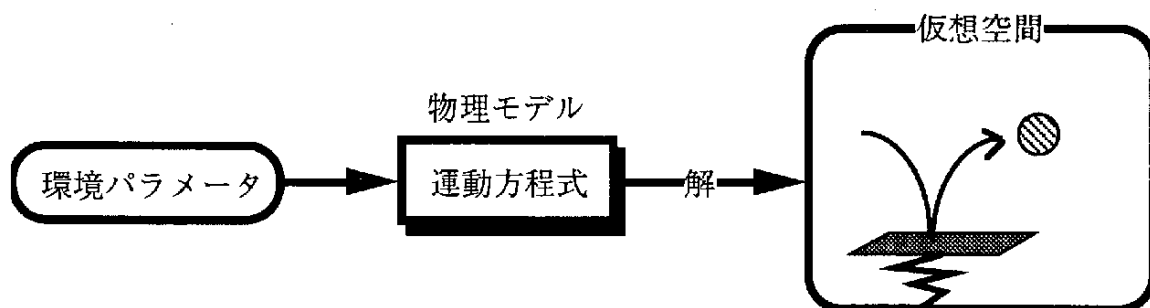


図3-8 画像の生成

(2) 音生成部

画像の場合と異なり、音の変化については、運動方程式のような確立した関係式が存在しないため、環境パラメータだけで音を生成するのは非常に困難である。そこで本システムでは、基本となるサンプリング音をデータベース上に用意し、それを環境パラメータのさまざまな変化に応じて変形させて出力する（図3-9）。

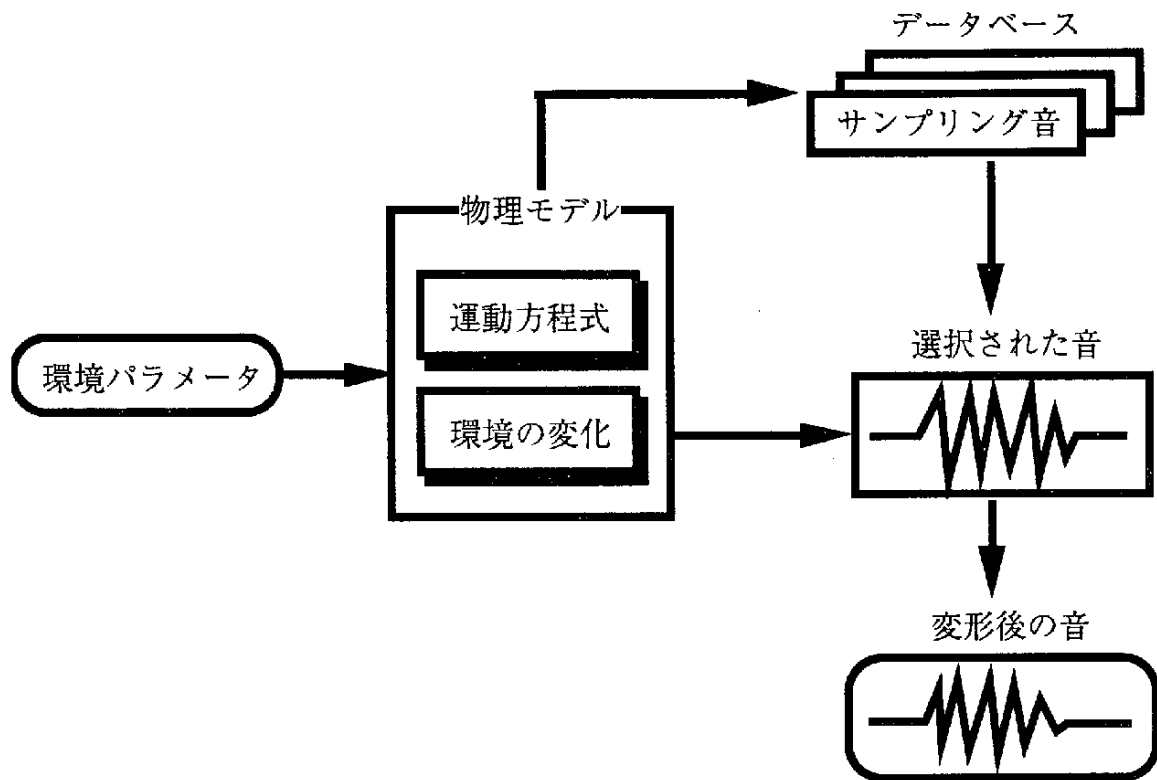


図3-9 音の生成

参考文献

- 松田・井口：「リングオペレータによるエッジとラインの抽出」，電子情報通信学会論文誌，Vol.J66-D，No.10 (1983)
- 松田・井口：「リングオペレータによる距離画像のラベリング」，電子情報通信学会論文誌，Vol.J66-D，No.10 (1983)
- 上坂吉則：「閉曲線にも適用できる新しいフーリエ記述子」，電子情報通信学会論文誌，Vol.J67-A，No.3 (1984)
- 井口征士：「画像と音響—信号処理の多様性」，システムと制御，Vol.31，No.5 (1987)
- 永野・梶真・森晃：視覚系の情報処理，啓学出版
- 和田・大山・今井：感覚+知覚，誠信書房
- 高木・島脇・田村：画像処理アルゴリズムの最新動向
- 田村秀行：コンピュータ画像処理入門，総研出版

第4章 メディア変換型感性データベースのプロトタイプの概要

本章では、今回の研究で提唱するメディア変換型感性データベースを構成するための要素となる、4つのデータベースないしは検索システムのプロトタイプについて、その内容を紹介する。

4. 1 イメージにもとづくイラスト画の検索システム

本システムでは、特定作者（永田萌）の作品に限定した作者心理をキーとしたイラスト画のイメージの解析および検索を行なう。データベース化されたイラスト画からは、絵画の重要な表現手段となる色とシンボルに関する特徴量が計算される。その特徴量から推測される作者心理と実際の作者心理との一致度から、作者心理を推定するのに有効な特徴量を見い出し、効率的なイメージ検索を行う。

4. 1. 1 システムの特徴と位置づけ

(1) 計算機による絵画表現の評価

表現方法の評価は、従来、芸術の評論家が、作品から作者の印象を読み取るために行ってきたことである。評論家は、さまざまな事物に対する広く深い経験や知識を持ち、絵画を観る目を持っている。しかし、表現方法が言葉で述べられることが多く、曖昧な点がある。

そこで、この表現方法の評価を計算機を用いて数量化することが考えられる。以下の前提条件を満たせば、計算機処理による評価の数量化が有効な手法となると考えられる。

- イ. 特定の作者の作品に限定すること。
- ロ. 印象、表現方法の変動の少ない作品群であること。
- ハ. 作者心理と表現方法が明示的に表現されうること。
- ニ. 大量の作品を対象にして評価を行うこと。
- ホ. 深い推論を必要としない作品群であること。

イ、ロは、人、計算機処理どちらの評価の場合も前提となる条件である（図4-1）。ハ、ニ、ホは、計算機処理による評価の条件である。イの条件により、事物に対する印象が一人の個人のものに定まり、その伝達手段である表現方法もその人固有のものに限定される。ロは、同一作者の事物に対する印象、表現方法の中であっても、それらが作品ごとに極端に異なることを避けるための条件である。ハは、計算機で数値処理が可能となるための条件である。計算機は記号しか処理をすることができないため、その事物と表現方法は記号化されなければならない。ニは、多量のデータを用いた統計処理が可能となるための条件である。多量のデータ処理を

得意とする計算機では、多数のサンプルの事物と表現の関係から、その全体的な傾向を知ることが可能となる。ホは、計算機上に記述された記号では推論できないような深い印象が、作品に込められていないための条件である。人間の深い知識と経験に基づいた、時には矛盾を含んでしまうような記号どうしの関係を求める処理は、計算機にはそぐわない。また、この条件は、ハとも関連性がある。

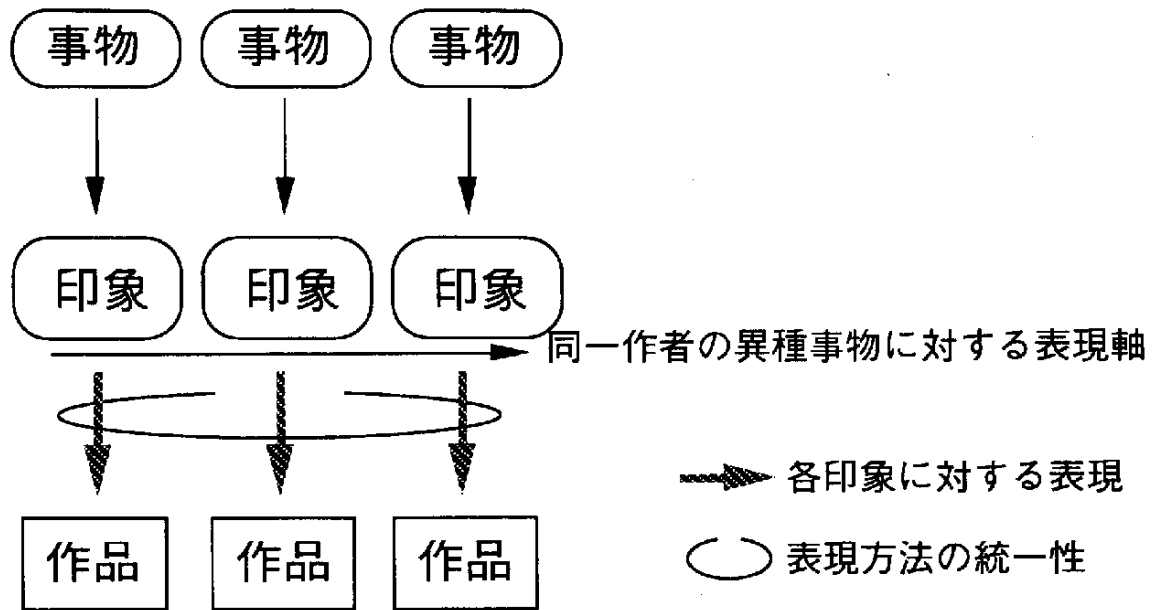


図4-1 同一作者による表現の統一性

(2) 本システムの概観

本システムは、イメージに基づく検索のほか、さまざまな解析用の機能も備わった総合的なシステムである。

本システムの使用目的は、効率のよい素早い検索にあり、本システムは、利用者が目的のイラスト画を探す際の手助けとならなければならない。ここで、本システムで可能な検索方法を述べる。本システムではあらかじめ全ての作品の表現内容が、選定語と呼ぶキーワードの形で登録されているので、作品の表現内容及びそのイメージをキーとしたキーワード検索が可能である。また、あらかじめ全てのイラスト画を観察し、主観的にシンボルを判断し、その名称を登録しているため、シンボル検索が可能である。また、本などの作品画集の名前でも検索を行なうことができる。

検索時の機能として、素早いブラウジング機能が挙げられる。絵のブラウジングは、実際に本をめくる感覚で画像などのデータを素早く閲覧できる機能で、多数の対象を扱う場合の効率のよい検索に必要な操作といえる。本研究で構築した検索システムは、この機能を実現している。また、本システムは、目的のイラスト画が見つかったのち、その場で詳細な観賞が可能となるように配慮している。大小の画像データを備えていて、ブラウジングには、小さい画像データを、観賞時には、大きい画像データを使用する(図4-2)。

また、このシステムでは、人が計算機と対話しやすいような配慮がなされている。人と計算機の対話においても、人間の会話のように、相手との確認の作業が必要となるが、その機能が備わっている。常に利用者に対してシステムの状態を示す。検索途中でなければ、利用者の指示をいつでも受けることができる。

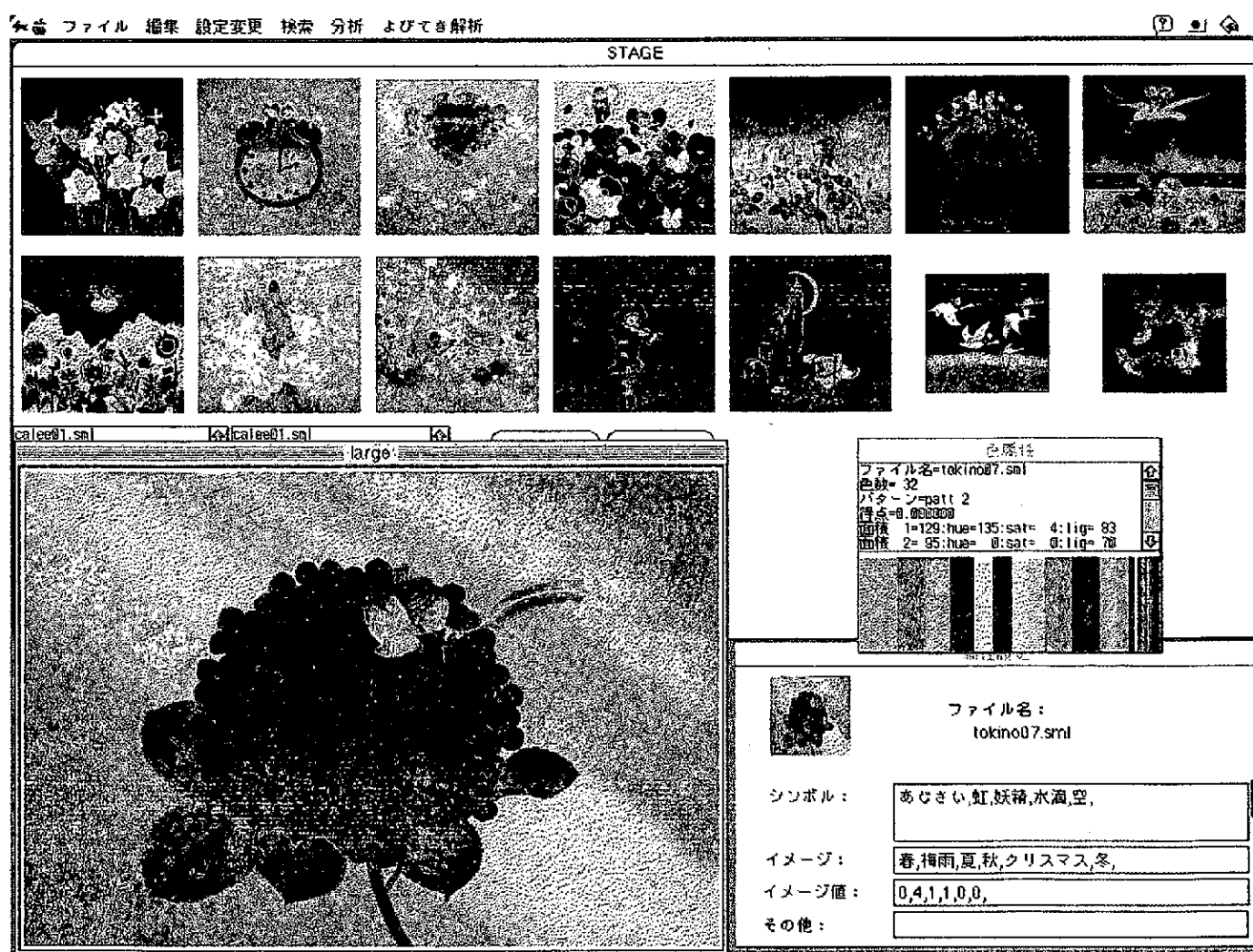


図 4 - 2 システム概観

4. 1. 2 作者イメージと物理的特徴

(1) イラスト画のイメージ

本システムでは、作品の表現内容を推定するための心理実験を行ない、その結果を作品の表現内容とした。ここでは、作者の表現内容は、その観賞者に理解されると仮定する。観賞者は、作品を観賞し、その表現内容を推定し、それに当てはまる言葉を選択する。ただし、観賞者は、あらかじめ各作品の題名を抽象するよう選んだ言葉に対する選定のみ行なうこととした。

以下に、ここで用いた、表現内容の選定対象となる言葉を示す。

選定語

(季節に関する言葉) : 春、梅雨、夏、秋、クリスマス、冬

観賞者は、この言葉の内容をその作品から読み取れる場合と読み取れない場合の二つに判別する。この心理実験を複数の人間に対して行ない、その平均を、その選定語の表現内容の程度を表わす数値とする。

実験においては、この言葉の内容を読み取った人の人数の全体に対する割合を求め、それを作者イメージ値とする。

$$\text{作者イメージ値： } M = \frac{(\text{選定語が適当であると判定した被験者の数})}{(\text{全被験者の数})}$$

(2) イラスト画の物理的特徴量

作者は、イラストを描く際に、絵画での基本的な表現要素を用い、効果的にそれらをアレンジする技法を用いて自分のイメージを表現する。ここで言う表現要素とは、色や材料などキャンバス上にあらわれる物理的な特徴量であり、技法とは、表現要素を使ったデザインの知識を基とした効果的な表現方法である。

ここでは、絵画での基本的な表現要素の物理的特徴量を基本特徴量と呼ぶ。また、その特徴を使ってデザインの知識に基づいた技法を表わす特徴量をデザイン特徴量と呼ぶ。概念図を図4-3に示す。

本システムにおいては、基本特徴量としては、色特徴、シンボル特徴の二つの特徴を用い、デザイン特徴量としては、基本特徴量を使って計算される色相分布特徴量、色調分布特徴量およびシンボル特徴量を用いることとした。ここで、色特徴とはキーカラーの集合のことであり、シンボル特徴とはイラスト画のなかに現われる個々のシンボルに関する名称と幾何特徴を指す。また、色相分布特徴量とは、キーカラーの彩度と色相および領域面積を使って計算される、色相の分布特徴を示すベクトル量であり、色調分布特徴量は、キーカラーの明度と彩度および領域面積を使って計算される、色調の分布特徴を示すベクトル量である。最後のシンボル特徴量とは、イラスト画に使われている材料の一般名称群のことである。

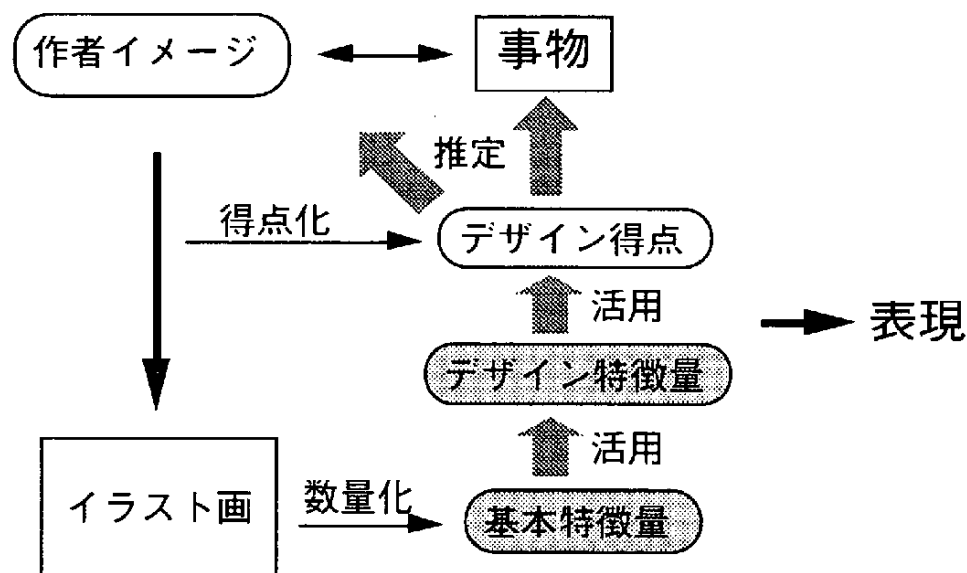


図4-3 基本特徴量とデザイン特徴量

4. 1. 3 検索方法

ここでは、本システムでの検索方法の処理の流れについて説明する。図4-4にその概念図を示す。

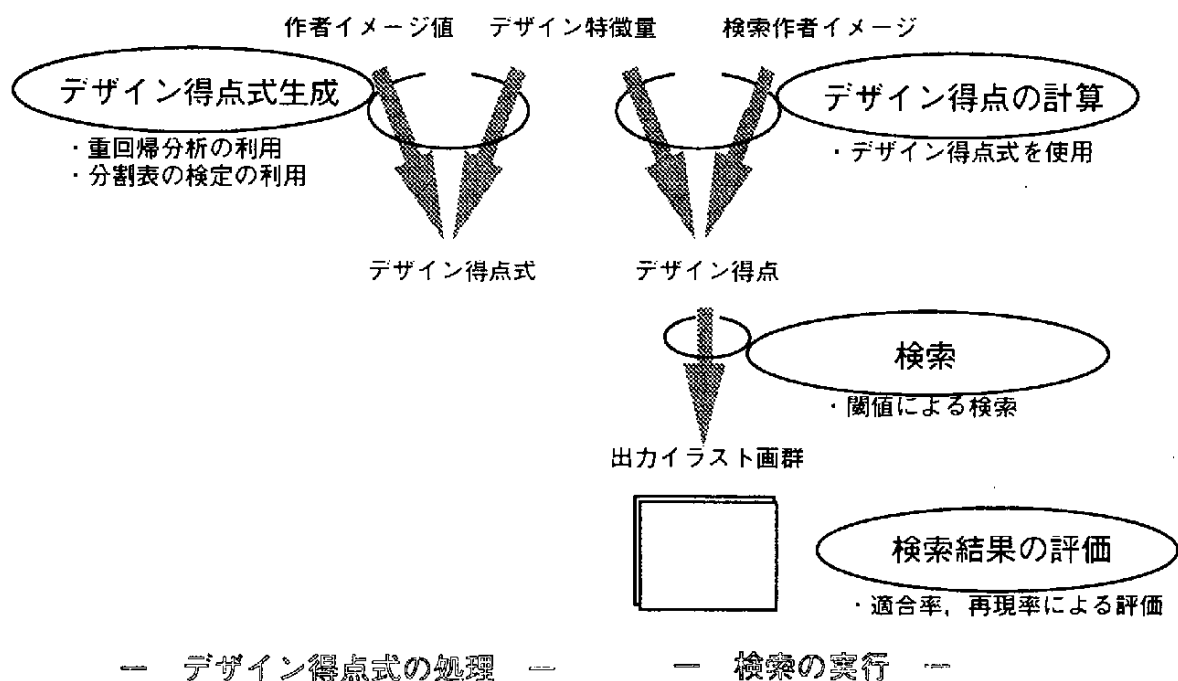


図4-4 検索に関する処理の流れ

デザイン得点式の処理では、デザイン特徴量から作者イメージに対するデザイン得点を計算する。デザイン得点には、色相分布特徴量から計算される色相得点と、色調分布特徴量から計算される色調得点およびシンボル特徴量から計算されるシンボル得点がある。それぞれの得点は、それぞれのデザイン特徴量から重回帰分析や分割表の検定を利用した方法を用いて計算される数値を、異なる手法で求めた得点と比較することもできるように、平均 0、標準偏差 1 に正規化したもので、作者イメージに近いほど得点が高くなる。

検索の実行では、その得点がある閾値以上であるイラスト画を選んでくる。検索したいイラスト画の作者イメージを計算機に入力すると、その作者イメージに対する得点が全イラスト画に対して計算され、得点が高いイラスト画順に表示される。ここでは、一つのデザイン得点を単独に用いる検索のほかに、複数のデザイン得点を組み合わせて用いる複合検索も行う。後述の実験においては、以下の 5 種類の検索実験を行っている。

- (検索手法 a) 色調分布特徴量を用いた検索
- (検索手法 b) 色相分布特徴量を用いた検索
- (検索手法 c) シンボル特徴量を用いた検索
- (検索手法 d) 色相分布特徴量とシンボル特徴量を用いた閾値検索
- (検索手法 e) 全てのデザイン特徴量を用いた閾値検索

なお、得点の閾値も予め決定しておく必要がある。そこで、ある得点を閾値としてイラスト画群を二つに分けたときの誤判別率を全ての場合について求め、誤判別率が最も小さくなる値を閾値として用いることとする。

また、実験結果の評価には、一般の閾値検索においてよく用いられる評価尺度である適合率、再現率を用いる。ここで、適合率とは、検索されたもののうち作者イメージに適合するイラスト画が何%含まれているかを示す尺度である。再現率は、適合するイラスト画のうち検索されたイラスト画が何%含まれているかを示す尺度である。この評価尺度によって、イラスト画の検索が、一つのデザイン得点を用いた場合と複数のデザイン得点を用いた場合ではどちらがどう有効であるかという判断が可能になる。また、どのデザイン得点を用いれば、効率良く検索することができるのかといったことの判断も可能である。

4. 1. 4 実験

(1) 実験環境

まず、今回の検索実験の対象となるイラスト画について述べる。対象は、特定作者が描いたイラスト画で、以下の 18 冊のイラスト画集から 501 枚を選んでいる。

1989年カレンダー、1990年カレンダー、1991年カレンダー、1992年カレンダー、花待月に、ILLUSTRATION GALLERY、雲から夢まで、

ITSUKA、京都夢見小路、コスモスパーク、時の果樹園、ラング・ド・シャ、
またゆら、不思議の森へ、うみいろのバケツ、幻想旅行、妖精村の夢工場、
永田萌の世界

以上の画集のイラスト画から、イメージスキャナ（GT-6000：エプソン製、解像度は300 DPI）を使ってデジタル画像データを生成した。その画像データは、RGB値で各色 8 bit に量子化されたものとして記録される。画像データのサイズは任意であるが、およそ縦横サイズが、400画素から600画素程度である。

前処理および検索処理、そのヒューマンインタフェース部の構築には、計算機（Macintosh Quadra 700：アップル社）を使用した。生成された画像データは、データ記録用光磁気ディスク上に記録され、ワークステーションを使って基本特徴量が計算される。

（2）実験結果

前項で示した検索手法 a ～ e による検索結果の一例を図 4 - 5 に示す。またこれらの結果の評価用に適合率と再現率の一覧表を表 4 - 1 に示す。表中の記号の意味は以下のとおりである。

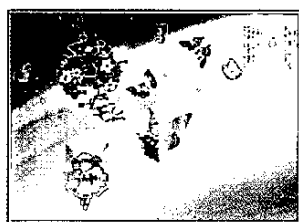
- [tone] : 検索手法 a を示す（色調を用いる）。
- [hue] : 検索手法 b を示す（色相を用いる）。
- [sym] : 検索手法 c を示す（シンボルを用いる）。
- [hue,sym] : 検索手法 d を示す（色相, シンボルを用いる）。
- [all] : 検索手法 e を示す（全てのデザイン特徴量を用いる）。

表 4 - 1 検索評価値（適合率と再現率）の一覧表

選定語	tone		hue		sym		hue,sym		all	
	適合率	再現率	適合率	再現率	適合率	再現率	適合率	再現率	適合率	再現率
[春]	0.15	0.03	0.27	0.73	0.40	0.94	0.46	0.91	0.46	0.73
[梅雨]	0.03	1.00	0.03	0.90	0.30	1.00	0.30	1.00	0.40	1.00
[夏]	0.28	0.77	0.39	0.86	0.62	0.80	0.63	0.83	0.48	0.95
[秋]	0.21	0.79	0.36	0.74	0.62	0.92	0.62	0.92	0.63	0.88
[クリスマス]	0.07	1.00	0.03	0.67	0.39	1.00	0.44	1.00	0.54	1.00
[冬]	0.22	0.84	0.21	0.87	0.65	0.84	0.67	0.82	0.49	0.90

[選定語：春]

[選定語：梅雨]



[選定語：夏]

[選定語：秋]

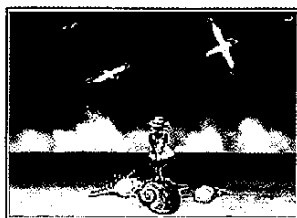


図 4 - 5 シンボル特徴量を用いた場合の検索例

4. 2 感性情報にもとづくイラスト画の自動生成システム

本システムでは、感性情報に基づくイラスト画の自動合成を試みる。これは、ユーザから感性情報として形容詞が入力され、この入力された形容詞に対応する様々なイラスト画がコンピュータ内で自動的に生成される、というものである。具体的には、画像データベースに集積された特定作者のイラスト画群から、感性情報とキーカラー、キャラクタの関係を画像処理手法により抽出し、コンピュータでイラスト画を描くための技法について、描画の技法をいくつか実現した。また、複数の画像に対し縮小・拡大・移動を協調的に行なうことで、構成要素間の相対関係を自動的にアレンジすることを考えた。

4. 2. 1 システムの概観

イラスト画を一目見たとき、我々はさまざまな印象を受けるが、「ロマンチックな」、「秋らしい」、「美しい」、「楽しい」などイメージを表す形容詞で表現することが多い。一方、イメージを表す形容詞からイラスト画で表現する作業は、イラストレータ（作者）の感性によるものである。そこで、コンピュータ上では、イラスト画の配色、描画技法、構成要素及びその配置などを簡単に変えることが可能であることから、人間の代わりにコンピュータでイメージの解析を行ない、自動的にそのイメージを表現するイラスト画を描かせることを試みる。

図4-6に示すように、作者が自分の表現したいイメージをもとにそれぞれの表現法で描いたイラスト画からその作者の抱いたイメージを推定し、そのイメージをコンピュータ内に構築する。一方、作者の描画表現法もモデル化し、得られた合成法に基づいて、イラスト画を自動的に合成する。そして、合成イラスト画と作者イラスト画のイメージを比較することで、合成イラスト画の評価を行なう。

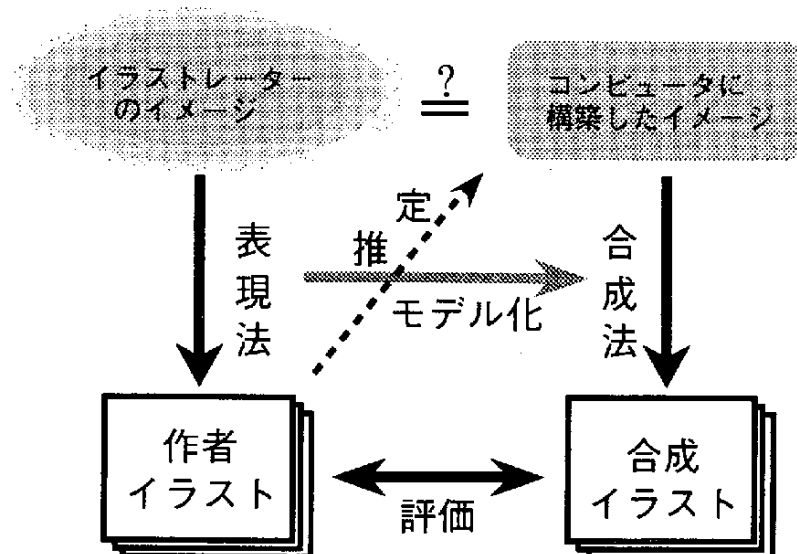


図4-6 システムの概念

4. 2. 2 イメージの推定と解析

本システムは、先に4. 1節で示したシステムの発展型とも言え、先のものが多数のイラストからのイメージ情報の抽出とそれを利用した検索を目的としていたのに対して、ここではその抽出された情報を元に、イラストの構成要素のデータベースからのイラストの合成を行うものである。したがって、イラストのイメージの推定と解析にあたっては、先のシステムで用いたのとはほぼ同様の手法を用いている。

ただし、それらに加えてイラストの合成には構図を決定するための情報が必要であり、これはイラストレータ特有の描画技法によって決まる。ここでは背景の配色とシンボルの配置に注目している。

(1) 背景の配色

配色に関する描画技法として、グラデーションとぼかしに着目した。

グラデーションは対象としたイラスト画の背景によく現われる技法で、背景色を少しずつ滑らかに変化させていくという描画技法である。抽出した情報は色変化の方向、始点とその色情報である。この作者の背景画像には、主に直線的なグラデーション、特有な楕円的なグラデーション、波動的なグラデーションなどの形状が見られた。また、ぼかしというのは、境目の色彩を次第に薄くし、境目をあいまいにする技法のことである。この作者は、平滑化的なぼかし、グラデーション的なぼかし、特有なノイズ的なぼかしをよく使っている。

(2) シンボルの配置

シンボル特徴量を利用して、複数のシンボル同士の組み合わせと相対位置の関係を取り出す。このとき、他のシンボルと組み合わせる必要があると判断されたシンボルについては、組み合わせるシンボルごとに相対座標を計算して保存しておく。また、典型的なシンボルについては、他のシンボルとの接触点の位置も記録しておく。例として、図4-7に夕陽と家の相対位置分析の結果を示す。図中では、夕陽の位置を基準にしている。

この例では、家が夕陽の $0.08 \sim 0.09h$ 下、右の $0 \sim 0.1w$ 、左の $0.1 \sim 0.2w$ の位置にあることが分かる。このように、統計分析の結果によって、シンボルの配置を決める。

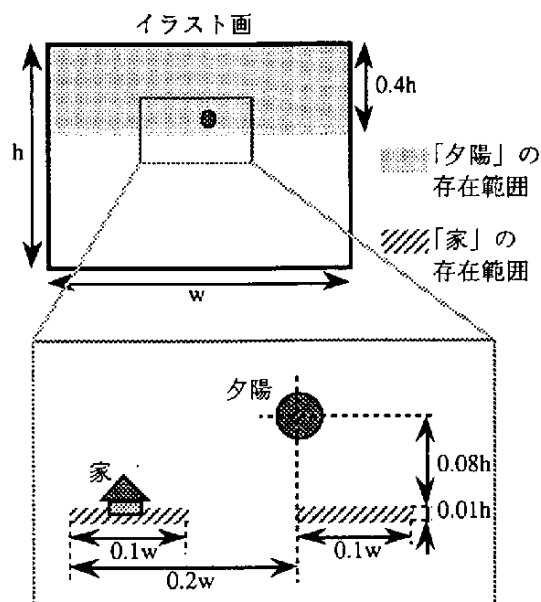


図 4 - 7 夕陽と家の相対配置

4. 2. 3 データベースの構築

本システムのデータベースは、以下に挙げたようなデータ構造でもって構築されている。

(1) 対象イラスト画の2次元配列データ

4. 1 節で示したシステムのものをを用いている。

(2) 感性情報を表わす形容詞テキストデータ

心理実験で推定したイメージを、テキストファイルとして保存する。

(3) シンボル画像の2次元配列データ

保存されたすべての対象イラスト画に対して、イラスト画の中にある典型的なシンボル（パーツ）を切り抜き、残っている部分を背景画像とする。切り抜いてできたシンボル画像をシンボルの名前とともに保存する。例えば、「妖精」という名前で、妖精の画像を保存する。これがシンボルの画像データベースになる。図 4 - 8 にシンボル画像の例を示す。

(4) シンボルの大きさ、配置、相対位置関係、組み合わせのデータ

(5) 背景の色情報データ

この (4)、(5) については、画像解析の結果をテキストデータとして記録する。

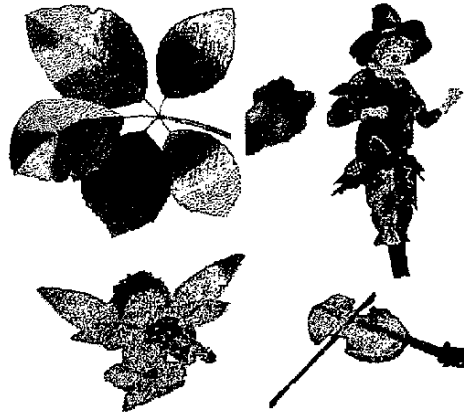


図 4 - 8 各種シンボルの切り出し例

4. 2. 4 イラスト画の自動生成

(1) 処理の流れ

従来のイラスト画の合成は、目や口などの部分写真を光学的に合成したり、テレビ技術の分野では、クロマキー手法により実時間で動画を合成したりする。こういった手法は IP(image processing)手法とって、画像から画像への処理手法である。本システムでは、シンボルの切り抜きに、この処理手法を用いている。一方、イラスト画の生成に関しては、CG(Computer Graphics)手法を用いる。この手法は、全体のイラスト画の共通点を利用して、複数枚のイラスト画を自動的に生成する。

本システムでは、二次元グラフィックス手法により、いくつかの画像を貼り合わせたり、また、必要なシンボルだけを拡大縮小したり、適当な位置へ移動したりして、入力された感性情報に対応するイラスト画を生成することができる。

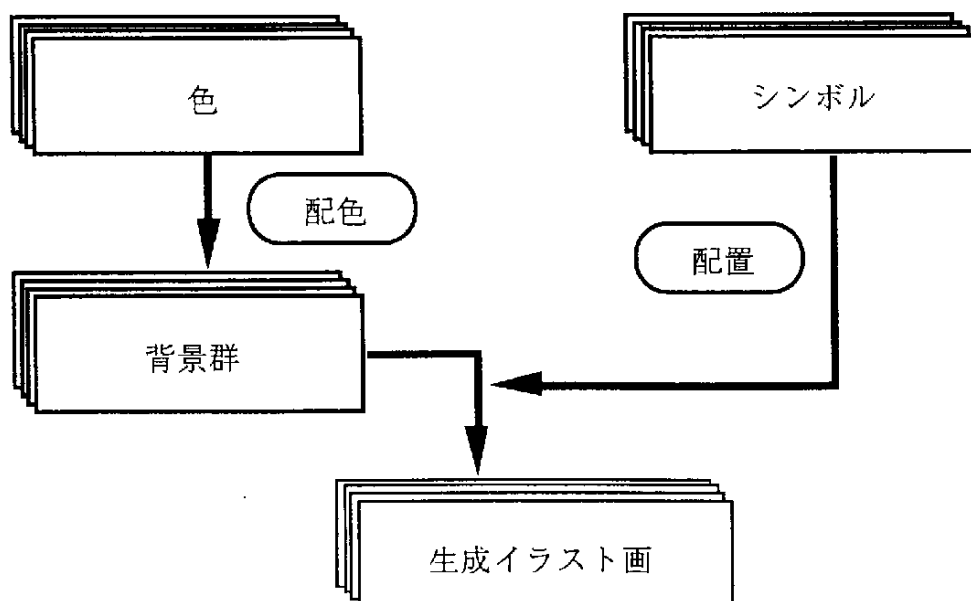


図 4 - 9 処理の流れ

まず対象イラスト画から典型的シンボルパターンを抽出し、データベース化する。次に、イラスト画を統計的に分析して得た背景の描画技法を実現し、背景的特徴量から背景群を得る。最後に、シンボルの組み合わせルールに従って、シンボル群と背景群を合成し、イラスト画を生成する（図4-9）。

（2）背景描画ルール

画像データベースの検索機能を利用して、対象イラスト画の背景色が一色のもの、二色のものと多色のものに分け、それぞれについて色を選別し配色を行う。その際には、以下のルールに従う。

イ．背景色が一色の場合

その一色が白であれば、ベタ塗りとする。それ以外の場合、背景色とその色のトーンを変えて得た新しい色の二色で直線的なグラデーションを描く。そのうち、30%の確率で、イラスト画に対してある範囲内でノイズ的なぼかしをかける。

ロ．背景色が二色の場合

第一色と第二色の面積の割合で背景を塗る。二色の塗る位置はランダムに決める。二色の境界付近のある範囲でグラデーション的なぼかしをかける。さらに境界線付近で平滑化するようなぼかしをかける。また適当な所で彩度を変える。

ハ．背景色が多色の場合

色の順序をランダムに決めて背景を塗る。描画技法としては、まず60%の確率で、背景に直線的なグラデーションをかけるが、このうち、さらに50%の確率で背景を変形する。また、10%の確率で、波的なグラデーションをかける。最後に30%の確率で、楕円的にグラデーションをかけるが、このうち、さらに50%の確率で、背景を変形する。なお、イラスト画全体の15%に対して、ある範囲内ではかす範囲をランダムに決めて、ノイズ的なぼかしをかける。

（3）シンボルの配置

イ．移動・拡大・縮小

シンボルを平行移動することによって、背景の任意の位置に配置することができる。シンボル画像は、その外接長方形領域の画像データと、領域内のシンボル部分を表すビットマスクとで構成されている。マスク外の画素は透明と考え、背景に置いた場合背景色となる。

また、イラスト画のサイズの変化とともに、シンボルのサイズも変える必要がある。これにより、少数のシンボルを有効利用することにもなる。サイズを変更するには、シンボルとシンボルのマスクを同時に拡大・縮小すればよい。

ロ. 組み合わせ

イラスト画を生成するには、シンボル同士やシンボルと背景の間の組み合わせを考える必要がある。ここでは、シンボルと背景との組み合わせはイラスト画のイメージへの影響は小さいと考え、シンボルと背景はランダムに組み合わせることとする。

シンボル同士の組み合わせについては、まず、共通の特徴を持っているシンボルを一つのシンボル型にまとめ、それらをさらに次の2種類に分類して考える。

主体 : イラスト画の中で一番大きなシンボル。分析の結果、ここに含まれるシンボル型は、大妖精・大動物・植物1～3の5種類であった。ちなみに大妖精とはサイズの大きな妖精のことであり、植物1～3はそれぞれ、上に付属要素を載せない植物、イラストの下部に配置する植物、その他の植物を指す。

付属要素：主体以外のシンボル。中妖精・小妖精・小動物・落ち葉・夕陽などである。

この2つの組合せルールを、イラスト画をモデル化することにより、次の表4-2のようにまとめた。

表4-2 組み合わせルール

大妖精	落ち葉だけ		9.7%
	その他		3.2%
大動物	小妖精を乗せている	落ち葉	3.2%
		木と家	3.2%
		その他	3.2%
	動物のみ		3.2%
植物1	植物のみ		6.5%
	中妖精が1つ		3.2%
	落ち葉と家と太陽		3.2%
植物2	上に中妖精	小動物	3.2%
		その他	12.9%
	上に複数の小妖精、家がある		3.2%
	そばに複数の小妖精		3.2%
植物3	上に中妖精がある	家と太陽	9.7%
		落ち葉	6.5%
		太陽	3.2%
		その他	9.7%
	その他		9.7%

シンボルの画像データベースから各シンボルを読み込んでくるときには、主体として使う場合はシンボルを拡大し、付属要素として用いる場合は縮小する。その後、ルールに従ってシンボルの移動などを行って、イラストを合成する。

4. 2. 5 生成結果

何枚かの合成イラスト画の例を図4-10に示す。ちなみに、上段左・中段右が「秋らしい」、上段右・中段左が「夏らしい」、下段左が「冬らしい」、下段右が「春らしい」という形容詞入力によって、それぞれ生成されたものである。

4. 2. 6 心理評価実験

次に挙げる6つの形容詞について、生成したイラスト画とそこから被験者が読み取ったイメージがどの程度一致するかを、評定尺度法により調べた。

「春らしい」、「夏らしい」、「秋らしい」、「冬らしい」、
「悲しい」、「楽しい」

詳細な結果は割愛するが、全体的に季節感を表す上段の4つについてはおおむね良好な結果が得られた。一方、感情を表す下段の2つについては、かなりのばらつきが見られた。また、色相よりもシンボルの方がより有効な特徴となっている傾向が見られた。

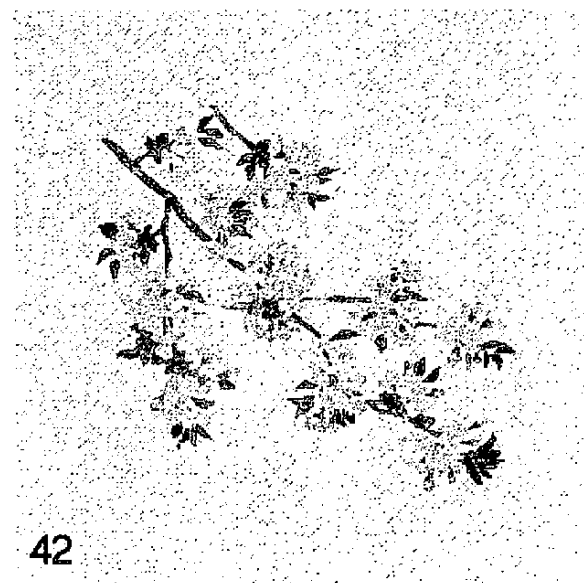
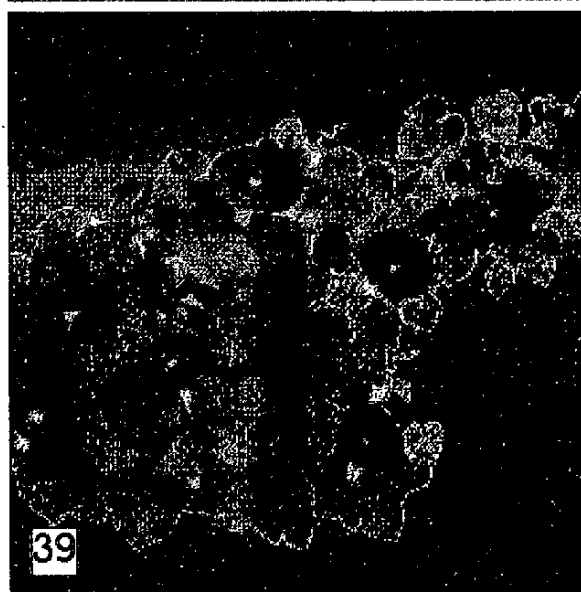
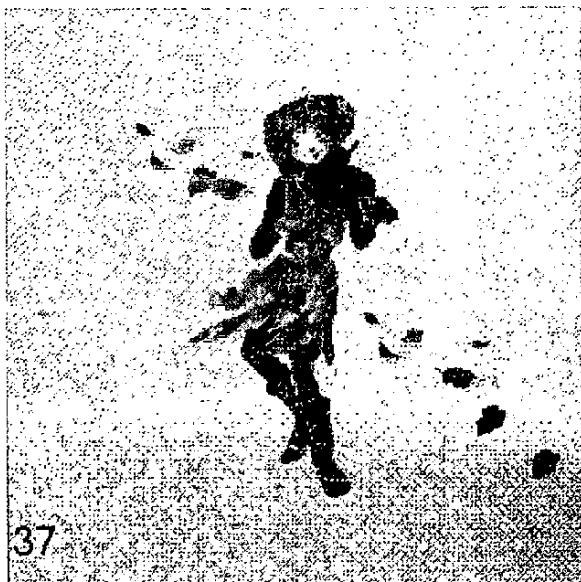


図 4 - 1 0 合成イラスト画の例

4. 3 音楽からの感性情報の抽出

本章の前半では、イラスト画からの感性情報の抽出とそれを用いた検索およびイラスト合成のシステムについて述べた。そこで、後半部すなわち本節と次節では、音楽における感性情報の抽出とそれを用いた音楽生成システムについて述べる。本節ではまず、拍子・調性・コード進行・メロディなど、音楽を構成する要素の認識について述べる。次に、そのような音楽要素を認識したうえで、それを形容詞とリンクさせて感想文を出力した結果を示す。

4. 3. 1 音楽プリミティブ

感性情報処理の目的は、“人間的な感情”に対する精密なモデルを立てることではなく、よりよいヒューマンインタフェースの形として“機械的な感情”を作成することにある。そこで、「どのようにしたら、より自然に人間とコミュニケーションできるシステムを作れるか」という主題をもって、音楽理解という問題を探っていくことにする。

さて、感情という問題に先立って実現しなければならないのは、音色（楽器）、メロディ、コード進行、リズム、演奏法などの物理的な音楽要素の認識である。少なくとも、これらを認識できなければ、音楽音響信号を単なる音、あるいは音の羅列を受け取ったに過ぎず、音楽として理解したとはいえるものではない。ここでは、このような音楽的要素を音楽プリミティブと呼んでいる。音楽プリミティブは物理的なデータから解析されるものであるが、それと同時に、ある種の感情（センチメント）と密接に結びついたものが多く、先述のパターン感性情報に分類されるものである。例えば、トランペット音のきらびやかさや、四七抜き音階の東洋的な雰囲気など、その理由はともかくとして、かなりの音楽プリミティブが直接的な形容詞（シンボル感性情報）と結びついていることは誰もが認めるところである。したがって、物理量である音響信号あるいはノート情報から音楽プリミティブの抽出ができれば、簡単ではあるが、「刺激→反応」というプロダクションをもって、その楽曲に対して感想文を出力することができる。

一方、コード進行やリズムなどがわからない音楽の未経験者が音楽を楽しんでいるという事実から、音楽プリミティブの認識を音楽理解の中間過程におくのはおかしいのではないかという意見もある。確かに、感性のすべてが個々の音楽プリミティブの単純な組み合わせから得られるものではない。が、少なくとも、音楽を楽しむ人は、程度の差こそあれ、音楽プリミティブに対する認識能力をもっているはずである。ただ、対象となる音楽プリミティブの呼び名を知らなかったり、説明ができないために、音楽プリミティブに対する認識が行なわれていないように見えるのである。このような事柄を考慮すると、音楽感性情報処理システムは、音楽経験者より、むしろ、音楽の経験の浅い人が、感覚的に音楽を扱うためのツールと位置づけることができよう。

音楽という世界を形作る制約条件は、基本的には、音の高さ、時間的間隔であるが、それぞれさまざまなレベルのものがある。音の高さに関しては、長調、短調など普段耳にする音階構造から、半音階・全音階、さらに周波数成分、すなわち、音色そのものが秩序の構成要素となるような音楽が存在している。また、時間的にも、楽節のようなグローバルな制約をはじめ、拍子・小節から、無拍子、さらに、時間そのものがパラメータであるというレベルに至るまで、現在、非常に範囲の広い世界で音楽が制作されている。もしも現代音楽まで含めると、扱うべき音楽の秩序は非常にあいまいなものになってしまうため、計算機で音楽理解を扱って行くときには、対象に対するある意味で厳しい制約を課す必要がある。また、情報処理として扱う以上は、いくら主観的な対象を扱いたいといっても客観性を排除してしまうことはできない。ここでは、最も親しまれているという点から、ポピュラージャンル、または、その範疇にはいる音楽を対象として、音楽プリミティブの抽出、および、センチメントの抽出を行なうことにする。

さて、音色、テンポ等の音楽プリミティブは音響から直接的に得られるものである。以下、高次の音楽プリミティブの抽出として、コード、および、メロディ、調性、リズムの認識について述べてゆく。なお、入力については、図4-11に示すようなデータ音高、オンセットタイム、長さ、大きさからなるノート情報を与えるものとする。

```
Header {
  UNDEFINED_UNIT 48 <---- 基準拍の長さが0.48秒
}
0 (A3 4 16)
1 (B3 2 25) (G#3 2 12)
2 (C4 6 22) (G3 2 18)
3 (D4 3 17) (F#3 2 12)
.
.
.
```

図4-11 ノート情報

4. 3. 2 リズムパターンの認識

リズムとは、音のいくつかが単位となって聞こえる繰り返し過程である。音楽においては、音の高低・長短・強弱・音質がリズムの決め手となる。Longuet-Higginsはメロディを構成する音の開始時刻と長さに注目した拍子の解析法を示した。Longuet-Higginsの示した方法では、基本的なリズムは構造全体を2分割（2連符）あるいは3分割（3連符）していく過程で生成されるという原理と、テヌート・スタカートの割り付けられる位置に関する条件が利用されている。

一方、本システムでは、音の大きさが採譜処理により測定されているので、強拍を追跡するという最も単純な方法をとることができる。一般に、音楽で使われる拍子は二拍子系、3拍子系に大別され、またビートとしては、4ビート系、8ビート系などがある。もし、拍子やビートが正しいものであれば、小節、あるいは、それぞれのビートにおけるアクセントの位置は一致したものとなるはずである。具体的には、装飾音を除く最小長さを単位とする立ち上がり時のパワーの配列を作成し、ビート n ($= 2, 3, 4, 8$) で割った剰余が一致するインデックス群の配列要素の総和に n をかけ、その中から大きいものを選び出して行くことにより、強拍のポジションを決定することができる。なお、モードが0でない音列群に最大値がある時は、対象が弱起であることを示している。

4. 3. 3 調性の認識

メロディを入力として主調を認識するという研究は、リズムの認識と同じく Longuet-Higginsによってなされている。Longuet-Higginsは5度圏による音の離れ（主調からの距離）と隣接音間の距離から調性を解析する精密な具体的方法を示している。処理的には部分部分で使われている音とスケールとのマッチングをとっていることにほかならない。

本システムでは、単声およびメロディがわかるときには、音高のオクターブのヒストグラムを作成し、それに対し24種の音階（主音(12)*長短調(2)）とのマッチングを行う。近親調、例えば、ハ調から見てヘ調とト調で使われる音の違いは、それぞれ、一つ分のフラット、一つ分のシャープだけである。したがって、マッチングの結果、主調とその近親調で、大きな値をもつものを選んでやることで、主調を決定することができる。

ところが、上記の方法だけでは、経過音の多い対象やメロディの抽出ができない多重音に対しては誤認識を起こしてしまうことがある。そこで、多重音についてはコード進行やテンションの解析を用いて調性の解析を行なうことにする。具体的にはまず、対象の最初のコードとテンションが許されうる調を複数個候補としてあげる。続いて、順次、その次のコードから許容できない調を消去していくことにより、調を決定するという手法を用いる。この際にGからCまたはDmからGのような解決を促す進行があるときには、調性の確からしさに対する重みを与えることにより、認識の効率化を計ることができる。一旦、調性が決定された後は、転調の可能性をトレースしてゆくことが可能である。また、後に示すように、数個の音列データで調を決定できるという利点もある。

4. 3. 4 メロディの抽出

メロディ自体の解析等は非常に積極的に行なわれてきたが、多重音、あるいは、複数の楽器からなる音楽におけるメロディ、あるいはサブメロディを抽出すること

は非常に難しいとされている。このタスクに関しては、基本的には2つのアプローチが考えられる。一つは信号処理からのアプローチである。例えば、歌謡曲のように歌のパートがメロディであるというときには、人間の声だけを抽出する信号処理が実現されれば、メロディの抽出は可能となる。二つ目は、音楽理論からのアプローチである。これは、ピアノあるいはコーラスなどの多重音のシーケンスから、音楽的制約を用いてメロディを抽出する手法である。これに近い研究としては、心理学の分野において、群化の研究（旋律構造の分離）が試みられている。しかし、これはどのように多声が記憶されるかを実験的に示したものであり、メロディ抽出のアルゴリズムを確立したものではない。

本システムではメロディの構造上の制約として、

- ・大きく聞こえる音の列である。
- ・和音でなっているときは、最も高い音がメロディの構成音である。
- ・極端な跳躍進行はない（音が飛び過ぎない）。

といった、簡単な信号処理と音楽ルールを組み合わせることにより抽出を行う。簡単ではあるが、このような制約だけで、多くの楽曲に対してメロディを抽出することができた。しかし、これらのルールは物理的な要素しか考慮していないため、ベースパートがメロディを担当するような場合やオブリガートが入った楽曲などに対処できないという欠点がある。

4. 3. 5 コード進行の認識

コード進行とは、メロディに対する伴奏部の和音シーケンスである。コード進行を調べることは、楽曲の雰囲気を知るうえで非常に重要な手掛かりになる。

コードからは、基本的に基音とドミナント音と長短を決定することができる。原理的には、それぞれのコードの展開系をモデルとしてマッチングを行なうことにより認識できると考えられる。しかし、実際にはアルペジオのようにコードを展開して演奏を行なったり、また、テンションが加えられたり、省略音などがあるため、単純なマッチングだけでは認識できるものではない。ここでは、局所的な範囲でコードネームを決定するステージと、隣接するコードネームから統合しうるものは統合するというステージの、2段階の方法でもってコードの割当を行なっている。

処理の流れという面から述べると、次のようになる。まず、伴奏部のベース音を仮定し、それをもとにどのような半音間隔で音が積み重なっているかを探索することにより、局所的なコードネームを決定していく。もし、ベース音を仮定したときにコードを決定する音があれば、その時点でコードネームを与える。しかし、コードを決定するまでの間、ベース音がオンの状態で持続されれば問題はないが、そうでないときには簡単にコードを決定することはできない。このような場合には、前になったベース音が、心理的には鳴っているものと仮定し、次に続く音の探索をしながら、コードの決定を行なう。この時、探索される音は、前のベース音に対するコードを決定する要素として使われるとともに、新たなベース音の候補としてフラ

グをつけ、コードの探索要素に加えられる。もし、探索途中でコードを決定する音が見つかった時には、途中でベース候補になった音に対し、ベース音候補のフラグをはずしてゆく。なお、省略音の問題に対処するために、コードを認識するための優先度が付けられている。例えば、Cの音がベース音という仮定があるときにEが発見された場合は、Cメジャーというコードが仮定される。もし、その後にAの音が来れば、Amとして認識される。照合の行なわれるコードは、ポピュラー音楽でよく使われるもの13個（Cを例に上げると、C C(2) C(3) C9 C7 Cmaj7 Caug Cm Cm(2) Cm(3) Cdim Cm7 Cm9）を用意している。実際は12通りの調性があるため、計156通りのコードが割り付けられる可能性がある。

さて、上記のように局所的にコードネームを与えて行く方法では、コードが過度に分割されてしまう恐れが多分にある。そこでそれを防ぐために、BdimならばBm7に読みかえてよいというような機能の代行に関するルールや、連続するコードでテンションが異なるだけなら統合を行なうといったルールを適用して、マーキングを行なうことにより、人間が与えるのと同じようなコードの割当が行なわれるようにする。

4. 3. 6 音楽プリミティブと形容詞のリンク

本システムには、コード進行・テンポ・音色などの音楽プリミティブとセンチメントとのリンクを記述するルールが、約30個登録されている。これは、音楽プリミティブに対するアンケートより得たものである。この数自体は少ないが、音楽が組み合わせの芸術であることを考えると、かなりのバリエーションの感想文を生成することができる。ルールの例を図4-12に示す。

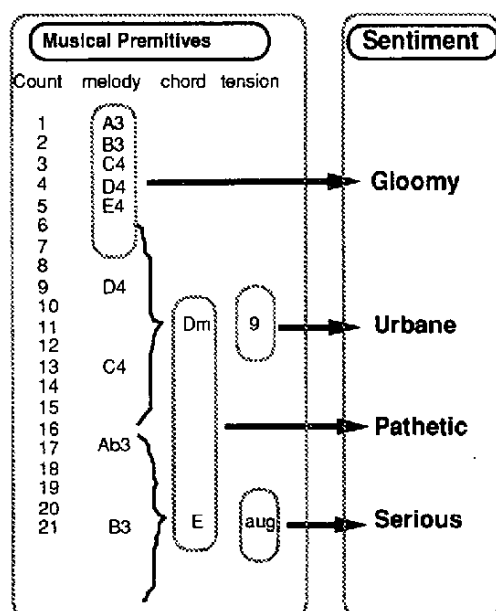


図4-12 センチメント抽出のためのリンクの例

4. 3. 7 実験および検討

ここでは、村松健作曲「封じ込められた光と影」を例に解析結果を示す。この対象のノートデータの一部を図4-13に示す。

Header {	29 (E2 1 16) (B2 1 14)
VL ON	30 (E2 2 16) (B3 1 8) (B2 2 15) (F#4 1 5)
UNDEFINED_UNIT 48	31 (E3 2 20) (C3 1 15) (C4 2 5)
}	32 (A3 3 27) (B2 4 47)
0 (A3 4 16)	33 (B3 1 27) (G#3 2 22)
1 (B3 2 25) (G#3 2 12)	34 (G3 2 31) (C4 6 25) (E2 2 14)
2 (C4 6 22) (G3 2 18)	35 (D4 3 19) (F#3 2 19)
3 (D4 3 17) (F#3 2 12)	36 (A3 4 16) (E4 1 8) (D2 1 9) (F3 1 6) (D3 2 6)
4 (A3 4 14) (E4 1 7) (D2 1 7) (D3 1 5)	37 (A2 2 12) (F3 3 7)
5 (A2 2 13)	38 (D3 3 31)
6 (D3 2 25)	39 (E3 2 15)
7 (E3 2 17)	40 (A3 4 22) (D4 2 18) (F3 1 11)
8 (D4 4 18) (A3 4 15) (D3 4 23)	41 (F3 3 23)
12 (G#3 3 21) (D3 3 35) (C4 2 14) (E2 1 7) (E3 1 7)	42 (D4 2 5)
13 (B2 1 21) (E2 2 7)	44 (G#3 3 21) (C4 3 16) (D3 3 28) (E3 1 7) (E2 1 6)
14 (E3 3 22) (B2 1 22)	45 (B2 2 20) (E2 2 7)
15 (F#3 2 19) (B2 3 25)	46 (E3 1 14)
16 (G#3 4 21) (D3 4 28)	47 (F#3 2 20) (B2 3 26) (E3 3 15)
17 (E3 2 5)	48 (G#3 4 22) (D3 4 30)
18 (B2 2 8)	50 (B2 2 8) (E3 2 2)
19 (E3 1 3)	52 (G3 2 24) (E3 1 26) (B3 2 11) (A1 3 12)
20 (G3 3 23) (B3 1 8) (C3 1 16) (B4 1 6) (E3 1 7)	(C3 1 11) (A2 2 11)
(E4 1 3) (A1 2 6) (A2 2 6)	53 (E3 1 27) (E2 3 16) (G#4 2 2)
21 (E3 2 16) (E2 3 14) (G#4 2 1)	54 (B2 2 26) (B3 1 10)
22 (B2 2 24) (B3 1 11)	55 (C3 3 18) (C4 1 7)
23 (E3 1 16) (C3 1 14)	56 (G3 4 32) (B2 3 18) (E4 1 8)
24 (G3 3 28) (E3 1 21) (B2 1 19) (C3 2 6)	57 (B3 1 9)
25 (B2 2 38) (E3 2 36)	58 (C3 2 17) (C4 2 7)
26 (C3 1 23) (C4 2 9)	59 (E2 2 16) (B3 1 7) (B2 1 6)
27 (B2 2 40) (E3 1 18)	60 (G3 4 29) (A2 2 19) (A1 3 17)
28 (E3 3 37) (A1 4 23) (A2 1 20)	61 (E2 4 18)

図4-13 解析対象となるノートデータ

図4-14に、このノートデータからコード進行を解析した結果の一部を示す。また、図4-15には、この曲に対し、システムが生成した感想文を示す。ここでは、曲全体に散らばったセンチメントのうち、多いものから順にピックアップすることにより、曲を支配するセンチメントを調べ、文を作るテンプレートに形容詞を埋め込む方法を用いて感想文を作成している。また、コード進行のパターンとメロディのピーク部分から「サビ」（最も盛り上がる部分）を抽出し、特にその部分に対する形容詞情報を述べる、という処理を行なっている。

全体的に悲しげで、「サビ」の部分で希望的な部分が見受けられるという評価は、対象曲の題名をよく表現したものと言える。

今後の課題としては、扱いうる音楽プリミティブを増やしてゆくこと、具体的に

は、その認識法およびどの形容詞とリンクさせるかという問題があげられる。また、より複雑な楽曲を対象とすることを考えると、音楽プリミティブ抽出のための知識の構造化やその獲得法に関する検討が必要となる。

from 5 to 8 chord = Dm9	from 5 to 12 chord = Dm9
from 9 to 12 chord = Dm	from 13 to 20 chord = Eaug
from 13 to 13 chord = Eaug	from 21 to 33 chord = Am7,9
from 14 to 20 chord = E7,9	from 34 to 34 chord = Eaug on A
from 21 to 28 chord = Am7,9	from 35 to 35 chord = C (2) on E
from 29 to 33 chord = Am9	from 36 to 36 chord = Em7,9
from 34 to 34 chord = Eaug on A	from 37 to 44 chord = Dm9
from 35 to 35 chord = C (2) on E	from 45 to 52 chord = Eaug
from 36 to 36 chord = Em7,9	from 53 to 65 chord = Am7,9
from 37 to 44 chord = Dm9	from 66 to 66 chord = Eaug on A
from 45 to 45 chord = Eaug	from 67 to 67 chord = Am
from 46 to 52 chord = E7,9	from 68 to 68 chord = Bm on E
from 53 to 60 chord = Am7,9	
from 61 to 63 chord = Em on A	
from 64 to 65 chord = Am9	
from 66 to 66 chord = Eaug on A	
from 67 to 67 chord = Am	
from 68 to 68 chord = Bm on E	

a) 初期結果

b) マージング後

図 4 - 1 4 コード認識結果

Whole atmosphere of this music is
 slow and leisurely.
 sorrowful.
 smart and serious. Level is 18.
 urbane. Level is 7.
 rural. Level is 6.

There is pathetic mood on chord from 5 to 33.
 There is pathetic mood on chord from 37 to 65.
 There is hopeful mood on chord from 69 to 97.
 There is melancholy mood on chord from 105 to 116.
 There is pathetic mood on chord from 133 to 153.

On release at 75, there is hopeful, urbane, smart, serious mood.

図 4 - 1 5 生成された感想文

4. 4 感性パラメータを利用した作編曲支援システム

4. 4. 1 システムの構成

コンピュータを「知的創造のための助手」として使うことを考えた場合、システムがどのような仕様を満たしている必要があるかは、どのような入力に対してどのような出力を得たいか、によって決まる。そこでまず初めに、ここで目指しているところの、感性情報を用いてより音楽性豊かな曲を作成できるシステムについて、次のような入出力を考えた。

Input : ○ モチーフになるコード進行

○ 感性パラメータ (形容詞に代表される感性語)

Output : ○ 入力された感性パラメータに応じた曲

これを実現するためには、以下のような要求を満たしていなければならない。

- (1) 感性パラメータを記号あるいは物理量として扱えるように変換するインタフェース部を備えていること。
- (2) 与えられたコード進行を分析する能力を持っていること。
- (3) 感性パラメータで入力された情報を楽典的知識に基づくルールに対応させ、知識処理できること。具体的には、感性パラメータの入力に応じたコードの展開やメロディ付けができること。
- (4) 知識の競合をうまく解決できること。
- (5) 生成された曲をユーザに音として返すことができること。
- (6) 生成された曲の構造あるいは印象を視覚的に分析・評価できればより良い。

これらを考慮して、図4-16に示すようなブロックダイアグラムを持った、感性パラメータを利用した作編曲支援システムを実現した。

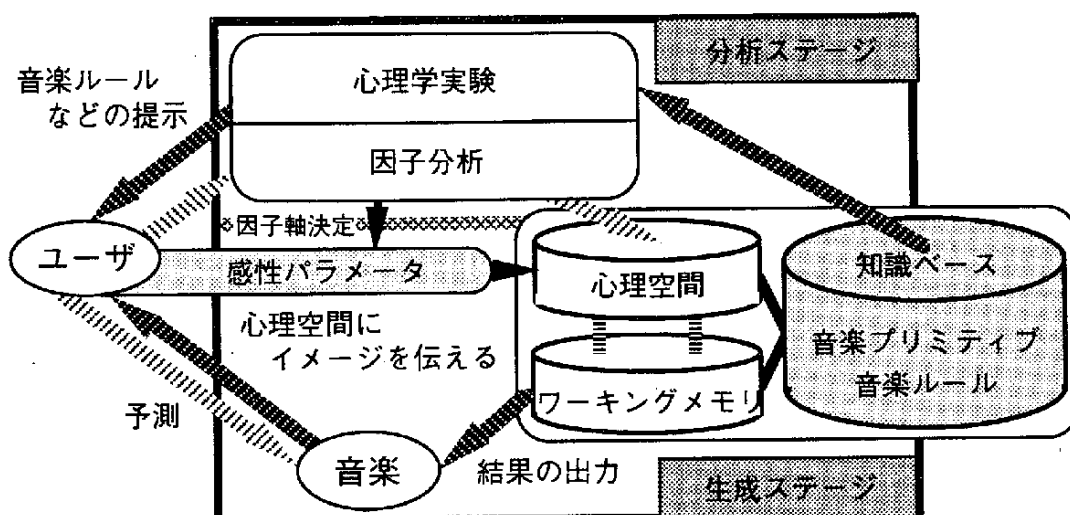


図4-16 感性パラメータを利用した作編曲支援システム

(1)の要求に対しては、心理実験の結果を因子分析して得た感性パラメータを利用することにした。上図の分析ステージがそれに対応している。ここでは、イメージとテクニックや知識の相関関係を、人間の心理に例えて心理空間と名付ける。まず、43種の感性語と音楽知識の関連性を心理学実験によって探る。次に、この結果を因子分析して因子軸（音楽の特性を表現するための必要十分な感性語）を抽出し、この軸を感性パラメータに割り当てて多次元の心理空間を構築する。その結果、ユーザが感性パラメータを使ってイメージを入力すると、心理空間の座標が計算され、そのイメージに最も近い音楽ルールおよびその組合わせを選択する事ができる。もちろん、いったん心理空間を構築してしまえば次から再構築する必要はないし、また心理空間そのものをデータとして蓄えるので、他人の感性で作編曲を行なうこともできる。

また生成ステージに対応する(2)(3)に関しては、プロダクションシステムと知識ベースを用いることにした。具体的には、まず、あらかじめ和音やカデンツや楽式といった音楽知識をIF...THEN...形式でファイルに保存しておき、これを必要に応じて参照する。そして、ユーザが与えた感性パラメータから選択された音楽ルールを、実際に適用して音楽を変形させてゆく。このとき問題になる(4)の競合解消戦略としては、ルールに感性パラメータより与えられる重み付けを行ない、それを確信度として最終的に最も確信度の高いルールが適用される、という手法を採用する。

(5)に関しては、MIDIを経由して音源をドライブできるようにした。このことによって、編集された結果を音として聴きながら、さらに手直しを加えて行くことができる。このようにして、動的で非因果的な処理を実現するのである。

4. 4. 2 分析ステージの処理

(1) 心理空間と感性パラメータ

分析ステージは、イメージを具体的なコード進行や作編曲作業におけるルールの組合わせ集合に対応づけるための、心理空間(Mental Space)を構築することを目的としている。具体的には、感性語と具体的事象の相関性について心理学実験を行ない、この結果を因子分析することによって、感性パラメータによって形成される心理空間を得るというものである。ここで、感性パラメータとは、一般に、感性量と物理量を結ぶ要素のことである。本システムでは、音楽を構成するコード進行などの要素を心理的イメージで置き換えるための尺度を指す。この感性パラメータを軸として張られる多次元の空間が心理空間であり、その概念図を図4-17に示した。

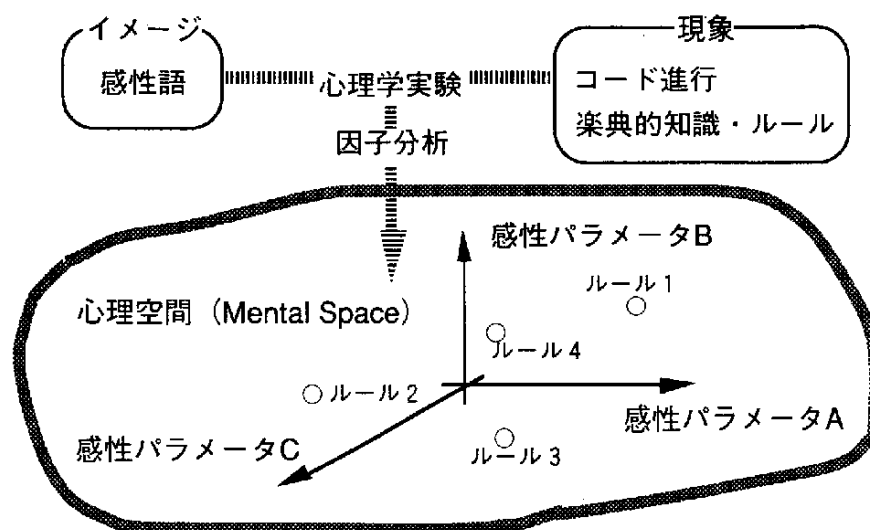


図4-17 心理空間の構築

(2) 心理学実験

さて、心理学実験を行なって、感性パラメータと音楽プリミティブの相関関係を求めるわけだが、一般的な心理空間を定義するという目的に限れば、すべての音楽プリミティブに対して実験を行なう必要はない。ここでは、コード進行と情動の相関モデルから因子軸を算出することにした。この因子軸を感性パラメータに対応させれば、そのパラメータを用いて、適当なコード進行を選択させることも可能になる。

具体的には以下の手順で心理実験を行った。まず、ポピュラー音楽やクラシック音楽で使用される代表的なコード進行およびカデンツを選び出して、相対音階に変換する(25個)。その中から代表的な例を表4-3に示す。

表4-3 実験に用いたモチーフの例

1. /Am/D/G/C/F/Dm/E7/E/
2. /G/C/F/B^b/E^b/A/Dm/C/
3. /G/C/Bm/Em/Am/D7/G/G/
4. /C/G/Am/Em/F/C/F/C/
5. /C/G/C/F/C/Am/Dm/G/
6. /C/C/B/B/B^b/E/A/D/
7. /Am/G/F/Am/C/D/E7/E/
8. /Am/G/F/Em/Dm/Em/Am/Am/
9. /Am/Dm/F/Em/Am/Dm/F/Em/
10. /F/C/Dm/C/F/C/Dm/C/
17. /C/Em/F/G/Am/F/G/G/
19. /C/Em/F/E/Am/Em/F/Dm/
23. /Am/Em/F/G/Am/Em/F/G/

次に、43種の形容詞対を並べた表を用意し、それぞれのモチーフについてモチーフの録音されたテープを聴きながら、形容詞対群との相関関係を1-10の得点で表し、この得点群とモチーフ群の相関行列について因子分析を行なう。これにより得られた音楽の特徴を最も良く表現できるパラメータは、以下の5つである。

イ. 因子1：緊張感 (Tension)

[緊張感のある (Tense) - 開放感のある (Lax)]

緊張感のあるコード進行のモチーフで特に因子負荷量が大きかったので緊張感と名付ける。

ロ. 因子2：明るさ (Cheeriness)

[朗らかな (Cheers) - 哀調をおびた (Plaintive)]

明るい感じのモチーフの因子負荷量が大きいのでこのように定める。

ハ. 因子3：洗練度 (Refiness)

[洗練された (Refined) - 素朴な (Simple)]

エレガントなイメージのあるモチーフで因子負荷量が大きく、素朴なイメージのモチーフは比較的因子負荷量が小さいので洗練度と名付ける。

ニ. 因子4：躍動感 (Liveliness)

[軽やかな (Lively) - 重々しい (Gravely)]

軽やかな感じのするモチーフで因子負荷量が大きいため躍動感と名付けるのが適当である。

ホ. 因子5：多様性 (Variousness)

[変化にとんだ (Various) - 単調な (Monotonous)]

シンプルなモチーフの因子負荷量が比較的小さく、複雑なモチーフで因子負荷量が大きいので多様性と呼ぶ。

4. 4. 3 生成ステージの処理

生成ステージは、感性パラメータによって与えられたイメージからコード進行を生成することを目的としている (図4-18)。具体的には次の2通りの生成法がある。

- A. ユーザが与えた基本的コード進行 (4小節相当) のケーデンスを解析し、ルールによって8小節のコード進行を複数生成する。そして、これらのカデンツ (終止形) の形式を分析して心理空間に配置する。ここから、ユーザは感性パ

ラメータによってイメージにあったコード進行を選択できる。

B. ユーザが入力した感性パラメータから、システムがあらかじめ内部に持っている基本的なコード進行（表4-3参照）が選択される。

Aの方はユーザの与えたコード進行を心理空間に配置する必要があるので、そのための前処理が加わったのであり、コード進行の選択以降は両者とも同じ処理を行なう。

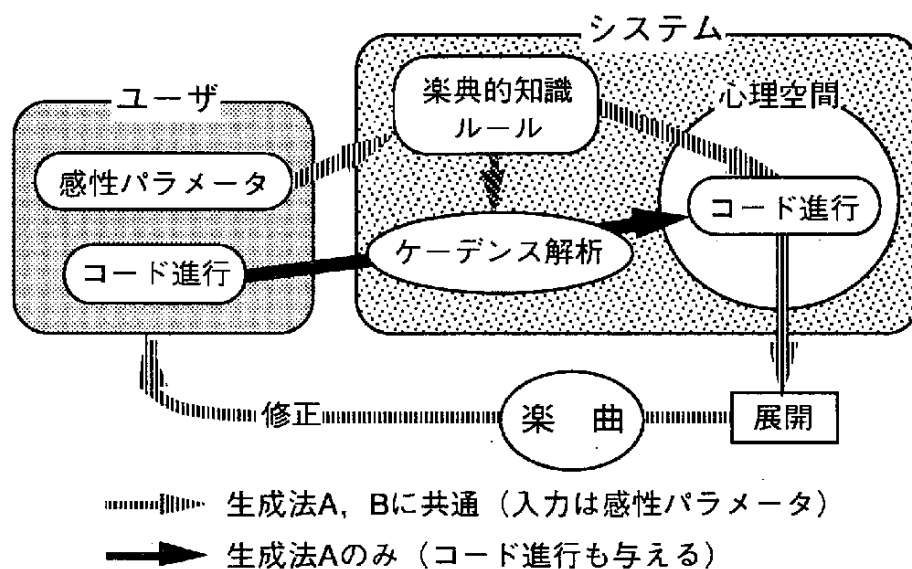


図4-18 生成ステージ

具体的には、例えば $A_m \rightarrow E_m \rightarrow F \rightarrow C$ というコード進行を与えた場合、システムは、以下のようなルールによる3通りの展開結果を提示する。

- 1) $A_m \rightarrow E_m \rightarrow F \rightarrow C \rightarrow D_m \rightarrow G \rightarrow C$
- 2) $A_m \rightarrow E_m \rightarrow F \rightarrow C \rightarrow D \rightarrow E \rightarrow A_m$
- 3) $A_m \rightarrow E_m \rightarrow F \rightarrow C \rightarrow A_m \rightarrow E_m \rightarrow A_m$

これらのカデンツは、

- 1) Subdominant $\rightarrow$ Dominant $\rightarrow$ Tonic（長調）
- 2) Subdominant $\rightarrow$ Dominant $\rightarrow$ Tonic（短調）
- 3) Tonic $\rightarrow$ Dominant $\rightarrow$ Tonic（短調）

である。分析ステージにおける心理空間の構築過程において、これらのカデンツをもつ基本的コード進行は、あらかじめ心理空間のなかに割り当てられている。したがって、生成されたコード進行は類似度が最も高い基本的コード進行と同じ位置に配置される。

以上のようにして、入力されたモチーフは編曲を施されて心理空間へ配置される。以後は、感性パラメータの入力によって生成されたコード進行も内部にあらかじめ持っていたものと同様に、感性パラメータによって選択可能となる。

また以下のような特徴を付加することもできる。これらは記号的信息が直接音楽の形式に結びついている例である。

- ・感性パラメータを入力するとき「明るさ」を強調すると、長調である1のシーケンスが選択される。
- ・「軽快さ」を与えた場合はシンコペーションまたは弱起が付加される。
- ・「なめらかな」ベース音のシーケンスを要求するとコードの構成音を調べて、続く2つの音程が最短距離で結ばれるような経路を探索する。但し最後の小節のベース音は必ずコードのルート音を取る。

4. 4. 4 実験結果

(1) ユーザーインターフェース

ユーザーインターフェースを、図4-19に示す。画面は、感性情報を扱う部分(上半分)とそれ以外のコマンド情報を扱う部分(下半分)に分れている。また感性情報は、感性パラメータの入力部(上部左)と、心理的評価グラフを表示する出力部(上部右)で扱われる。言語で入力された感性情報の結果が、音楽として耳へ返されるのと同時に、心理的な安定感のグラフとして目に返される。

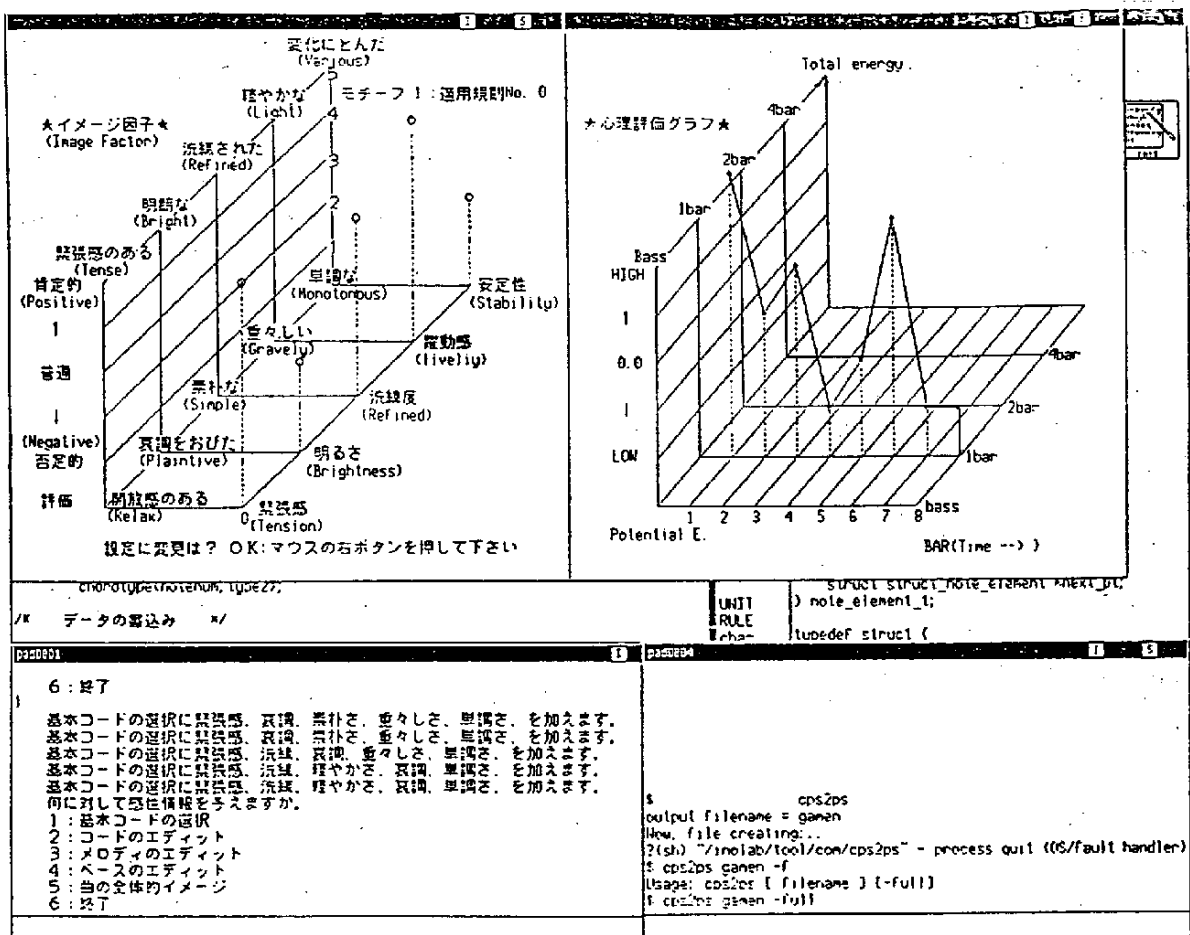


図4-19 ユーザーインターフェース

(2) 結果と考察

方法：まず、4小節のコード進行のモチーフを与えて、8小節のコード進行を要素とする心理空間を得る。そして、感性パラメータを用いて、生成されたコード進行のうち一つを選択させる。最後に、出力結果が入力したイメージに該当するかどうかを検討する。

結果：Am → Em → F → C というコード進行を入力し、感性パラメータを任意に設定することにより、

1. Am → Em → F → C → Dm → G → C
2. Am → Em → F → C → D → E → Am
3. Am → Em → F → C → Am → Em → Am
4. Am → Em → Dm → E → Dm → G → C
5. Am → Dm → G → C → Am → Dm → G → C
6. Am → G → F → F → Dm → E → Am

という6つのコード進行が得られた。1番～3番は与えられた4小節を元に生成されたもの、4番～6番は心理学実験で使われたものであらかじめシステムが持っていたコード進行であり、与えられたコード進行に類似していたものである。表4-4に、それぞれのコード進行が得られたときの感性パラメータの値を示す。

表4-4 コード進行が得られたときの感性パラメータの設定

番号	緊張	明るさ	洗練	躍動感	安定
1	2	5	2	4	3
2	4	2	3	3	4
3	3	3	3	5	3
4	4	4	3	3	4
5	3	3	4	2	3
6	2	2	4	2	3

値が大きいほどその性質を強く持っていることを示す

検討：1番と4番のコード進行は、カデンツが長調の型すなわちスケール上の根音であるCのコードで終わっており、他のものが短調型であることを考えると、「明るさ」を要求したときにこれらが選択されるのは妥当である。同じカデンツを持っていたとしても、5番が「明るくもなく暗くもない」と評価

されるのは、同じコードが循環しており長調とも短調とも判断がつかないからである。

また、3番のコード進行が、「躍動感」を要求したときに生成されるのは、強進行が連続しており、はつらつとした印象を与えるからである。

5、6番が洗練を要求したときに出力される理由としては、5番は長調とも短調とも判断がつかず、6番も弱進行（F→F）が存在して他の進行にはない不安定さを持っていることが考えられる。ここから、我々が素朴と感じるのは安定進行の時、洗練されていると感じるのは少し不安定な要素があるときである事がわかる。これは、テンションが存在するとエレガントな感じや不安定な感じを受けるのに似ている。

最後に、2番のコード進行は、実験した全ての中で最も特異なパラメータ設定で得られた。緊張感、安定感が共に高い。一見、背反するように感じられるが、別の次元の事象を表現しようとした結果、このようなパラメータ値に対応したものであると考えられる。この原因は、このコード進行が Subdominant → Dominant → Tonic と進行し、非常に安定している半面、DやEのコードが含む第3音は、F#、G#共にスケール外の音であり、普通の短音階とは異なった旋律的短音階を形作っているからである。よって、C → D → E という部分が緊張感のある展開になっている。

以上、ここに見るようにコードの機能を解析して心理的空間に配置されたコード進行は、実際の音楽的構造をうまく代表する位置におかれているといえる。

参考文献

- 電子情報通信学会編：コンピュータグラフィックス，オーム社
日本図学会編：CGハンドブック，森北出版
内藤・佐藤・井口：「イラスト画のイメージ検索に関する研究」，第35回自動制御連合講演会3021，pp.421-422
浅野哲夫：計算機幾何，浅倉書店
小林重順：カラーイメージスケール，講談社（1984）
R.Nevatia著，南敏監訳：画像認識と画像理解，啓学出版（1982）
千々岩：色彩学，福村出版（1987）
大田昌孝：応用グラフィックス，アスキー出版
深澤孝哉：色の技法，アトリエ出版社（1991）
視覚デザイン研究所編：配色ノート（1988）
加藤・栗田：「マルチメディア対話における個人差の考察」，情報学基礎，15-7（1989）

鷺尾泰俊：推定と検定，共立出版

藤田・古郡・畑：「心理空間を利用した画像検索方法」，コンピュータビジョン，
1987.11

亀田・藤崎：「テーマ・キー概念・キーワード間の階層構造を利用する新聞記事情報
の分類・検索システム」，情報処理学会論文誌，Vol.28，No.11
(1987)

田辺・大谷・石井：「多次元心理空間を用いる類似画像検索法」，電子情報通信学
会論文誌，Vol.75，No.11 (1992)

音楽の友社：新音楽辞典 (1975)

新原・今井・井口：「歌唱の自動採譜」，計測自動制御学会論文集，Vol.20，
No.10 (1984)

Y. Nakamura and S. Inokuchi : "Music Information Processing System and its Application to
Comparative Musichology", 6th IJCAI (1979)

J.A.Moorer : "On the segmentation and analysis of continuous musical sound",
Rep.STANM-3 (1975)

片寄・今井・井口：「ニュー・エイジ・ミュージック分析のための採譜システム」，
第35回情報処全大，pp2783-2784 (1987)

片寄・今井・井口：「感性情報抽出のための採譜システム」，第2回人工知能学会
全国大会，pp585-588 (1988)

岡謙太郎：「楽譜印刷」，コンピュータと音楽，bit別冊，共立出版 (1987)

片寄・今井・井口：「ルール駆動型写譜システム」，春季信学全大，D-413 (1988)

片寄・今井・井口：「音楽における感性情報抽出の試み」，人工知能学会誌，
Vol.3，No.6 (1988)

J.Canny : A Computational Approach to Edge Detection, IEEE Trans., Vol-PAMI, No.8

原・井口：「複素スペクトルを用いた周波数同定」，計測自動制御学会論文集，
Vol.19，No.9

畑中・市川：「音声認識におけるアルゴリズム」，信学誌，Vol.69，No.4

D.Jaffe : Ensemble Timing in Computer Music, Computer Music Journal, Vol.9, No.4

片寄・井口：「知的採譜システム」，人工知能学会誌，Vol.5，No.1

H. Katayose, T. Fukuoka and S. Inokuchi : "Expression Extraction in Virtuoso Music
Performances", Proc. of Intl. Conf. on Pattern Recognition, pp. 780-784

加藤・井口：「小節単位処理に基づいたピアノ楽譜の自動認識」，電子情報通信学
会論文誌，Vol.J71-D，No.5

T. Taguti and M. Ohta : "Agogics as a Metrical Rhythm and Phrasing", Proc. Of 1st. Intl.
Conf. on Music Perception and Cognition

A. S. Tanguiane : "A principle of Relativity of Perception and its Applications to Pattern
Recognition in Analysis of Performed Music", Proc. of 1st. Intl. Conf.
on Music Perception and Cognition

第5章 将来に向けての課題と展望

5.1 システム提案

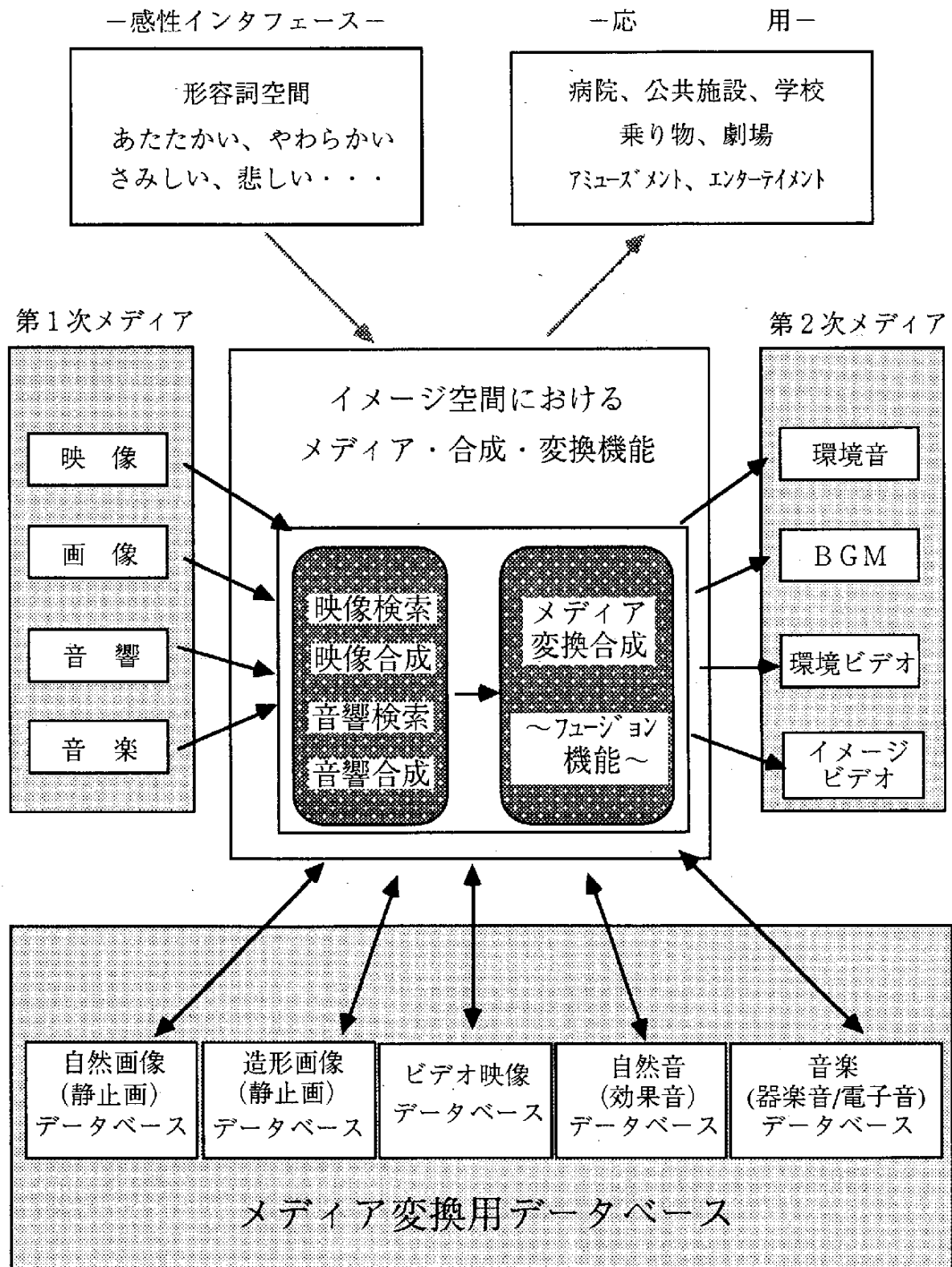


図5-1 メディア変換型感性データベースの概念図

第4章では、メディア変換型感性データベースを構成するための以下の4つの要素を取り上げた。

- ・画像から感性情報を抽出する。
- ・感性情報から画像を生成する。
- ・音楽から感性情報を抽出する。
- ・感性情報から音楽を生成する。

これらを同一の形容詞空間上で結合したものが、本研究で提唱するところのメディア変換型感性データベースであり、その概念を図5-1に示した。

これらの具体機能としては、次のようなことが考えられる。

(1) 画像に相応しい自然音および音楽の検索

今回制作中の「画像および音響データベース」を用いて、提示画像の内容に相当しい自然音環境およびイメージ音楽の検索を行い、視覚・聴覚を融合するメディア環境を創出する。キーワードとしては「音の出るフォトアルバム」である。

このシステムを実現するためには、二つの方法が考えられる。

- イ. 画像に含まれる自然音の音源情報、例えば、海、小川、小鳥、自動車などを、画像データベースのヘッダーにシンボル情報として登録しておき、そのヘッダーを手がかりに、音響データベースから海の波、小川のせせらぎ、小鳥のさえずり、自動車の走行音などを引き出す方法である。
- ロ. キーカラー、空間周波数、人物像の有無といった画像から直接抽出可能な、物理情報、図形情報、パターン情報とイメージ語との関係を登録しておき、イメージ語を手がかりに音楽データベースからBGMを創出する方法である。

(2) 信号変換による画像から音環境の生成

画像に、具象、抽象があるように、音響の世界にも、具体音、造形音がある。さらに分類すると、シンボル情報に変換可能な音声、音源の理解できる自然音・器楽音、波形合成により具体音を模擬した合成音、具体音には関係なく単にアルゴリズムに合成された造形音（抽象音）などがある。（1）の音響データベースが、主に前2者を対象にしているのに対して、直接画像から得られるパラメータにより、合成音・造形音を生成するメディア変換を行う。特に、（1）ロ. に示した物理情報をキーとした抽象的音環境に有効である。

(3) 音響情報に相応しい画像・映像の検索提示

音響情報に相応しい画像・映像の提示には、さまざまな目的があるが、大別すると以下のようなになる。

イ. 音の知覚・認識を助けるための視覚情報の提示

音響情報は、画像や映像と言った視覚情報に比べて「一過性である。主観的である。」などの特徴を有する。従って、不明瞭な音から意味のある情報を知覚するためには、視覚情報を提示し、カクテルパーティ効果や音響パターンの可視化を積極的に利用することが考えられる。これには種々の音響パラメータの中から、抽出すべき情報に相応しいパラメータを選択的に提示することが望まれる。これらの情報変換機能（プログラム群）をデータベース化し、目的に応じて利用する。

ロ. 音楽情景の心理効果を高めるための視覚情報の提示

音楽は、極めて主観的であるが故に、音楽に相応しい情景を視覚的に提示することにより、感性・感情の抑揚を刺激できる。具体例としては、カラオケにおけるビデオ映像を挙げることができる。悲しい失恋の歌に、ハワイのワイキキ海岸のビキニ女性の映像を重ねるわけにはいかない。多くの映像データベースとビデオエフェクタを併用することにより、音楽に相応しい映像シーンの生成を行うことができる。

5. 2 これからの課題

前節では、本研究が目指すべきメディア変換型感性データベースのあるべき姿について述べたが、現状においてはまだまだ課題も多く、一朝一夕にこれを実現するのは困難である。本節では、それらの課題について検討する。

5. 2. 1 要素技術における課題

(1) 扱う対象の制限

第4章で述べたシステムは、いずれも扱う対象がかなり限定されたものであった。将来的には、このような制限が可能なかぎり緩和されることが望ましいが、現状では、そのような広範な対象を統一的に取り扱う手法は確立されていない。また、無理に統一的に処理しようとすれば、それぞれの対象の特性を損なう可能性もある。よって、まずは現在のようなある一定の制限下でもって、各対象をより自然に無理なくハンドリングできるように、各手法の研究を進めていくことが肝要であると思われる。

(2) 処理過程構築の効率化

現状では、各対象を形容詞空間にマッピングするためには、やはり人間の分析能力に頼っている部分が多いと言わざるを得ない。小規模なデータベースならばともかく、ネットワーク規模で展開する大規模なものを実現するには、この部分の効率化が不可欠である。

5. 2. 2 システムとしての課題

(1) グローバルな形容詞空間の構築

このデータベースは、感性をキーワードとして、音楽と画像を同一のグローバルな形容詞空間にマッピングすることにより、メディアの壁を超えた統合的環境を目指すものである。現状では、それぞれの画像や音の情報がマッピングされている形容詞空間は、まだ重なりに乏しくほとんど独立に近い形で存在しており、これらを含括するものを構築する必要がある。ただしその実現には、要素技術の進歩を待つところが多く、その意味からすると、現在の急務は各要素技術をまず充実することにあると言わざるをえない。

(2) 公共的感性と個性的感性の共存

一口に感性と言っても、大衆が感じる感性（公共的感性）は客観的と言ってもいいほど普遍性があり、知識の延長線上に位置するものとも考えられる。一方、芸術家の持つ感性のような特定個人の主観のみにもとづく感性（個性的感性）は、そのような公共性や客観性のない情報である。これらは、ともに重要であり、感性データベース上ではうまく共存させられなければならない。そのためには、例えば前節で少し触れたような、自分の個性や独自の技法を収めた個性プロフィールといったものの導入により、利用者の好みによりカスタマイズ可能なシステムにする必要がある。

参考文献

井口征士：「感性情報処理が目指すもの」，情報処理，Vol.35，No.9 (1994)

————— 禁 無 断 転 載 —————

平成7年3月発行

- 発 行 財団法人 データベース振興センター
東京都港区浜松町二丁目4番1号
世界貿易センタービル7階
TEL 03-3459-8581
- 委託先 財団法人 イメージ情報科学研究所
大阪府大阪市北区堂島二丁目2番28号
泉谷堂島ビル2階
TEL 06-347-0641
- 印刷所 有限会社 ダイヤ印刷
大阪府大阪市西区本田一丁目6番25号
TEL 06-584-1580

