

個別メディアの生成とその動的統合化

デモ 5

佐藤 剛

村尾 洋

榎本 肇

芝浦工業大学

1 はじめに

一般にマルチメディアによる作品やサービスを提供することは簡単ではない。

本論文では個別メディアとして、複数種の独立した音(音声、楽音)及び画像を取り上げ、音声、楽音を生成し、他のメディアである動画像との統合を行う効果的な方法を探求する。その際に、各個別メディアが相互かつ並列的に生成される必要がある。そこで、各メディアに時的制約として同期関係をクライアントが指定することで、時間的な調整、修正を行う。

2 個別メディアの生成

個別メディアの情報としては、複数種の独立した音(音声、楽音)と画像を用いる。画像分野は本研究室で開発されてきた拡張機能言語 Extensible WELL(Window-based E Laboration Language)により作成される動画像^[1]、音声分野としては、MIDI データを用いての楽器、WAVE データによる人や自然な音などの音声を使用する。また、それぞれのデータを作成する環境としては、画像はワークステーション(WS)、楽音、音声はパーソナルコンピュータ(PC)とし、それぞれを作成するのに適した分野ごとに分け、役割を持たせる。

● 動画像生成

動画像は WS 上で Extensible WELL を用いて作成されるが、生成法としては、人間の画像、動物、仏像などの静止画像はスキャナーで取り込む。取り込んだ静止画像に動きを表すパラメータを与えることによって変化を与える。変化させた画像を主要時点に置き、主要時点間の動きパラメータを補間することで一連の動画像を生成する。

● 音情報生成

音声、楽音データ情報の個別生成をおこなっていく際、音声である WAVE データは、PC 上で、サウンドカード等を用いて、あらかじめ録音しデータとして保存しておく。楽音である MIDI データは、誰もが簡単に作成できるわけではない為に、動画像データと、楽音データの作成とは異なるユーザが行う場合もありうる。その為、ネットワークで繋がっている PC、WS でファイルを共有し、PC、WS 両方で、ユーザにその境を感じさせないようにし、統合化が容易なシステムとする。

2.1 音信号転送

音を送る側をクライアント、受ける側をサーバとする。クライアント側からサーバに接続要求を出し、サーバ側が接続許可を出すことで、クライアント側から音の転送が始まる。サーバ側は、接続許可を与えるときに、クラ

イアント側に「OK」という文字を送ることで接続を許可する。クライアント側はこの2文字が送られてくるのを確認してから音声を送信する。いくつものサンプリング周波数やステレオ、モノラル等の、各種の品質の音を取り扱えるようにした場合、ネットワークの効率によっては、この時点でサーバ、クライアントの両者間で音質を取り決め合うこともできるようにする。

音が途切れないようにするために、録音とデータ送信、あるいは再生とデータ受信を同時におこなうことで、それを回避できる。実際にこれらの処理を行なうためには、マルチスレッド処理をする必要があるが、同時に行なう際には、ファイルの共有が重要となる。再生、録音、送受信を、全く同じファイルで共有すれば、必ず、どこかで不都合が生じる。

Windows 上で音、画像といったマルチメディアサービスを利用する際には、MCI(Media Control Interface)、高レベルオーディオサービス、低レベルオーディオサービスの3つがある^[5]が、前者2つのサービスはファイル自体を指定して扱うため、ファイルの共有が出来ないので、3つ目のサービスを利用する。大まかな流れとしては、メインメモリ上にバッファを設けて、そのバッファに録音再生する音のデータを格納していき、そのバッファをデバイスに送ることで、あとはデバイス上で並列処理を行なう。

音声用のバッファは、サーバ側、クライアント側で、最低二つずつは必要となる。一つは録音、再生用のもの、もう一つは転送用のものである。そして、この二つのバッファをそれぞれの役割で交互に切り替えて、ダブルバッファすることで、ファイルの共有を避ける。

録音した音が、ほぼ同時に再生されないと音の遅延が生じる(図1)。このような音の遅延は、音がとぎれて聞こえるようになる。サンプリング周波数を 11.025kHz、音の量子化ビット数を、8 ビット、モノラルで、1 秒間の音のデータ量は、88.200 ビット程度だが、実際には、同一ネットワーク上をファイルが送受信しているため、ディスク処理に時間がかかり、とぎれる時間が多くなる。そこで、サーバ、クライアントそれぞれのバッファを増やす、1 回にネットワーク上に流すデータ量を減らす方法によりこれを回避する。

ネットワーク上での動作は、音声を受信するサーバが、常にクライアント側からの接続を待機するようにする。クライアントからの接続要求が来たら、ユーザに接続を許可かを問い、許可、不許可をおこなう。許可されると、接続中となり、いつでもクライアントからの音声を受信する。音声を送信するクライアント側は、サーバへ接続要求を行う。サーバに許可が出されると、クライアントのマイクからユーザが音声を入力する。入力された音声データは、ネットワーク内で、共有ファイルとして、他の端末からも利用できる。

3 統合化

動画像分野、音(音声、楽音)分野がそれぞれの担当役割機能によりそれらを動的情報として生成し、統合をしていく方法が図2に示されている。すなわち各個別メデ

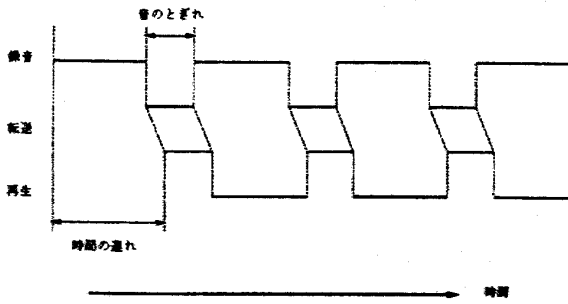


図 1: 音声転送

アについては、時間（再生する時間）に沿ったデータの流が存在する。このデータの流には時相的制約があり、動画像と音の同期を取る必要がある。そのため同期関係を指定する同期用特徴点マーカーをユーザが指定する。例えば、人が「あ」と発音した場合、動画像には、口を開けた点と、音声の発音が始まった点のそれぞれが、同期用特徴点マーカーとなる。それが全メディアにわたって、時間に沿ったデータの流れにおいて隣接し合う時点マーカー間のメディアについて、動画像の再生、音の再生を速くもしくは遅くするといった時間的伸長又は圧縮を行う^[2]。ユーザは複数の場合もありうる。以上のような個別メディアの統合化の方法として、動画像描画において用いられる総称的方法^[4]を、すべてのメディアの生成についても適用することができる。たとえば複数の楽器の合奏において、個々の楽器の演奏データの生成において総称的に行える他に、時相的制約において、楽音やピッチなどの制約として表現できる。総称的方法により、多種多様な動画像を生成することができる。

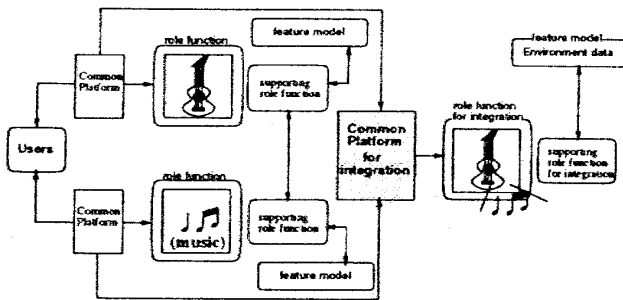


図 2: 音声と動画像の統合

同期再生に必要なデータは、WS上で作成された動画像データとPC上で作成された音声、楽音データの2つの異なる場所にあり、再生を行なう際にネットワークを経由してデータが行き来する。実際には同期情報から動画像の総動画像枚数などにも条件がある。このような制約条件の確認やWSとPCでの通信の際のデータ洩れなどのチェックを行なう必要がある。制約条件においてはユーザがデータを作成していく過程において、ある条件を満たしていなければエラーを返し、データ洩れにおいてもデータ不足のエラーを返すことでユーザに知らせる。例えば音声情報を元に動画像を作成していく際に、1つの音を選択し画像を作成したが、音声の長さが短く同期が取れない場合には、動画像を1枚追加することでユーザを支援し整合を行なうことができる。これらの異なる分野同士の同期再生を行う際に必要な情報としてテンプレートを使用する。MIDIデータの「ド」「レ」「ミ」の情報を、動画像へ反映させ、長さに合わせて動かすことができる。

4 ストーリー性

原理的には、あるユーザが、各個別メディアを生成し、それらを統合して、1つの作品を作り上げることができる。しかし、1人で大きなストーリー性を持つ作品を作るのは難しい。そこで、1つのストーリーに、必要な個別メディアの仕様^[3]、脚本、シナリオを計画設計し、複数のユーザで全体が生成できるように役割を分担する。その場合、重要となるのはシナリオであり、生成手順^[2]としては、一般に

1. 役割のプロット(筋)の作成
2. 役割間の意図関係の指定
3. 各シーンのメディア情報→プロット→の系列の計画設計し、仕様とする→原作、脚本→台本
4. プロットの各役割の登場場面、セリフとアクション、衣装、証明、黒子などの支援役割の協調作業の演出台本
5. 各種台本の統合関連図→作成者用

本論文の事例では、1つのストーリーからの生成ではなく、そのような本格的な作品製作に対し異なるユーザが別々に生成した作品に関し上記の一部を組み合わせることで、1つのストーリーを生成することができることを示している。図3は、異なるユーザが生成した背景の動画像と、鳥の動画像のストーリーに基づく統合について示している。

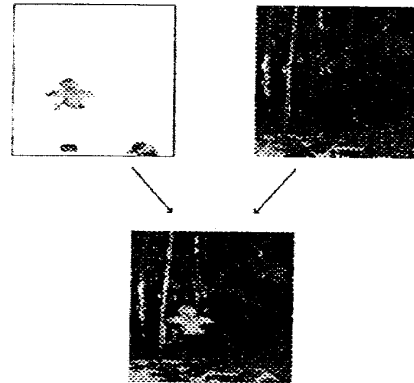


図 3: ストーリー性

5 まとめ

本論文では、複数の独立したメディア情報を、それぞれ個別に生成、さらにそれらの統合化を、分野に依存しない Extensible WELL で、画像分野、音声分野の役割を持たせて行う構造について述べた。数種の事例について結果を示す。事例は、仏像、気球、動植物の動き、楽器演奏など、本研究室の卒研生の労作による。

文献

- [1] T.Miyagi, A.Hori, H.Sugama, Y.Murao, H.Enomoto "Realistic image generation using model driven processing in an interactive system", Proc. IS & T/SPIE's Symposium on Electronic Imaging: Science & Technology, Feb.1997
- [2] H.Enomoto, Y.Murao "Interactive Realization system of visual reality using hierarchical model driven concurrent processing", Proc. IS & T/SPIE's Symposium on Electronic Imaging: Science & Technology, Jan.1998
- [3] 増田 征貴 "Extensible WELLにおける対話プロセスの表現", 情報処理学会第56回全国大会, 3K-04, 1998.3
- [4] 太田 徹 "総称的オブジェクトによる動画像表現-対象の動き表現とそのモデル化-", 情報処理学会第56回全国大会, 4X-04, 1998.3
- [5] 翔泳社 "Windows95 API バイブル 3 ODBC、マルチメディア編" 1997.5