

汎用データフォーマットによるマルチメディアデータの表現*

4 V-9

平中幸雄 嘉藤雅文†
山形大学工学部

1. はじめに

(1)音声や画像などは、人間から見たときに同じ情報なのに、ソフトやバージョンによってデータ互換できないことが多い。本質的に柔軟なデータ表現はできないものだろうか。(2)データ・音声・画像などからなるマルチメディアデータとして一体のものとして扱いたい。(3)線をつなげば音が出るオーディオ機器のように、複雑なプロトコル不要で通信できないか。これらの問題を、自然で柔軟性のあるデータフォーマットを作成し、これを計算機世界での基本語にすることで、解決することを考えている。具体的なフォーマットを、試行・実験を行いながら、検討中である [1]。これまでの検討結果、現在のフォーマット、マルチメディア一体表現、多チャンネル伝送の基本方式について述べる。

2. 汎用データフォーマットの条件

ソフトウェアは柔軟に何でも対応できると期待されているが、逆にハードウェアより制約が大きい。この状況を脱するには、限定された表現形式を使うべきである。その条件を考えてみると、

(1)現状で共通に理解できる要素、表現を基礎にすること、(2)できる限り柔軟であり、拡張性があること、(3)仕様の変更は、変更前の仕様の拡張形で実現すること、(4)音声・画像・データなど表現形態を統合できる、つまりマルチメディアデータとして扱えること、(5)デジタル通信の利点である、異種データの多重化能力を持つこと、(6)一部の表現が解釈できなくとも、解釈できる部分だけを利用するようデータ解釈(受け取り)側の自由度を大きくすること（これは伝送誤りなどの影響を最小限にすることにつながる）。

これらの条件にあうデータ表現法の基本として以下を考えている。

- A) すべて文字で表現する。二進データは解釈が個別的となり、将来の拡張法を先に規定しないとイケないので、常に制約を作ることになる。文字表現であれば、人間世界で行われているのと同じ拡張を最初から前提にすることができる。たとえば、数値や名前の桁数は無制限である。文字コードとしては ASCII 文字表現を基本とする。漢字文字などは JIS コードのようなモード変更方式による拡張が妥当に思える。
- B) 正規化 SGML 方式で、すべてのデータを `<tag>data...</tag>` の形で表現する。データの表現は、相対的にデータを規定することと考えられ、理解される相対指定要素 (tag) のもとでデータを与えるのが妥当であろう。現実的メリットとして、データに対して XML 処理系が利用できる [2]。
- C) 音声データのサンプリング間隔など、解釈側として必要とする情報がある場合は、パラメータ表現するが、その値が特に指定されないときの標準値を決めておき、細かなことにとらわれない利用ができるようにする。
- D) データの細かな指定要素は付加データとし、解釈できないときは捨てても大きな問題はないようにする。
- E) 同じ tag のデータは、外側 tag が違っていなければ、連続しているものと見なす。

`<A>1,2,3</A><B>a,b,c</B><A>4,5,6</A>` は、`<A>1,2,3,4,5,6</A>` と `<B>a,b,c</B>` が同時並置されているのと同じとする。A を音声、B を画像とすると、音声・画像の割り込み混合データ表現ができ、同期もデータ列順序として表現できる。

- F) 異なるデータは最も外側の tag まで一旦戻って切り替える。

`<ch1><audio>1,2,3</audio></ch1><ch2><audio>4,5,3</audio></ch2>`

では、二つの audio は別々の音声として処理される。この方式により、異種データの混合伝送ができ、ch1、ch2 をそれぞれテレビチャンネルと考えると、すべてのテレビチャンネルデータをひとつの通信路で伝送でき

* Multimedia Data Representation by Using Universal Data Format

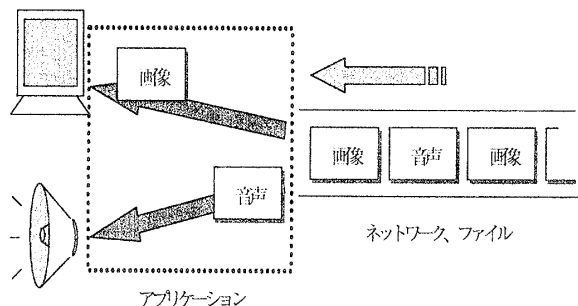
Yukio HIRANAKA and Masafumi KATO, Yamagata University

4-3-16 Jonan, Yonezawa, 992-8510 Japan, corresponding e-mail: zioi@ieee.org

† currently, NJK Co.Ltd.

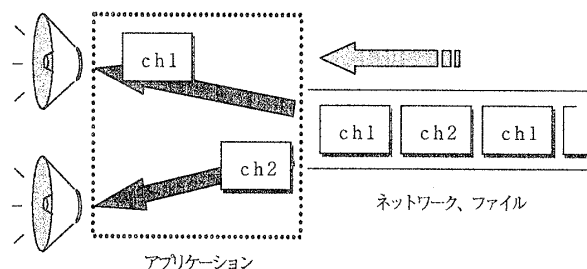
る方式ができたことになる。

- G) 伝送時間短縮のためのデータ圧縮はこのデータ表現法の外側で考える。実際には、V42.bis などの文字圧縮技術を利用するのが適当である。
- H) 伝送エラーなどによる解釈の問題は、アプリケーションもしくは通信路の選択の問題とするが、大量データでは行番号やタイムスタンプなどの挿入により、回復不能な部分は小さくなるようにする。



```
<picture><line>0,0,2,3,...</line><line>...</line>...</picture>
<audio>10,12,...</audio><picture><line>4,3,2,3,...</line>
<line>...</line>...</picture>
```

図1 マルチメディアデータの表現



```
<ch1><audio>1,2,3</audio></ch1>
<ch2><audio>4,5,3</audio></ch2>
```

図2 多チャンネルデータの混合

3. 具体的表現

テキスト:	<text> text data </text>
テキスト (文中に<tag>を含む場合は、<verb> </verb>で囲む):	<text> ... <verb><tag></verb> ... </text>
音声 (サンプリング間隔は 8kHz を標準とする):	<audio>10,12,...</audio>
ステレオ音声(10d2 は合計 10 の強度で左右差が 2、L=6,R=4 を意味する):	<audio> 10d2,12d3,...</audio>
音声 (サンプリング間隔指定):	<audio><sampling_rate>44.1kHz</sampling_rate>10,12,...</audio>
ビットマップ画像:	<picture><line>0,0,2,3,...</line><line>...</line>...</picture>
ビットマップ画像 (行番号の指定):	<picture><line><#>1</#>0,0,2,3,...</line>...</picture>
カラービットマップ画像 (HLS(明度色相彩度)カラーモデルなどで表現する [3]):	<picture><line>0,0,2h30s0.3,3h35s0.4,...</line>...</picture>
タイムスタンプ (時間を規定しようとする場所に挿入する):	<audio><time>1999y3m31d10h20m11.9s</time>10,12,...</audio> <audio><time>+2s</time>10,12,...</audio> その前の基準時間 (default はスタート時間) に対する相対指定
図形プリミティブ:	<picture>circle(r10,x30,y20,ts=5,te=10),line(...)</picture>
データ (データごと tag を決める。知らない tag はすべてデータ tag とみなす):	

音声、動画像を単一通信チャンネルを通して伝送する図1の形態を、Windows95 PC 上で、API を作成して実現した。通信路でのデータ圧縮を行っていないため、ロスがあるが、通信部分を簡単に実現できた。フォーマットの細部はまだ検討が必要なものが多く、基本方針を見失いやすいが、実際の応用を少しずつ試みながら、目的の実現を図っていきたい。

この研究の一部は通信・放送機構平成十年度新規事業創出型研究開発制度による受託研究として行った。

参考文献:

- [1] 嘉藤雅文・平中幸雄, 汎用マルチメディアデータフォーマットの研究, 情報処理学会東北支部 1998 年度第4回研究会, 資料番号 98-4-24, 1999.3.
- [2] 星武史・平中幸雄, XML による柔軟なデータベース表現方式, 情報処理学会東北支部 1998 年度第4回研究会, 資料番号 98-4-25, 1999.3.
- [3] 日本図学会編, CGハンドブック, p.426, 森北出版, 1989.