

異種複合データ取り扱い可能計算機 TOM³アーキテクチャ

坂村 健

(東京大学理学部情報科学科)

梗概 (Abstract)

本小論では、数値、文字、図形、画像、音声など表現形態が異なるデータを統一的に取り扱うことのできる計算機のアーキテクチャについて述べる。このような計算機はまだ存在せず、サーベイ等を行うことは出来ない。また系統立てて概観することも難しい。というよりはこの種の計算機の枠組は未だ与えられていない。個別の芽となるような要素技術に関する研究が各所で進められているのが現状である。そこで本論では、著者が現在デザイン中の異種複合データ取り扱い可能計算機TOM³を述べ、そのアーキテクチャを提示することにより異種複合データ取り扱い可能計算機のイメージを固めることにしたい。

1. はじめに

従来の計算機システムが取り扱うデータは、主として数値データやテキストであった。最近では、その他に図形や画像、あるいは音声データを取り扱えるようなシステムも登場してきているが、それらはあくまで個々の応用に専用化されており、これら性質の異なるデータを統一統合して取り扱うことのできる計算機は未だ存在しない。けれどもこれら表現形態や性質の異なる異種データを統一してとりあつかうことのできる計算機システムを作ることができれば、現在既に計算機が使われている分野においてもさらに有効に使えるようになる。それだけでなく新たな応用分野も開けてこよう。

また、逆に考えると応用の広がりがある現在の計算機に対して異種の情報の取り扱いを強く望んでいるようにも思える。とくにオフィス・オートメーション等を中心とする新しい応用分野がそれを要求している。また、その他プログラム開発エイドはもちろんのこと各種科学技術シミュレーション、対話型の教育及び訓練システム、エンターテイメント・システムなどその応用には限りがない(図1)。

以下、本論では異種データをメディアという立場から論じその性質を整理する。また異種のメデ

ィアの扱い方に焦点を当て統一的に扱うことの意味から始め、異種複合データ取り扱い可能計算機の定義を与える。その後このタイプの計算機のイメージを与えるため、現在デザイン中のTOM³(Tokyo university Multi-Media Machine、トモサンと読む)のマンマシン・インターフェース、およびそのアーキテクチャ・フィロソフィー、そして実現に当たっての考察について述べる。

2. 異種メディアの統一的处理

情報やデータの実体を伝達する手段を計算機でなく人間の側から見た場合、メディア(Media)という概念で考えることが多い。新聞、郵便、電話、放送といった制度的な角度からみたり、マスメディア、ミニメディア、個人メディアなど対象とする規模から分ける事もある。けれども、情報を受け取る主体はあくまでも人間なのでその感覚と認識をもとにメディアを分類すれば、視覚からはいつてくる映像や画像、聴覚からはいつてくる音声や音響、これらの認識の結果として生ずる言語などに分けることができる。これを計算機で取り扱うデータに置き換えれば人間が目で感じる図形データや画像データ、耳で感じる音声データ、音響データ、頭で感じる数値、文字データということになる。

これらの異種メディア総てを人間のように統一的に取り扱える計算機は原理的には実現可能である。例えば、画像、音響などのアナログ情報はデ

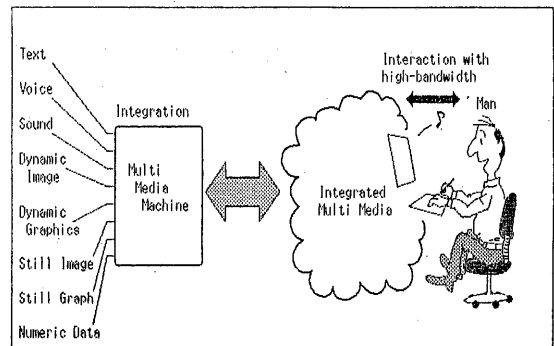


図1 TOM³

デジタルな形式に変換してしまえば計算機の内部での取り扱いは自由であるので統一的な取り扱いは理論的には可能といえる。しかしこのような計算機は現在存在しない。これは、必要性が認識されなかったことも一因ではあるが、技術的困難さによるところが大である。その困難さは、第一にもともと計算機で取り扱える情報が1、0のデジタルであるため、自然界の帯域幅の広いアナログな情報をとり扱うのは不得意であったこと。さらに近年の半導体技術や各種記録再生技術の著しい発達により将来量的な困難さは克服することができたとしても、統一的に取り扱うための方法論が全くわかっていないなどの質的な問題が山積していることなどによる。例えば、性質の異なるメディアに対してどのような内部表現、データ構造を取ればよいのか明らかでない。

ここで異種メディアを統一的に取り扱うとはどういうことかを考えておこう。人間が複数のメディアから情報を受け取る時は例えば目で数字、文字、図形、画像を一度に見て関連づけ、それを認識、理解する。それゆえ計算機の側で異種メディアの取り扱いをするためには例えば、画像とテキストとが関連をもって提示できる必要がある。

つまり、異種メディア間のいわゆる同期が重要である。とくに、人間が接するメディアは実世界の写像ゆえに空間の概念はもちろんのこと、動画像や音楽のように時間の概念（それも実時間）が重視されるものも多い。だから時間的な取り扱いに対し充分配慮しなければならない。また例えば音楽を楽譜にして表わすなど空間的概念と時間的な概念を変換して提供することが人間にとっては有益であることも多い。

ここで異種メディアの統一的な取り扱いを形式的に述べると、

異種メディアの実体を各々 $M1, M2, M3, \dots, Mn$ とすると、

1. 基本メディア演算 $OP1, OP2, OP3, \dots, OPk$ が $M1, M2, M3, \dots, Mn$ に適用可能で結果がもとの集合に含まれること。
 >>>例えば、MOVE, COPY, DELETE, CREATE, APPENDなどの操作がどのメディアに対しても適用でき意味のある結果を得られることなど。
2. 異種メディア間演算 $HOP1, HOP2, HOP3, \dots, HOPl$ が $\phi, M1, M2, M3, \dots, Mn$ の組み合わせに

表1 メディアとテクノロジーの対応

	入力	出力	処理	通信	認識	合成	蓄積
数値	テンキー	数値ディスプレイ		データ通信			ファイル データベース
文字 (テキスト)	Dvorakキーボード 変形キーボード 仮名漢字変換 2ストローク 連想入力	文字ディスプレイ	エディタ フォーマット トランスレータ	テレックス 電子メール	印刷文字認識 手書き文字認識	多重フォント生成	テキストファイル データベース 知識ベース
図形	デジタイザ タブレット ジョイスティック マウス	CRTディスプレイ レーザプリンタ	回転、移動 拡大、縮小	グラフメール	図形認識 手書き図形認識	コンピュータ・ グラフィックス	図形ファイル グラフィック・ データベース
画像	TVカメラ イメージスキャナ	テレビジョン CRTディスプレイ 3次元ディスプレイ	符号化 特徴抽出 雑音除去	ファクシミリ通信	画像認識 シーンアナリシス パターン認識一般 画像計測	画像生成 シーンシミュレーション	フレームメモリ 光ディスク 磁気ディスク 画像データベース
音声	マイクロフォン	スピーカ	エコーキャンセラ 声質変換 音声編集	電話 放送 音声電子メール	音声認識 話者認識 声紋検出	音声合成	テープレコード 音声ファイル 音声データベース
音響	マイクロフォン ピックアップ 楽譜	スピーカ 楽譜	残響処理 雑音除去 音楽編集、混合	放送	楽音認識 旋律認識	音響合成 音楽合成	テープレコード ディスクレコード コンパクトディスク

対して存在すること。

>>>例えば、APPEND SIMULTANEOUS (同時に出す)などの操作が動画像と音響などに対して出来ることなど。

3. メディア間変換演算COP1, COP2, COP3, ..., COP_nが存在すること。

>>>例えば、文字テキストを音声に変換して出力するなど。

4. 各メディアの持つ時間的属性と空間的属性を変換する時間空間変換演算TSOPが存在すること。

>>>例えば、動画を静止画の並びとして表示できることなど。

このような特性を兼ね備えた計算機を一般的に異種複合データ取り扱い可能計算機と呼ぶことにする。つまり、異種メディアを統一に取り扱うことが可能な計算機を意味する。先にも述べたように、完全な意味での異種複合データ取り扱い可能計算機は存在しない。しかし、芽となるような個別の要素技術は現在実用化されているもの、研究のものなどいくつかはある。表1にメディアとテクノロジーの対応例を示す。また、それらの要素技術を用いてシステムを構築する部分的な実験のようなものは例えば基礎的な研究では、

- Xerox PARCのLearning Research Groupが70年代一杯をかけて行なったパーソナル・ダイナミック・メディア[1]に関する研究。
- MITのArchitecture Machine Groupが行っている一連のメディア・テクノロジーやマンマシン・インターフェースに関する研究。例えば、Spatial Database Management System [2]やMovie Map[3]など。

また、応用システム構築の研究では、

- Brown大学の動的対話型ドキュメントシステム[4]の研究。
- Computer Corporation of America社のView System[5] (実用的なSpatial Database Management System)。

などがある。また、一部の成果を元に

- Scitex社 (イスラエル) のVista 雑誌レイアウト編集システム
- DEC社のIVISマルチ・メディアシステム等が商用化されている。

以下でイメージを与えて行くTOM³のアーキテクチャは、このような部分的なシステムと違い、異種メディアの完全な統一化をシミュレートする

ための実験システムである。

3. 異種複合データ取り扱い可能計算機TOM³の概観 (アーキテクチャ・フィロソフィ)

異種メディアを統一して取り扱う計算機を利用者の側から眺めた場合は、従来の計算機と比べて①異種メディアが同時に融合して提示され、利用者にとって自然な形で情報を提供できる。②それに対して人間が反応することにより人間機械間の双方向インターフェースの著しい向上が図れる。といった点で大きく異なる。つまり、情報を人間にとって最も適した形で提供することにより帯域幅の広いマンマシン・インターフェース (マンマシン間で一定時間に多量の通信) が可能になるわけである。ここで、マンマシン・インターフェースの向上といった場合は①ユーザーにとって親しみ易い。②使用することにより疲労が蓄積したりせず快適な環境を提供する。 (必要以上の注意力を要しない) ③精神的にいらついたりすることがない。 (反応速度、提示速度が適正。操作や提示方法が合理的) ④一定時間により多くの仕事をこなすことができる。⑤仕事の間違い、ミスが減少する。などを指す。以下では統一化のイメージと実現の際の問題点およびTOM³で用いている方法論について述べる。

3.1 異種メディアの同時的な提示

各種の異種メディアを統一に取り扱ったマシンでは、テキスト、画像、図形、音響などが互いに同期してリアルタイムに出力される。利用者は各メディアの特性を生かして提示される情報を理解把握することができる。例えば、テキスト、画像、音響の併用で利用者はより早く深い理解が出来る。これは、システムの実現の上からは基本的に計算機自体が高速に動作し利用者が快適に感じる位の反応速度を保証する必要がある。とくに動画や音響では重要である。

さらに、実現の際のソフトウェア的なポイントとしては、異種メディアをどのようにしてお互いに同期させて提示するかという問題が大きい。例えば、素材の文章と動画像そして音響をどのように出力すれば利用者に対して最大の効果を上げられるかということでこれは主としてアプリケーション側の問題である。マシン側としてはサポートのために何らかの異種メディア提示用のプログラミング言語を用意する必要がある。同様な意味で

異種のメディア同士の編集を助けるメディア・エディタも必需品である。このような言語やエディタは、メディアという実世界に近いものを扱うので一般のプログラミング言語と比べて極度に抽象化された対象を操作できる必要がある。また、実世界においては欠かせない空間的や時間的な概念についてもサポートの必要がある。

3.2 双方向インターフェースの向上

異種メディアの同時的な提示は高度ではあるがいわばマシン側から人間への一方向的な通信手段といえる。さらにこれに対し、利用者の側からの反応も異種メディア上で展開できるようにすれば、双方向でマンマシン・インターフェースのレベルを上げることができ人間機械間に広帯域の通信路が開かれる。

例えば、最近ではビットマップ・ディスプレイを用い、オーバーラップ付きのマルチ・ビューウィンドウをサポートする機能をもったワークステーションがある。これも数値、テキスト、図形などの融合を目指しているので広い意味では異種メディアの取り扱いを考慮したマシンと言えよう。しかし、マンマシン・インターフェースの向上はこれで総てというわけではない。マルチ・ウィンドウは限られた表示画面をより有効に使うための一手段を提供したに過ぎない。上記のようなワークステーションではアイコンという図形シンボルを導入しマウスなどのポインティング・デバイスを用いて実世界との比喻により指示を与える方式を取るものが多い。従来の文字による指示（コマンド等）に図形という異種メディアを導入したわけである。目新しさ（新規性）や取っ付きの良さ（速い習熟速度）は注目に値する。しかし、実際に使ってみればすぐわかるが、アイコンの注視を要するので眼精疲労がはげしく、マウスもポインティングに際し隔靴搔痒の感がありマンマシン・インターフェースとしてはまだまだ改良の余地があるように思われる。例えば、音声、音響のより積極的な活用が重要と思われる。

3.3 TOM³における方法論

ここで上記のような問題点を踏まえた上でデザイン中のシステムTOM³がどのように異種メディアを統合しているかを述べておく。

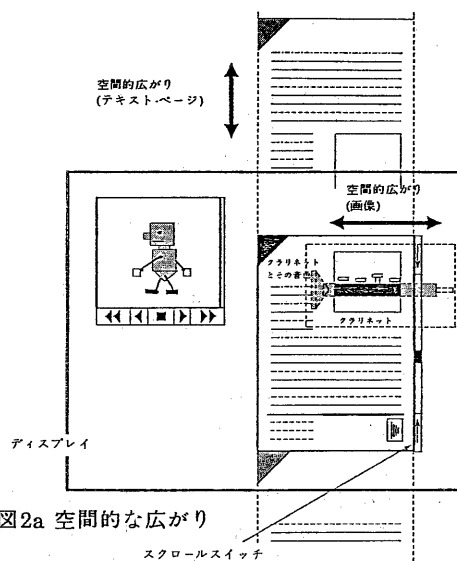


図2a 空間的な広がり

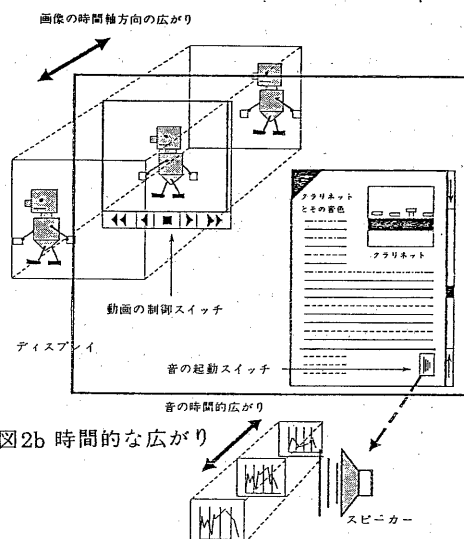


図2b 時間的な広がり

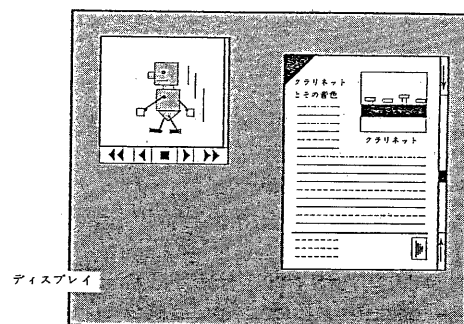
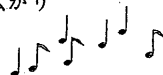


図2c 空間および時間的な広がり



I. TOM³におけるメディアの取り扱いと提示

TOM³は一応オフィスなどにおけるプロフェッショナルの仕事のサポートをまず応用例として考えているので扱うメディアのタイプとしては、① テキスト（文章一般）② グラフ（ビジネスグラフを主体とした図形）③ テーブル（表形式一般）④ イメージ（一般画像）⑤ サウンド（音響、音声）を基本的に想定している。

基本的な概念として、空間的な広がりや時間的な広がりやをサポートしているので利用者は実世界との対応のうえでマシンと対話することができる。空間的な広がりやを扱った例を図2aに、時間的な広がりやを扱った例を図2bに、両者が統合された姿を図2cに示す。定量的には示せないが利用者が計算機に相当自然な形で接することが可能になる。このような環境の実現のためTOM³ではメディアを空間的、時間的の属性をもったオブジェクトと考え、このオブジェクトを操作する高レベルのプログラミング言語を提供する。

II. TOM³におけるマンマシン・インターフェース

TOM³では対話型の双方向マンマシン・インターフェースの向上を目指した具体例として、手書きシンボルというメディアを導入したテキスト・エディタを備えている。例えば、修正前の図3a、タブレットからリアルタイムで書き入れられる手書きシンボルを用いて修正中の図3b、そして修正後の図3cを示す。テクニック的には手書き文字認識の技術を採用している。従来の紙というメディアでの校正手法に準じているため使い勝手は、定性的なシミュレートの結果ではアイコンとマウスによるエディタよりもかなり良いという結果が得られている。

III. 空間的概念と時間的概念の相互変換

人間のもつ時間的認識能力と空間的認識能力は性質が異なるので時間と空間をいれかえてやると理解の助けになる場合が少なくない。単純な例では、距離を示すのに100kmというより、車で1時間半といったほうが把握が容易なことが挙げられる。これは人間が空間を時間に置き直して理解しているためである。逆に時間を空間で表わしたほうが良い場合もある。TOM³では空間、時間変換の具体例として、動画像や音楽

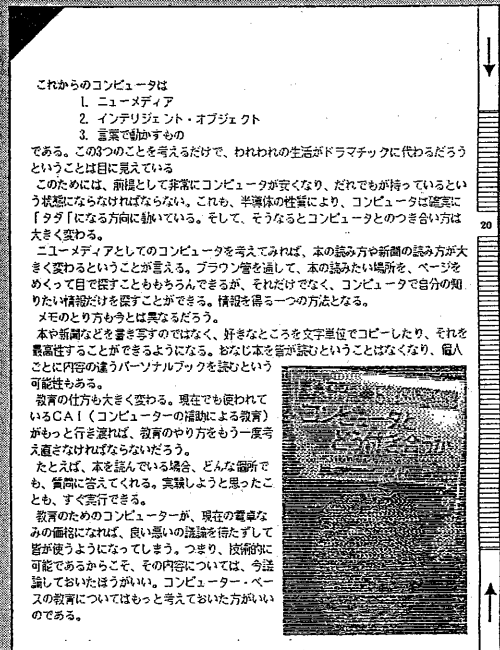


図3a TOM³の校正記号エディタ 修正前

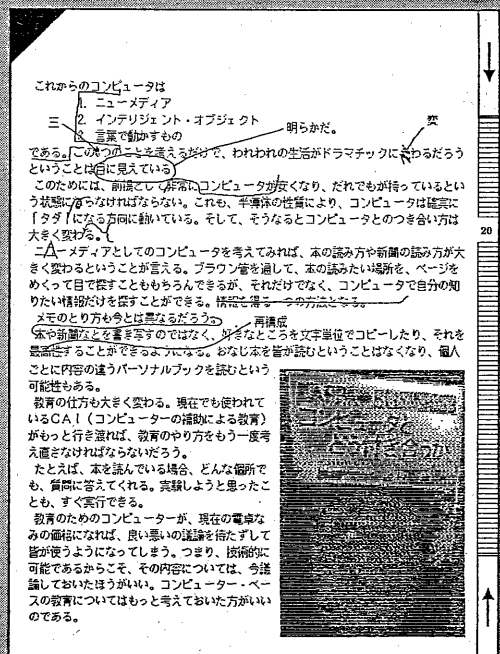


図3b TOM³の校正記号エディタ 修正中

など時間に支配されているメディアの編集を容易にするため時間的要素を空間的要素に変換したイメージ・エディタやミュージック・エディタの実験を行なっている。図4aと図4bにイメージ・エディタが動画像のシーン編集を行なっている例を示す。

4. 実現に当たっての考察

前章までの一般的な概念の構築から離れ、実際にTOM³をインプリメントするに当たっての考察についてここでは述べる。しかしインプリメントにおいては一般論と異なりニーズのみでなく、サイズも問題となる。つまり概念のインプリメントにおいて一意的に方法が決められない。ハードウェアコストを重く見る見方も有るし、ソフトウェア開発コストを重く見る見方もある。またパフォーマンスを考えても、速度を重視したり、処理のバリエーションを重視したりと千差万別である。これらの問題のバランスを取るのがアーキテクチャであるが、TOM³は考えるべきサイズのパラメータが通常のマシンの数倍有り、現時点ではまず実験マシンをインプリメントして評価法を確立する参考として使用している段階である。当然この考察は我々のインプリメント方針を反映するという束縛下においての物になってしまう。以下では出

来るかぎり考察の客観性を求めるため、またこの発表の性質を鑑み、実験の初期段階で既に表面に現われた問題点を中心にした。

4.1 フィジカル・ビュー

TOM³を実現するためのハードウェアの接続に関する、基本的なフィジカル・ビューが図5のようになる。

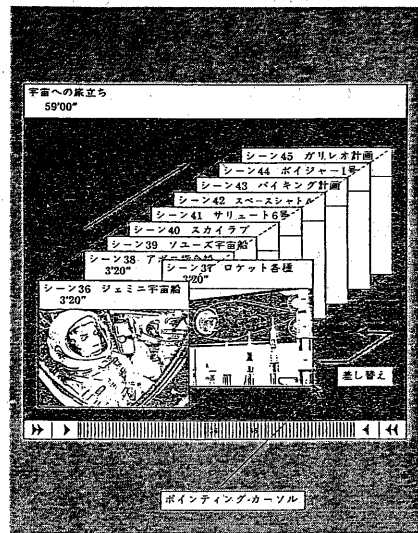


図4a イメージエディタのシーン編集例1

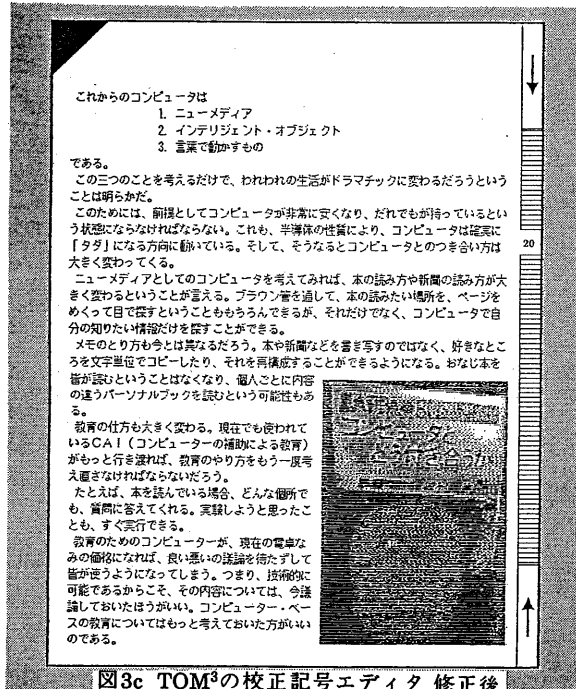


図4c TOM³の校正記号エディタ 修正後

図4b イメージエディタのシーン編集例2

ビデオなどの特に高周波数特性を要求されるメディアで特に処理が必要ないものはTOM³内部のスイッチングによってビデオ信号のまま処理する。TOM³の画面自体は、図6のようにグラフィック・メモリーからのビットマップ画像と、ビデオ画像をデジタル化したものを、TOM³の合成指示に従ってデジタル・ビデオ・ミキサーによって合成したものであり、画面に映るイメージのそのままを保持するフレームメモリーと呼べるような物は持っていない。イメージ上の矩形領域の高速転送と論理演算を行うBITBLT機能はグラフィック・メモリー内で使用され、ウィンドウ間及びビットマップとビデオの表示上の合成には合成指示に於ける画像間のプライオリティ指定と、ピクセルごとの合成指示アトリビュートが使用される。

図5 のようになると言っても、もちろん使用状況毎に各デバイスの利用度が異なるから、コストパフォーマンスを考えるとこのように可能なデバイスをすべて直接TOM³に接続するという事は、現実には考えられない。例えばスキャナでビットマップを読み取るという作業がそう頻繁でないならば、スキャナ・サーバーのようなスキャナとストレージ・デバイスのみを持つマシンを図7の様にマルチメディア・ローカル・ネットワーク（MLAN）につなぎ、複数の使用者が共用するほうがよい。

しかしシステムの柔軟な構築性から言っても、ロジカルな情報のロケーションとフィジカルなシ

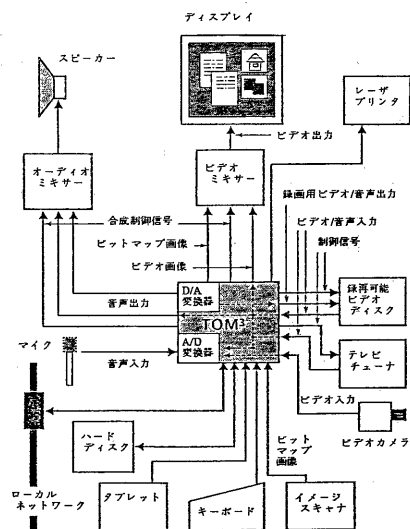


図5 TOM³の基本的フィジカルビュー

ステム構成が、できうるかぎり関係ない事が望ましい。このためにはMLAN全体に於けるオブジェクトのユニークさの確保が必要とされる。同時に、ワークロードの分散を考えてコストパフォーマンスを上げるにはどうしても任意の異なるメディアのサーバー・ソフトウェアを同一ハード上に乗せられることが必要だ。結論として、サーバーもTOM³でなければならない。

このことより、研究方針としてアーキテクチャー上マイナスになる面は有るがまず考えられる大体のメディアを処理できるマシンを構築し、それらの分散によるメリット/デメリットを評価する方式を確立する事とした。

4.2 ハード面の評価

図5 でも判るように、これらのデバイスの個々のコンピュータへの接続に関しては（そしてあるデバイスに限ればその個々の接続法に関しては）そのほとんどが既に使用例の有るもので、またいくつかに関しては確立された技術である。つまりパーツとして見るかぎり技術的には可能と言い切っていだろう。

しかしこれらをシステムとしてまとめ上げることは別の難しさを伴う。特に異なるメディア間に於て厳密な同期（人間の感覚に対する同期も含め）を必要とされるTOM³は、リアルタイム・マルチプロセスをある速度（遅いのは当然として速すぎてもいけない）で確実に処理する機能を持たな

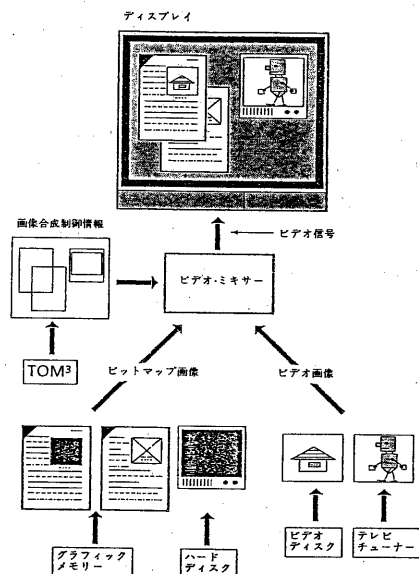


図6 TOM³におけるビットマップ画像とビデオ画像の合成

ればならない。同期用のバッファを介さない独立の信号線を確保するなどの対応が考えられる。

これに関しての議論はまだ行える段階に至っていないが、ハード面において以下のパラメータがパフォーマンス決定上の要となると言える。

(図8 表2 参照)

1. ストレージ・デバイスのデータ入出力平均速度及び分散
2. ストレージ・デバイスのデータシーク平均時間及び分散
3. データ伝送路の速度
4. MLANの平均速度及び分散(コリジョンの発生を考慮に入れる)
5. 画像メモリーの書き換え速度

4.3 ロジカル・ビュー

フィジカル・ビューに関連して、ロジカル・ビューが重要となる。TOM³のロジカル・ビューとしては、多種類あるメディアを上位レベルで同一一般命令で処理できる利点により、オブジェクト・オリエンテッドが上げられる。但しただのオブジェクト・オリエンテッドではここで要求されるような厳密な同期を取るには適していない。許容誤差なども含んだ実時間の概念を持ち、クロックリードや同期指定命令が使え、しめ切りによって実行環境が動的に変化するような、新しいオブジェクト・オリエンテッド・モデルが必要とされる。

一方TOM³のインプリメント時にこのロジカル・ビューをどこまで正確に写像するかについては、コストパフォーマンスの問題が有り慎重に考えなければならない。オブジェクト・オリエンテッドを使えば、TOM³の外部仕様からプログラムへの写像が直接的なためソフトウェア開発コストは下がるが、実行時に図9 のようなオブジェクト・テーブルを大量に操作する必要があり、このためにパフォーマンスが低下すると見込まれる。これについて現在はこれ以上議論するだけのデータを公表出来ないが、考慮すべき点としては以下のような事が上げられる。

1. オブジェクトとして扱う、データ構造上の最低レベル

例) 文書をオブジェクトへ素直に写像すれば最低構成要素は文字となる、しかし一文字について数十バイトを使ってオブジェクトとして扱うのは効率が悪いだろう。

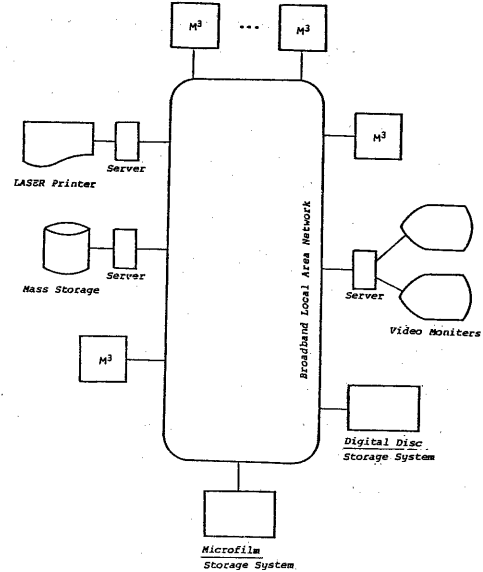


図7 MLAN

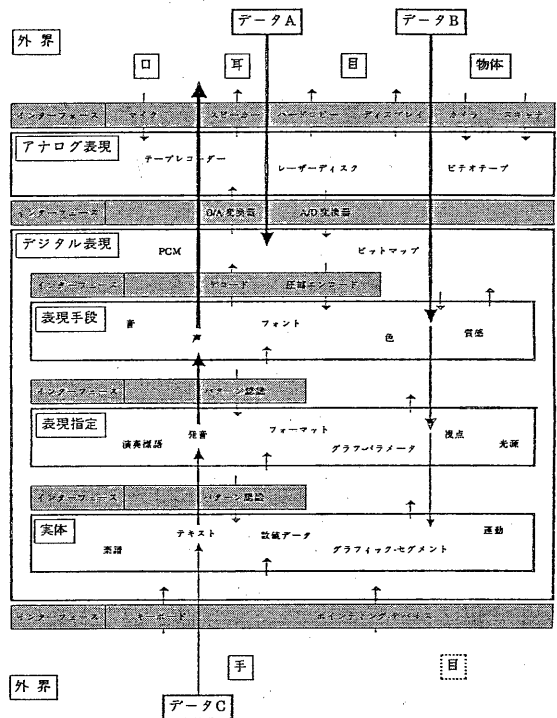


図8 TOM³の内部表現における抽象化の階層

表2 カラービデオ画像の場合のストレージ・トレードオフ

ハードメディア	アクセス時間	一画面の 転送時間	容量	一画面の コスト
アナログ 光ディスク (書き換え可能) (リム・パル)	1 sec (1)	1/30 sec (連続)	10000 画面 (ディスク 1枚当り)	1
磁気ディスク (固定)	30 msec (1/30)	1/2 sec (1MB/sec)	200 画面 (100MB/台)	1000
RAM	1 μ sec (1/1000000)	1/10 sec (5MB/sec)	2 画面 (1MB/フェイブル)	10000

・カラービデオ画像 (NTSC、4fscサンプル、8bit)、
1枚 (約0.5MB)

2. データによって最適の表現 (オブジェクト トへの分解) を決定できる評価法

例) 図10 の様に図形の性質によって、
適当な表現は異なる。空間周波数
などを指数としてこれを一意的に
決定できないか。

3. オブジェクト・テーブルの物理的位置の 最適分散

例) 図11 の様にデバイスの接続を考え
た場合、各デバイスでデータ本体
へのフィジカル・アドレスのとり
かたは全く違うと考えたほうがい
い。また一般的な操作に関する実
行ルーチンも完全にデバイス・デ
ィペンドな物になる。この場合デ
バイス・マネージャーとして適当
な I/O プロセッサを間にはさみ、
そのローカル・メモリーにデバイ
ス・ディペンドなオブジェクトに
関するテーブルを持つのが適当で
ある。

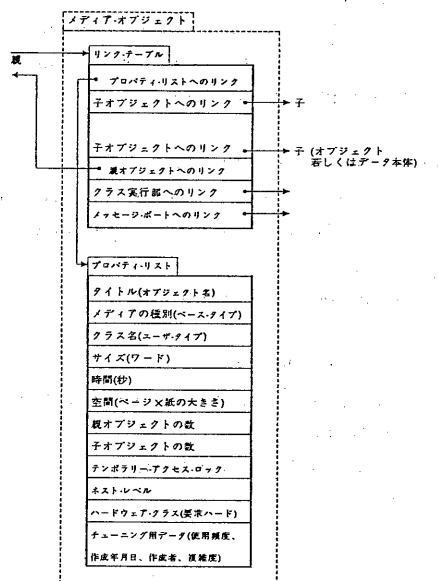


図9 TOM³のオブジェクト・ディスクリプタ

5. 終りに

実験構築中の異種複合データ取り扱い可能計算
機 TOM³ のアーキテクチャについて述べた。異種
複合データ取り扱い可能計算機は異種複合メデ
ィアが取り扱える計算機ということで、それだけで
従来の計算機とは全く異なる。また、その実現に
あたっては従来の技術の寄せ集めの考えでは無理
で、そのための新たなメカニズムの考案が重要で
あることが実験を通して判ってきている。

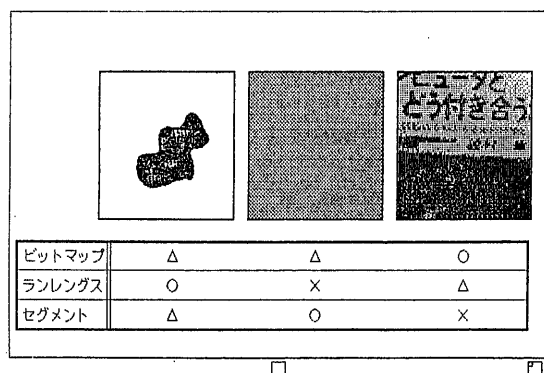


図10 オブジェクトのいくつかの表現法

参考文献

- [1] A.C.Kay and A.Goldberg:
"Personal Dynamic Media"
COMPUTER, March 1977
- [2] William C.Donelson:
"Spatial Management of Information"
Siggraph '78 Proceedings
- [3] Andrew Lippman:
"Movie-Maps: An Application of the
Optical Videodisc to Computer Graphics"
Siggraph '80 Proceedings
- [4] Steven Feiner et al.:
"An Integrated System for Creating and
Presenting Complex Computer-Based
Documents"
Siggraph '81 Proceedings
- [5] Mark Friedell et al.:
"Context-sensitive, Graphic Presentation
of Information"
Siggraph '81 Proceedings

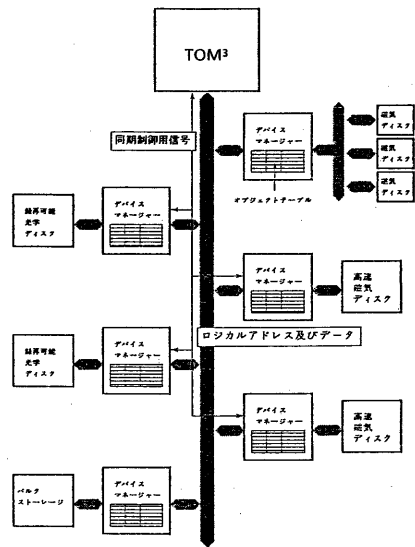


図11 デバイスマネージャ