

マルチメディア情報をプログラマブルに操作可能にする構造化技術

～マルチメディア携帯端末の実装と応用～

大盛 善啓、辻本 修一

(株) 東芝 研究開発センター 情報・通信システム研究所

〒210-8582 川崎市幸区小向東芝町 1

概要: 本論では、マルチメディアデータに対して、データ構造を利用したアクセス手段を可能にする構造化技術と、その技術を実装した拡張マルチメディアレコーダ (EMMR) について述べる。EMMR は、入力された手書きペン・テキスト・音・静止画等のマルチメディア情報を、意味のある単位に区切るにより構造化して、各単位で関連付けて記録する。記録されたデータが入力された時の状況を、マルチメディア情報によりリアルに再現するための検索 IF も備える。更に、EMMR で管理するマルチメディアデータはオブジェクト指向によって記述されるので、マルチメディアデータを連携して利用するアプリケーションの支援を可能とする。

Enhanced MultiMedia Recorder (EMMR): a tool for programmable access to organized multimedia data

Yoshihiro Ohmori, Shuichi Tsujimoto

R&D Center, Toshiba Corporation

1, Komukai Toshiba-cho, Saiwai-ku, Kawasaki 210-8582, JAPAN

+81-44-549-2243

{ohmori, tsuji}@eel.rdc.toshiba.co.jp

Abstract: In this paper, we describe multimedia structuring technique which provides access method for multimedia data, and implementation of this technique to the "Enhanced MultiMedia Recorder" (EMMR). The EMMR records heterogeneous multimedia data, such as pen annotations, texts, sounds, and pictures, and organizes them by dividing them into meaningful units to relate these units. The EMMR has a searching interface for replaying the environments when the data were recorded. Furthermore, the EMMR has simple representation of recorded multimedia data, and supports the applications that use multimedia data.

1. はじめに

近年、デジタルカメラ、デジタルビデオ、タブレットなどのデジタル入力機器が普及している。これらの機器で入力したデジタルデータは、アナログデータとは異なり、ハードウェアの制約などを受けることが少ないため、異種メディアのデータを、組み合わせて利用することが容易になった。

従来、この特徴を利用した、新しいマルチメディアデータの使い方の提案がいくつか見受けられる。その一例として、手書きの注釈やサウンド等の複数のメディアのデータを同時に記録し、記録後に、手書きの注釈メディアをインデックスにして、他のメディアであるサウンドなどに、アクセスする研究が挙げられる[1] [2] [3]。しかし、これらの研究では、ユーザが GUI を介して、通常は検索困難な時系列データへのアクセスを容易にすることに焦点が置かれていた。本論ではこれに加え、データベースなどのアプリケーションが、マルチメディアデータの関連付けを利用して時系列データにアクセスすることを目的に、構造化されたマルチメディアデータにアクセスする枠組を提案し、拡張マルチメディアレコーダ (Enhanced MultiMedia Recorder (EMMR)) に実装した。

本論では先ず、マルチメディアデータの新しい利用に関する研究動向について述べ、従来の問題点を示す。次に、この問題を解決するマルチメディアデータ構造化と、オブジェクト指向による実装について述べる。最後に EMMR を使った、マルチメディアデータ検索の例題を示し、その結果について述べる。

2. マルチメディアデータを関連付け利用する研究の動向

マルチメディアデータの新しい使い方の提案として、我々は、複数のメディアのデータを簡単に記録 / 検索する「マルチメディアレコーダ (MMR)」を開発済みである(図 1 参照)[1]。

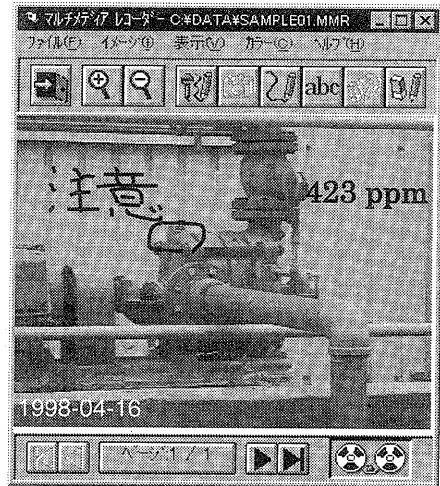


図 1 マルチメディアレコーダの外観

MMR は、静止画、手書きの注釈、テキスト、サウンドの各データを関連付けながら、同時に記録し、記録後に手書きの注釈やテキストをインデックスにして、同時に記録されたサウンドデータを検索できる。この様子を図 2 に示す。

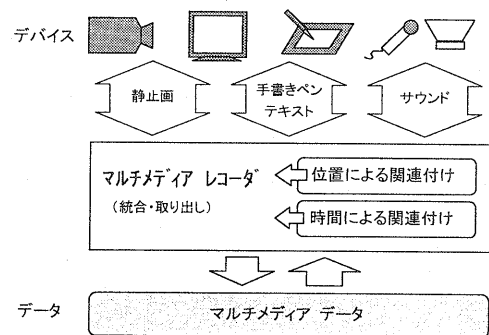


図 2 異種メディアデータの関連付け

MMR は、デジタルカメラとサウンド機能を持つ携帯型ペンコンピュータ上に実装され、携帯型パトロール端末「MULTI-IN®」のマルチメディアデータ入出力部として商品化済みである[4]。MULTI-IN®により、プラント巡視などのモバイル環境で、複数のマルチメディア情報を効率良く収集することが可能になる(図 3 参照)。

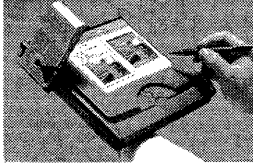


図 3 MULTI-IN®

プラント巡視員は、デジタルカメラで、検査対象の配管の写真を撮影し（図 1）、その画像の上に例えば丸印や「注意」の手書きの注釈を入力して写真の機器の故障箇所を示したり、手書きメモやキーボード入力により注釈を付加できる。注釈の記録中には、サウンドも同時に録音されており、注釈入力と、それと同時に録音されたサウンドデータとが注釈に関連付けて記録される。

データの記録後、ユーザが、例えば、丸印の手書きの注釈をペンでクリックすると、MMR はこの注釈が入力されたときに録音された音声、例えば「水がもれています」を再生する。これにより、プラント巡視員は、注釈を見ることによって配管の状況を概観できるだけでなく、その注釈と関連付けられた音を聞くことによって、注釈の意味などの補助情報を知ることができる。

MMR の他にも、マルチメディアデータの新しい使い方を提案する研究がいくつか行われている。例えば、Wilcox らは手書きの注釈とオーディオを構造化して記録して、検索を容易にする電子ノートを開発している[3]。Weber らも同様に、ユーザによって入力されたテキストのキーワードと、それと関連付けられたビデオシーンを同期させて再生するビデオデータ検索システムを開発した[2]。

3. 従来のシステムの問題点

これらのシステムは、手書きなどの注釈と、その注釈が入力された時に同時に録音されたサウンドなどの時系列データとを関連付けることに特徴があり、主に時系列データの頭出しに主眼が置かれていた。このように比較的単純なデータ利用を

する場合には、データ構造は、注釈と時系列データを関連付ける程度で十分であった。

しかし、「手書きの注釈と関連付けられたサウンドデータを削除する」などの、データ操作を行うには、上記の方法では不十分であり、削除すべきサウンドデータの範囲を GUI を介してユーザが指定するなどの煩雑な操作が必要となる。

このように従来のシステムでは、データベースなどから、ファイル単位で入力データを扱うことはできても、その中のデータ単位の操作を行うことはできなかった。本論ではこの問題を解決するために、時系列データを操作に適した単位に区切る構造化方法を開発し、オブジェクト構造記述により、データ構造の表現を実現した。

4. マルチメディアデータ構造化

拡張マルチメディアレコーダ (EMMR) は、従来のシステムが持っていた問題を解決するために、MMR を拡張して、マルチメディアデータ構造化を行う。EMMR ではデータ利用の目的に応じて、データの単位を自由に変更して関連付けられるようにマルチメディアデータを構造化する。

図 4は、EMMR によって、構造化されたマルチメディアデータの一例を示す。EMMR では、データを「セグメント」と呼ぶ単位に区切って管理する。セグメントには、2 種類あり、独立セグメントと再構成セグメントがある。

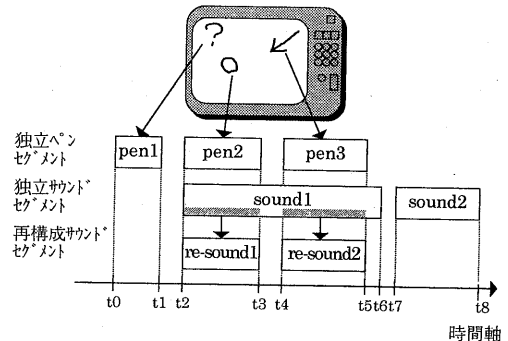


図 4 セグメント生成の例

4.1. 独立セグメント

独立セグメントは、時系列データを意味のある単位に区切ることにより生成され、データの操作に必要な単位を提供する。独立セグメントで扱う時系列データは、例えばペンデータだけ、というように、単独のメディアである。これにより、時系列データは連続した時系列データではなく、セグメントというまとまった単位として扱われる。

「意味のある単位」とは、操作が簡単になるように、データに固有の性質を利用して、データを時間的、空間的にまとめたものである。例えば、タブレットが生成する座標値列として入力されるペンデータは、一定時間内に連続して入力されたストロークをまとめることにより、ペン独立セグメントを生成し、データの操作性を向上させる。例えば、図 4 のコンピュータ画面上に描かれた矢印の手書きの注釈は、t4 から t5 の間に矢尻のストロークと柄のストロークが描かれたものである。EMMR は、これら 2 つのストロークを“pen3”独立ペンセグメントとしてまとめることにより「矢印」として、構造化している。

サウンドデータは、音圧値の連続データとして入力され、サウンドの入力開始、終了のユーザによる指示により区切られる。例えば、図 4 の t2 から t6 の間に録音されたサウンドは、“sound1”独立セグメントとしてまとめられ、構造化される。

図 4 には示していないが、EMMR は、テキストデータに関しても同様にして連続して入力されたテキストを独立テキストセグメントとして生成する。

4.2. 再構成セグメント

再構成セグメントは、データの操作性をさらに向上させるために、異種メディア間の関連を利用して生成される。EMMR では、関連付けの方法として「手書きの注釈と同時に録音されたサウンドデータは、その手書きの注釈と密な関係がある」というヒューリスティクスに基づき、あるメディアのセグメントと同時に記録されたデータを互い

に関連があるとしている。

例えば、図 4 の“re-sound1”再構成サウンドセグメントは、“pen2”独立ペンセグメントの単位を用いて、サウンドデータを再構成したものである。具体的には、“pen2”独立ペンセグメントが記録された t2 から t3 の間と同時に録音された、サウンドデータの一部が“re-sound1”ペン再構成セグメントとして生成され、“pen2”独立ペンセグメントと関連付けられる。

逆にサウンドセグメントの単位を用いて、ペンデータを再構成することもできる。図 4 には示していないが、例えば“sound1”独立サウンドセグメントの単位を用いて、ペンデータを再構成すると、“sound1”サウンドセグメントと、同時に入力されたペンデータ (t2 から t6 の間に記録された“pen2”と“pen3”の集合) が再構成ペンセグメントとして生成され、“sound1”独立サウンドセグメントと関連付けられる。

5. オブジェクト指向によるセグメント表現

セグメントは、データの操作を容易にする単位と単位間の関係とを提供する。しかし、データベースや他のアプリケーションがセグメントを利用するには、セグメントを表現可能する必要がある。EMMR では、オブジェクト指向の枠組を用いて、セグメントとセグメント間の関連を実装した。

図 5 は、図 4 に示したセグメントと、セグメント間の関連を、オブジェクト木で表現したものである。図中の四角形はすべてオブジェクトを表し、オブジェクトにぶら下がるように線でリンクされたオブジェクトは子オブジェクトを表す。独立/再構成セグメントは、オブジェクトとして実装される。EMMR 自身もオブジェクトとして実装され、すべてのデータオブジェクトは、EMMR オブジェクトからリンクを辿って行けるように記述される。

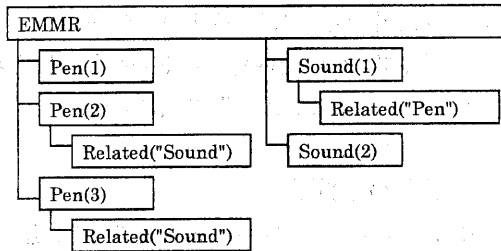


図 5 EMMR のデータ表現

独立セグメントは、図 5に示すように EMMR オブジェクトの子オブジェクトとして実装される。例えば、図 4に示した、“pen2”独立ペンセグメントは、オブジェクト指向の表現を用いて、EMMR.Pen(2)と記述される。このように独立セグメントをオブジェクトとして実装することにより、独立セグメントが表現可能になる。

再構成セグメントは、図 5に示すように独立セグメントの子オブジェクトとして実装される。これにより、「親オブジェクトと関連付けられたデータ」というセグメントの関係が表現可能になる。例えば、図 4に示した、“related1”再構成サウンドセグメントは、EMMR.Pen(2).Related(“Sound”)と記述される。これは、EMMR.Pen(2)と同時に録音された、“Sound”メディアのデータ、という意味を表す。

同様に、EMMR.Sound(1).Related(“Pen”)再構成ペンオブジェクトは、EMMR.Sound(1)独立サウンドデータセグメントと、関連付けられたペンデータ、すなわち、“pen2”と“pen3”独立ペンセグメント

の集合を表す。このように、再構成セグメントを、独立セグメントを表すオブジェクトにリンクされた子オブジェクトとして実装することにより、データ構造も表現可能になる。

オブジェクトには、データの操作を可能にするために、メソッドやプロパティが実装されている。例えば、“pen2”セグメントの消去は、EMMR.Pen(2).Delete メソッドの発行により行われる。その他にも、データの内容を参照する Data プロパティや、データの内容を表示したり再生する Activate メソッドなどが用意されている。

6. 例題

前章で述べたデータ表現により EMMR で入力されたマルチメディアデータはプログラマブルに利用することができ、マルチメディアデータを連携して使用するアプリケーションの構築を支援することが可能になる。

この特徴を示すために、我々は図 6に示す簡単な実験を行った。この実験では、検索アプリケーションが、テキストとキーワードを用いて、マルチメディアデータを検索し、これと関連付けられたサウンドデータを削除する。図 6の例では、巡視の際に入力した“23”というテキストデータを検索キーにして、これらのデータを持つマルチメディアデータを検索する様子を示している。

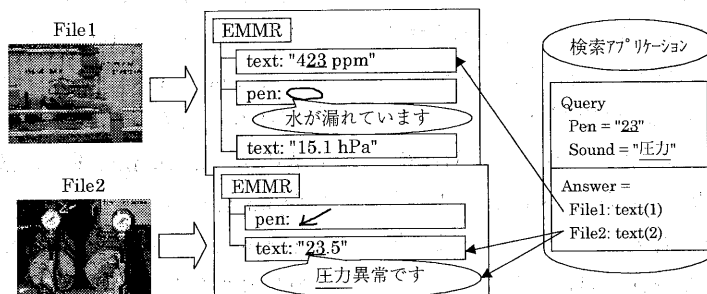


図 6 EMMR を使ったマルチメディアデータ検索の実験

左の File1 と File2 は検索の対象となるデータファイルを表す。中央にはそれらのデータ構造が図 3 と同様の方法で描かれている。右は検索アプリケーションを表す。

従来のシステムでは、データはファイル単位でしか操作できなかったが、EMMR を介することにより、検索アプリケーションはデータの構造も利用することができるようになる。

検索の手順を以下に説明する。検索アプリケーションは、まず、テキストオブジェクトを参照し、“23”の文字列を含むオブジェクトを検索し、検索結果のオブジェクトのファイル名とオブジェクト番号を“File1: text(1)”, “File2: text(2)”のように、検索アプリケーションのディスプレイに表示する(図6参照)。この検索は、各テキストオブジェクトについて、EMMR.Text(n).Data プロパティを参照することにより実現される(n は任意のオブジェクト番号)。EMMR が独立セグメントを提供する機能を有するため、アプリケーションが個々のデータの内容を参照することが可能になる。

次に、ユーザはこれらの検索されたデータの中から、“圧力”の音声を含むデータを検索する。この検索は、ユーザが、検索されたすべてのテキストオブジェクトに対して、それと関連付けられたサウンドデータを再生して、ユーザがひとつずつ確認することにより行う。この確認の操作は、EMMR.Text(2).Related(“Sound”).Activate メソッドを、“File2”に対して発行して「圧力異常です」とサウンドを再生し、ユーザがそれを聞いて確認することにより実現される。これにより、ユーザは“File2: EMMR.Text(2)”を目的のデータとして検索することができる。EMMR が再構成セグメントを提供する機能を有するため、関連付けというデータ構造を利用したデータの参照が可能になる。

このように検索されたサウンドデータに対して、EMMR.Text(2).Related(“Sound”).Delete メソッドを発行することにより、検索アプリケーションは目的のデータを削除することが可能になる。

このように、EMMR の提供する機能により、テキストとサウンドで構成されるマルチメディアデータを検索キーにしてマルチメディアデータを検索し

て、それを削除するなどの利用が、データベースなどから、プロパティ参照、メソッド発行などのプログラマブルな方法により実現できることを示した。

7. まとめと課題

拡張マルチメディアレコーダが提供する、構造化されたマルチメディアデータへのアクセス手段は、データ単位とデータの構造を操作するアプリケーションで有効であることを、実験により示した。

本論で示した例題では、サウンドデータの検索については、ユーザの補助を介して実現したが、将来的には、音声認識を用いて検索を完全に自動化することを計画している。また、ペンデータについても、手書き図形認識を用いて、自動検索することも計画している。

謝辞

本論の執筆にあたって、多大なご支援を頂くとともに、有用な洞察を頂いた Dr. Gareth Jones に深く感謝致します。

参考文献

1. Imai T., Yamaguchi K., and Muranaga T., Hypermedia Conversation Recording to Preserve Informal Artifacts in Realtime Collaboration. in Proceedings of ACM Mutimedia'94, 417-424
2. Weber K. and Poon A., Marquee: a tool for real-time video logging, in Proceedings CHI '94, 58-64.
3. Wilcox L., Schilit B., and Sawhney N., Dynamite: A dynamically organized ink and audio notebook, in Proceedings CHI '97, 186-193.
4. マルチメディア情報統合管理ソフトウェア「MULTI-IN」, http://www.toshiba.co.jp/product/ipcsg/denkei/multi_in/index_j.htm