

ヒューマンインタフェース環境の演出的構築法

大場章男†、桜木恵子、渡辺敏郎

ソニー 総合研究所

テレビ番組や映画に代表される動画像メディアでは、カメラワークや編集技法など種々の表現様式が確立されており、映像メタファとして利用されている。我々はインターフェイス環境の構築にこれらの表現様式を利用することを考え、我々が開発した動画像一覧表示技術：Video Browser を用いてテレビ番組の演出構造の抽出実験を試みた。実験からテレビ番組メタファとして利用できる枠組みとH I 環境構築に有効な演出技法の知見が得られた。また、今後の展開として対話的な演出構造の抽出、利用技術：Macro Browser を提案する。

Stagecraft on A Human Interface Environment

Akio OHBA, Keiko SAKURAGI, Toshiro C. WATANABE

Corporate Research Laboratories, Sony Corporation.

6-7-35, Kitashinagawa, Shinagawa-ku, Tokyo 141 Japan

In the media based on dynamic visual representation, such as TV programs and movies, some expression styles in camera working and editing are established, which are used as an image sequence metaphors. We have suggested to apply such expression styles for human interface design in computers; our first trial was an experiment of extracting production structure in typical TV programs by using the Video Browser, which is a method of generating and showing a video index of a TV program.

In this paper, we describe some results of the experiment, which are useful for human interface design. Then, we propose the Macro Browser, which is a technique of utilizing the production structure or the stagecraft in an interactive information environment.

† 現所属：ソニー（株）コンピュータ・エンタテインメント事業準備室

1. はじめに

メタファは人間の認知理解能力を特徴づける修辭学的技法であり、コンピュータのシステムイメージをユーザに伝達する技術として広く応用されている。コンピュータのユーザインタフェースに用いられるメタファは、「小さなメタファ」と「大きなメタファ」とに大別することが出来る。

「小さなメタファ」とは、特定の機能やデータの特徴を単純な図像で表現したアイコンなどであり、言語でいえば、語や文のレベルで意味を伝達する。「大きなメタファ」とは、デスクトップメタファのように枠組みとなる世界を設定してその世界全体から対象の性質や出来事の意味を推定させるもので、言語でいえば、言述や作品のレベルで意味を伝達する。前者を狭い意味でのメタファとすれば、後者は修辭学でいうアレゴリーに相当する。本報告では、コンピュータのH Iをアレゴリカルに構成された情報環境と見なす立場をとる [1]。

さて、テレビ番組や映画に代表される動画像メディアではカメラワークや編集技法など種々の表現のための手法が確立されている。また、これらの手法を用いた映像の表現様式はユーザにとっても身近であるので、インタフェース環境を構築し映像として表現する時にこれらの様式を利用することは有益である。

本報告では、我々が開発した動画像一覧表示技術：Video Browser を用いてテレビ番組の演出構造の抽出実験を行なった。実験から得られたテレビ番組メタファとして利用できる枠組みとH I環境構築に有効な演出技法の知見について述べた後、今後の展開として対話的な演出構造の抽出、利用技術：Macro Browser を提案する。

2. H I環境の構築法

メタファを用いたインターフェイス環境は一種の世界モデルを持ち、日常経験に対応した仮想現実環境と考えることができる。ユーザはこの仮想現実環境の中でメタファで表された機能やタスクを日常経験になぞらえて理解し、操作することになる。この仮想現実環境の構築法にはHMDを用いたVRシステムに代表される実体的な構築法と、ここで述べる演出的構築法とに分けられる。

実体的構築法においてはシステムの機能をユーザが想起するメタファにより近付けてミスマッチを防ぎ、自由度を上げるように設計される。この方法ではモデル全体の自由度が多く、実体の動作

に近付けるため計算コストが大きく、制御も困難になる。

演出的構築法はシナリオを主体に環境の自由度を制限し、ユーザを誘導していく方法である。この方法は自由度が制限されているだけ計算コストも小さく効率がよく、ユーザも理解しやすい。

我々は将来の家庭内に入るより広範なユーザを対象にするH I環境には演出的な構築技術が重要であると考え、特に動画像表現における演出技術を中心に研究することにした。テレビ番組ではシンボライズされた道具立てや設定を用いたシナリオにより、非常に短い時間で様々な状況設定を効率的に視聴者に認識させている。我々は、テレビ番組を解析してそこで用いられている演出構造を抽出し、その枠組みを基本フォーマットとしてH I環境の設計に利用することを試みた。

以下に対話型動画像解析法を用いた演出構造の抽出実験について説明する。

3. 対話型動画像解析法

我々は比較的安価な装置で大量の動画像データを処理し、解析することができる二つの対話型動画像解析法を開発し、テレビ番組や映画などの動画像メディアからの演出構造の抽出実験や、視聴者観測に適用した。以下にその原理と用法について説明する。

3.1. Video Browser

まず、テレビ番組や映画から演出構造を抽出するのに用いたビデオブラウザ：Video Browser と呼ぶ動画像一覧表示技術[2][3]について説明する。

動画像情報は2次元の画像情報が時間軸方向に並んだ図1-aのような直方体モデルで表すことができる。Video Browser では図1-bのように動画像の直方体モデルを一定時間間隔で斜めに切断し、斜めの切断面にあたる画像情報を図1-cのように並べて、帯状に表示することにより、Video Index を生成する。斜め切断面を生成するために、Video Browser では縦スリット状のサンプリング窓を一定時間間隔で左から右に等速度で移動し、動画像から画像データをサンプルしてIndex 画像を生成している。そのためIndex 上では時間の流れが帯の方向（左から右）へ連続的に表現され、シーンの長さやシーンチェンジの時刻等の時間情報をVideo Index 上の長さや位置とし

て正確に読み取ることができる。また、動画画像情報内の静止画像成分はそのまま再現されるため正確な時間指標のもとに映像の概略も把握することができる。

図2に Video Browser による Video Index の実例を示す。この例では1スライスが5秒、1行で1分、全体で2分のビデオ動画を概観していることになる。インデックス下部の白い波形は音声波形で Video Index の時間軸に正確に対応している。シーンチェンジのタイミングや音声のバースト波形などから映像と音声の同期関係を見ることができる。Video Index 上では動きは歪みとなって表示される。

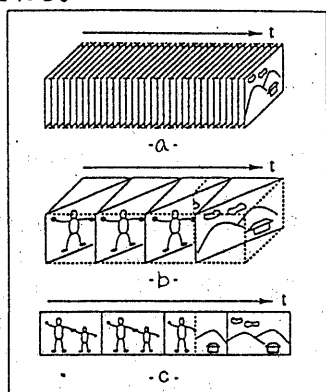


図1 ビデオブラウザの原理

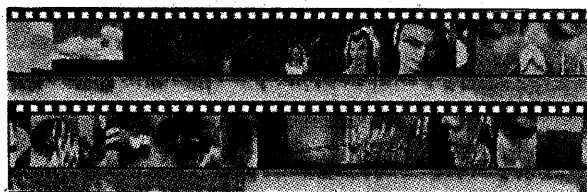


図2 Video Index 実例

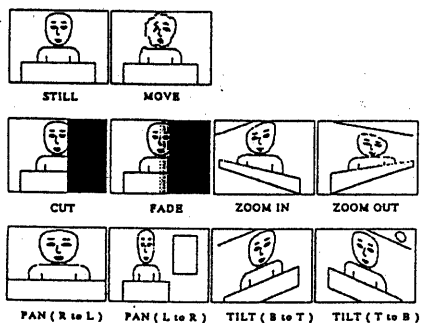


図3 Video Browser の見方

図3は特徴的なシーケンスでの歪みの例を表した図である。ズームイン、ズームアウト、パン、スクロール、カットイン、フェードイン等のカメラワークや編集操作はこのように画像の歪みとして読むことができる。

3.2. Slit Scanner

スリットスキャナ[3][4]はビデオブラウザがシーンの移り変わりや時間配分、内容の概略といった文脈レベルの解析に用いるのに対して、一つのシーン内の部分的な動きを抽出する文節レベルの解析手法である。ビデオブラウザと同様にスリット状の窓からサンプルしたデータを展開し、解析用画像を生成する。(図4参照)

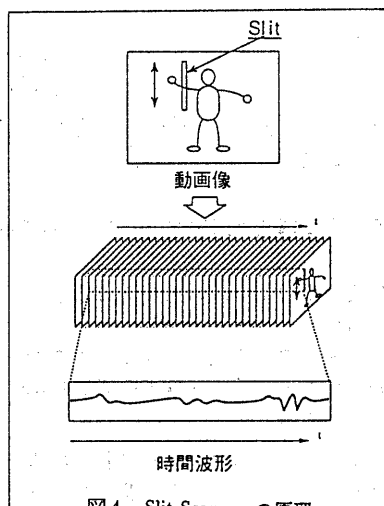


図4 Slit Scanner の原理

スリットは対話的に対象画像上に設定され、解析されるシーン内で固定されている。スリット上の画像がフレームごとにサンプルされ、解析画面上の時間軸に展開されるため、簡単な画像処理によりスリット上の動きを時間波形に変換できる。

当初、手振りや身振り等のノンバーバルな慣用表現の動き抽出を想定していたが、本報告では脈拍等の生理指標と併用して視聴者（ユーザ）観測に用いる。

4. 演出構造の抽出実験

図5に演出構造の抽出実験[4]に使用したシステム構成を示す。システムはビデオ入力ボードを持つグラフィックワークステーションを中心に各外部機器がネットワークで接続された構成になっている。このシステムでは、対象となる動画画は

一旦、VTRに録画されてから解析される。VTR及び視聴者観測用カムコーダ（後述）はビデオ制御インターフェイス（Vbox）を介してワークステーションから制御できる。上記の Video Browser や Slit Scanner はワークステーション上で実行される。

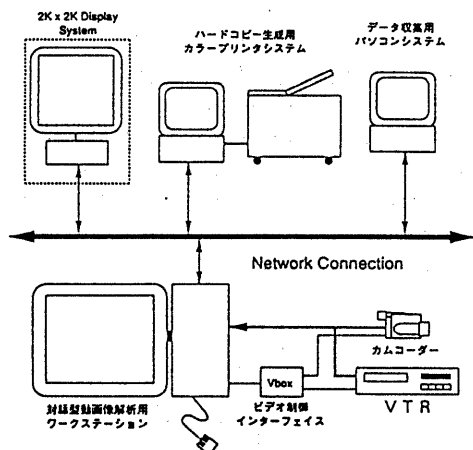


図5 演出構造抽出実験用のシステム構成

実験者はGUIを用いることにより、VTRの制御、Video Indexの生成、Video Indexをもとにしたシーンの抜き出し、スリットの設定、動きの時間波形の生成等の一連の作業を対話的に行うことができる。

ネットワークには他にデータ収集用のパソコンとカラープリンタシステム、及び2000x2000の解像度を持つ超高解像度カラーディスプレイシステムが接続されている。パソコンシステムは音声や脈拍、脳波などの時間波形データの収集に使用し、プリンタシステムと超高解像度カラーディスプレイシステムは通常のディスプレイでは一覧表示できない長時間にわたる映画などのVideo Indexの一覧表示に用いる。以下に本システムを用いた演出構造の抽出実験について述べる。

4.1. 抽出実験 - Video Browserを用いた番組調査

HI環境の設計において情報をユーザーにいかなる様式で提示するかが重要な問題となってきた。また、新聞やTVニュースといったメタファを用いて情報を提示する手法もいくつか研究されている。

我々はまず一般的な情報提供番組であるニュース番組とクイズ番組について演出構造の抽出実験を行ない、その構成と働きについて分析を行なった。またテレビ番組に比べ、より密度の濃い演出が行なわれていると思われる映画作品に対しても同様の実験を試みた。以下に実験内容と実験から得られた知見について述べる。

(1) ニュース番組

代表的な2つのニュース番組の同日の放映についてVideo Indexを作成し、その構成について比較分析した。採用した番組は番組A：平日9～10時までの60分番組（NHK）と番組B：平日10～11時20分までの80分番組（民放）である。

まず、両番組の共通点について述べる。両番組ともスーツ姿のキャスター（アナウンサー）のバーストショット中心に取材映像、特派員報告、図表の使用等、基本となるシーンの映像構成・様式に大きな隔たりはなく、トップニュース・スポーツ・ニュース・天気予報といった共通の枠組みも見ることができた。共通の構成要素をもとにVideo Index上で両者を比較して見ると番組Aは番組Bに比べ、各シーンが長くなっていることがわかった。これは番組Aがある程度まとまった解説を試みているのに対して番組Bはどちらかと言えば散文的なコメントが多いと言ったメインキャスターの差異もあるが、(a) 番組AがBに比べ1時間放映時間が先であり準備時間が短いことと、(b) 番組Bが視聴者に分かり易いニュースを目指しているのに対して、番組Aがニュートラルな報道姿勢とキャスターの能力に重きを置いていること等の編集方針を反映していると思われる。

両番組ともリアルタイムである特派員報告のシーンが長くなっていることから素材に近い加工されていないシーン（分節）ほど長くなることが読み取れた。カット割が多く、良く編集された構成の番組Bがテンポが早くわかりやすくなっているが、じっくり分析しながら見る視聴者にはかえって煩雑になり、素材に近い冗長性のある映像の方が有効になると思われる。

(2) クイズ番組

次に我々はエンタテインメント性のある情報提示番組と考えられるクイズ番組を対象映像に選び、その構成と演出の枠組みを生成したVideo Indexをもとに調べた。実験に用いた番組は米国製の幼児番組内の小番組（全3min 30sec）と日本製の

動物の生態を題材にしたクイズ番組（全 60min）である。クイズ番組はその構成法、演出法がかなり様式化され、共通のフォーマットを持つ番組も多い。実験に用いた二つの番組はほぼ同様のフォーマットで構成されていた。以下に両番組をプロット、ステージとカメラ割、キャストイングの各カテゴリーから見た構成の共通性と働きについて述べる。

プロット： Video Index から両番組とも司会者の登場－解答者紹介で始まるオープニング、出題ゲストの登場－出題－各解答者の解答で展開する中盤、出題ゲストによる正解の発表へと続く後半と共通の構成を持っていることがわかった。司会者登場から始まるオープニングはショウ番組の典型的な構成であり、視聴者の状況認識を助けている。両番組とも各解答者がそれぞれ特徴ある解答をし、その間に視聴者自身も考えるという構成になっている。

各解答者がいかに特徴的な解答するかというエンタテイメントの部分と、クイズを介した情報提供の部分とが両番組のプロットの中心になっている。

カメラ割とステージ： カメラの、被写体の画面サイズ、カメラワーク、スイッチング等の決定は「カメラ割」と呼ばれる。テレビ番組ではテレビジョンシステムの物理的な制約（小画面、低解像度）をこのカメラ割によりカバーし、同時に演出に基づいた物理的な視点設定をも行なっている。Video Index の歪みから両番組のズームイン、パンといったカメラワークが同じ状況で用いられていることがわかった。両番組ともズームインが多用されていた。ズームインは全体（ステージ）の中での被写体の位置や関連を表現するのに有効で、視聴者の視点位置認識を容易にしていると考えられる。番組の中では特に解答者の解答シーンでズームインが多用され、解答者の位置づけを明確にしている。

両番組のステージも司会者、解答者、出題者、メインカメラ等の位置が同じになっており、プロット、カメラワーク、とともにステージも共通の枠組みで作られていることがわかった。

キャストイング： 調査した二つの番組のキャストはともに、司会者、クイズを出題するゲスト、解答者（複数）で構成されている。この構成での特徴は解答者が両番組とも知名度が高く、性格付けのはっきりしたキャラクターを割り当てている点である。ここでの解答者の役割は正解を出すことではなく、各解答者が持つキャラクターを反映

した視点からの「らしい」解答をすることにある。視聴者の物理的視点は基本的にステージをみる観客の設定になっているが、比較的理解しやすい立場から解答を出す解答者の視点を介して対象となるクイズを見ることがキャストイングの構成の中で意図されていると考えられる。

「プロット」、「ステージ」、「キャスト」が、それぞれ時間的、空間的、社会的な基本的な構成を示す枠組みになっている。クイズ番組はこれらの枠組みのなかでシナリオやカメラ割りにみられる様式化した演出技法を用い、視聴者の社会的、空間的な視点設定を行っていると考えられる。二つのクイズ番組に見られる基本的枠組みと演出技法の共通性は、それらがテレビ文化のなかで規格化された道具であり、基本フォーマットとして再利用可能な資源となっていることを示している。

(3) 映画

映画作品は大画面であり、制作費、制作日数ともにTV番組に比べて多く、演出構造もTV作品とは違った特徴があると期待される。我々は1991年製の米国映画（全 138min）の Video Index を作成し、その構造を調べた。（図6参照）

Video Index は1スライスが10秒、1行で4分、全体で34.5行となった。データ量は1200 x 1800 画素となり、A3版のカラーハードコピーを用いて一覽し、評価した。調査した映画がアクションシーンの多いSF映画であり、シリーズ第2作でもあるため、最初から非常にテンポの早いアクションシーンが続き、Video Index 上ではそれらのシーンがほとんど静止画成分を再構成しない歪みとして表現された。全体を概観すると赤色の部分が点在する青色を基調にした歪みの多い部分と黄緑色を基調にした歪みの少ない部分とにセグメント化できた。点在する赤色は炎や爆発シーンであり、これらを包含する青色を基調にした部分は動きの早いアクションシーンであった。黄緑色の部分は中盤に現われ、前半後半のアクションシーンの間をつなぎストーリーの背景を説明する部分となっており、スローテンポの画面構成になっている。

Video Index の解析から各シーンにおいて画面の色調が統一されていることがわかった。特にアクションシーンでの青色を基調にした色調は炎や爆発シーンをより強調する効果があり、映画の制作における色調制御が演出構造において重要な働

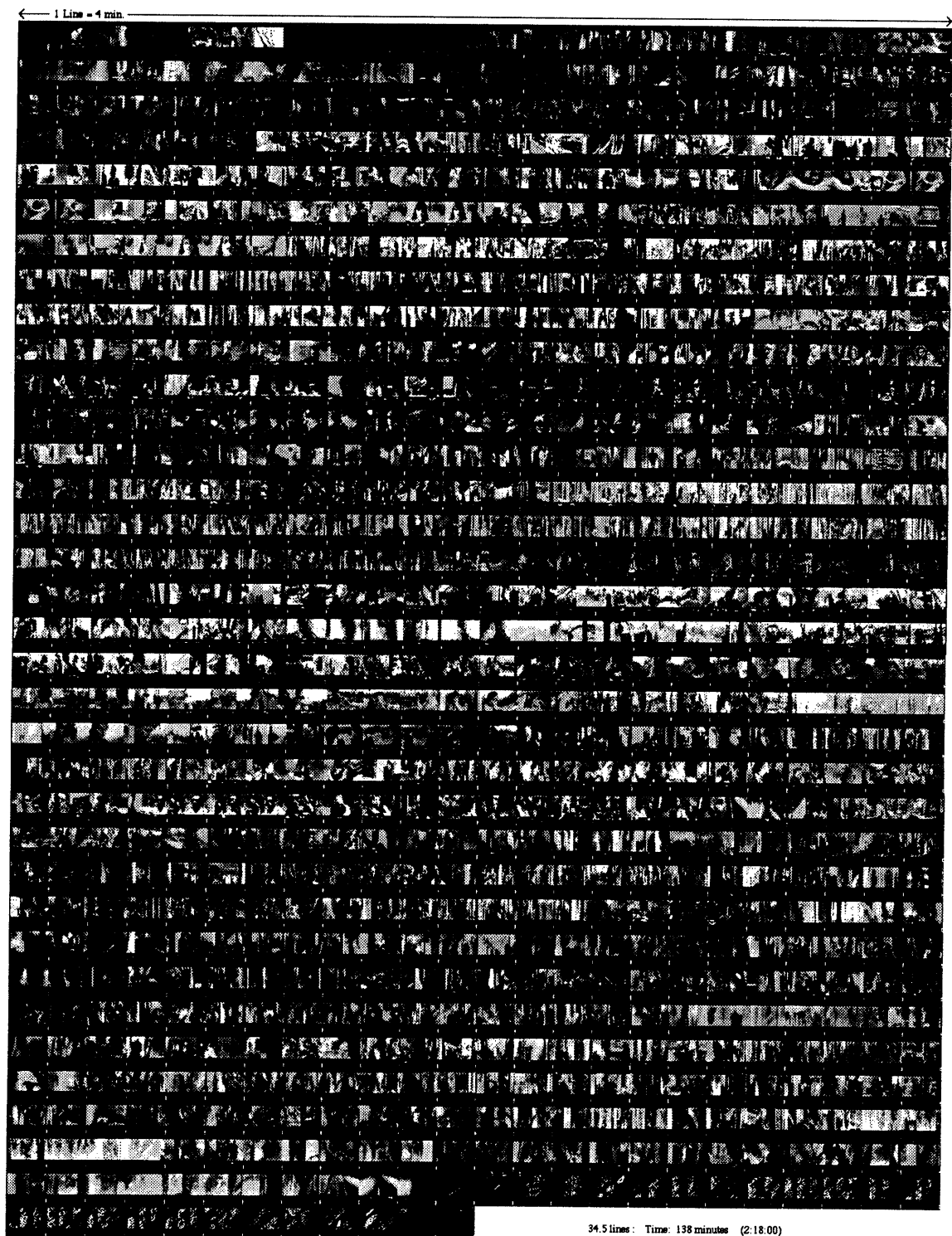


図6 映画作品の Video Index

きをしていることがわかった。Window Systemでは Window のカラーデザインをユーザに委ねている場合が多い。映画における色調の演出的制御はインターフェイス環境の能動的な色調制御の有効性を示唆している。

4.2. 演出構造に対応したユーザー観測

演出構造の抽出実験では Video Index 上で各シーンをセグメント化し、そのセグメントに対応する主観評価によって演出構造を読み取っている。しかし、演出構造によって視聴者が実際にどのように反応しているのか、また、それらの反応はどのタイミングで起きているのか、等を知ることができない。そこで、我々は実際の視聴者の反応を観測するために図7に示す測定システムと処理フローを構成した。

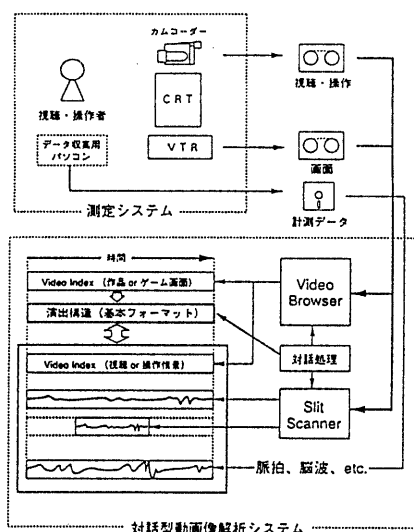


図7 ユーザー観測システムの処理データの流れ

測定システムでは動画情報提示と共に視聴者の反応をカムコーダーによって録画したり、脈拍、脳波などの生理指標を実時間情報と共にパソコンへ記録する。実験システムでは、視聴者の反応を録画した動画画像から Video Browser や Slit Scanner を用いて Video Index や動作の時間波形を作成し、測定した生理指標と共に提示動画画像の Video Index と対比し、演出構造に対応した視聴者の反応を観測する。

図8は提示動画画像を見ている2人の視聴者のまばたきを Slit Scanner を用いて時間波形に展開した例である。これらの波形データは生理指標デー

タと共に視聴者の反応として利用することができる。視聴者反応の観測実験は現在はまだ予備実験の段階であり、提示動画画像に対応したデータは今後の課題である。

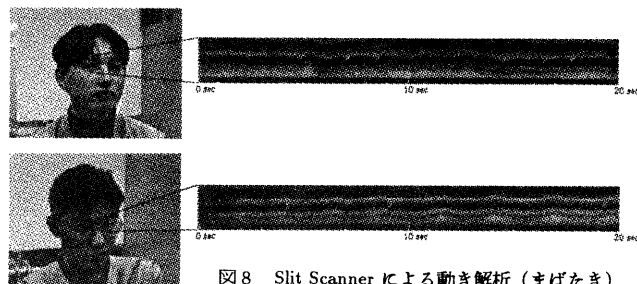


図8 Slit Scanner による動き解析（まばたき）

5. 演出構造の利用、対話型演出構造への展開

我々は演出構造の抽出実験からいくつかのニュース番組やクイズ番組を解析し、その全体的な枠組みや映像シーケンスの配置などを考察し、その共通点や相違点を見いだした。同じカテゴリーの番組には規格化された共通の演出的枠組みがあり、その共通の演出構造はテレビ番組メタファとして視聴者の状況認識を助けていることがわかった。共通の演出的枠組みは基本フォーマットとしてテレビ文化のなかで資源として相互に利用され、番組制作における重要な表現技術となっている[5]。

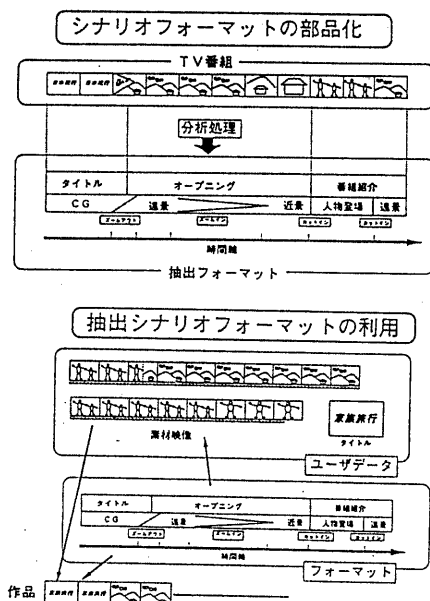


図9 抽出したシナリオフォーマットの応用例

Video Browser を用いて抽出した時間的な演出構造はフォーマットとして動画表現に利用することができる。図9はテレビ番組から抽出したフォーマットを家庭用ビデオ作品の作成に応用する例を示している。例では抽出したフォーマットを下敷きに対応する素材映像を選択し、編集処理している。Video Browser を用いた演出構造の抽出ではテレビ番組や映画のような一方の情報提示メディアの演出構造を解析できる。

しかし、我々が目指すH I 環境はパソコンやゲーム機に代表される対話的な情報環境であり、対話的な演出構造が必要になる。図10は対話的な演出構造の作成の為に対話的な情報環境から対話構造を抽出するマクロブラウザ：Macro Browser の概念図である。Macro Browser はコマンド列とそのタイミングや表示画面、システムの状態を Video Browser と同様に時間軸上に展開表示するブラウジングシステムである。

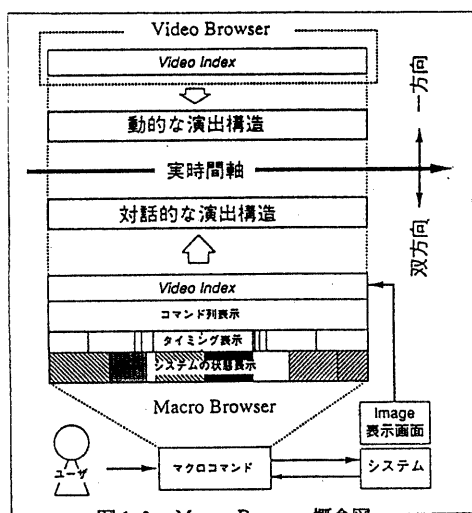


図10 Macro Browser 概念図

Macro Browser では、ユーザの操作履歴をファイルに記録し、その履歴をユーザの認識しやすいある一つのコマンド（これをマクロコマンドと呼ぶ）として定義する。マクロコマンドが実行されるとファイルからコマンド列が読みだされ、操作時と同じイベントが再現される。その際にユーザは、時間軸上に展開された実行コマンド群と出力映像変化とシステムの内部状態の変化を一覧することができる。

この Macro Browser 機能を利用することによ

り、目的とするタスクのシナリオ（演出構造）をユーザが理解し、対話的に編集できるようになることが期待できる。試行錯誤的に作成したり、部分的なマクロを組み合わせて実行したりして得られた一連の操作は、そのまま一つのタスク記述であり、ユーザモデルであると考えられる。従って、このメカニズムを用いることにより、一般ユーザもH I 環境の演出的状況設定に直接参加することができて、個々のユーザに適した有効なタスクモデルを創造することができる。

6. おわりに

Video Browser を用いたテレビ番組の演出構造の抽出実験から、テレビ番組には共通する演出的枠組みがあり、テレビ番組メタファとしてH I 環境の演出的構築に利用できることを示した。また、Video Browser を対話的な演出構造の抽出ができるように拡張したマクロブラウザを提案し、対話的なH I 環境の演出的構築法への可能性を示した。

今後は、特定のタスクを選んで具体的なH I 環境を構築し、演出的状況をインタフェースとしたシステムにおけるユーザとタスクモデルとの相互作用を観察して、その有用性を立証する予定である。さらにその知見に基づいて、より一般化した情報環境の構築指針を得ることを目指している。

なお、本報告の一部は通産省のFRIEND 21プロジェクトの一環として実施されたものである。

参考文献

- [1] T.C.Watnabe: "Style and Viewpoint in Television Media Interface" International Symposium on Next Generation Human Interface (1991)
- [2] A.Ohba: "Interactive Video Image Coding & Decoding" Next Generation Human Interface Architecture pp41-46,(1990)
- [3] 大場章男: "動画像一覧表示技術: ビデオブラウザ" 1991年テレビジョン学会年次大会 pp219-220,(1991)
- [4] 大場章男、渡辺敏郎: "テレビジョン番組の動画像解析" FRIEND 21 第三回研究成果発表会 pp135-144,(1991)
- [5] 新井一: 「シナリオの技術」ダブィット社 1986