

デジタル エッセイ

「動画フォーマットと私」

坂口 裕靖

映画では、フィルムという物理媒体に映像と音声空間的に分離して保存されており、かつ長さ方向が時間にそのままマップされ（まあ、映像の方は間欠駆動により量子化されていますが）、同期が外れないようにシステムアップされていました。その後色々な付加機能が提案されますが、いずれもフィルムの空き部分、例えばパーフォレーションのスキマに二次元バーコードで圧縮した音声を書き込むとか、パーフォレーションの外側、よく破れちゃう部分に二次元コードで圧縮した音声を書き込むとか、映像領域とアナログ音声領域のスキマに同期信号を埋め込んで、外部音源と同期を取るなど、いずれの方法もフィルムの空間領域に信号を埋め込むことで、同期の問題を回避するというやり方が踏襲されています。ことほどさように、同期という問題はシビアで根

源的でした。

この流れはビデオシステムに持ち込まれ、映像トラックと音声トラックが空間的に分割されて、テープ上に記録されるようになります。筆者が会社に入った頃、まだ1インチのVあTRが主流で、2インチのデッキが何か残りつつ、D1が導入され始めたというあたりでした。これらはいずれもテープの長手方向に時間軸を取り、空間的な位置合わせで時間的な位置合わせを行っていると考えてよいでしょう。アナログの場合は読み出したそのまま信号となっていていきます（TBCが壊れた状態のアナログビデオの信号、今思うと大変面白いものがありました...）が、デジタルの場合は数学的に補正してから出すことができるため、データを位置的に分散保存するなどの工夫が行われています。このため、同じ

データが細かに前後することになり、長手方向に一樣な時間軸がとれるわけではありません。いずれにしろ、こういった物理デッキの場合は入力する信号と出力する信号が決まっていたのが当たり前でした。さて、映像や音声をトラックとして抽象化し、ファイルの頭からのバイト数を時間（の関数）と見立ててデータを配置していくという形で動画ファイルフォーマットが出てくるのは自然でしょう。初期のquicktimeやaviは、そもそもOS自体がTCP/IPのプロトコルスタックを持っていなかった時期に提案されたこともあり、ネットとかマッシュプリーパラレルといった状況は一切想定されていない、大変シンプルなものでした。当初は非圧縮というか無圧縮というか無配慮というか、である画像・音声ファイルが、適当なカタマリで適宜配置されるようなフ

ワンポイント バスワード statcast

グランドプリンスホテル新高輪国際館パミール。品川のパミールは会社の近くにあるので、あのあたりでイベントが行われる際には結構お世話になっております。先日はAWS summit Tokyo 2016（エイダブリエスサミットウキョウ、トゥーサウザンシクステーション... 本当にこういうアナウンスが流れるんですよ）で行ってきました。今回は事務所から歩いて行って、途中梅林で肉ソース焼きそばの大盛りを頼み、シャツにソースのシミを付けつつ、CSLの前を通り過ぎ、坂をえっちらおっちら登って、幾つもの高輪教会を通りすぎて東京高輪病院の入り口に到着。そこから歩いて入ろうと思ったんだけど、大変迷いました。素直に一旦新高輪のロビーに入って、中を移動したほうが早かった。あれ、クルマで移動すべきところであって、歩くところじゃないですな。今回のAWS以下略は、どちらかという企業の情シス向けの色合いが強く、「クラウドでも大丈夫ですよ」的なセッション

が多くて、実はちょっとがっかりでした。それでもキーノートはいくつか面白いプロジェクトが紹介され、その中でも筆者はstatcastが気に入りました。まあ不勉強で知らなかったわけですけど、球場にカメラとかレーダーとかを仕掛けて、画像認識および電波でプレイヤーとかボールとかの動き・速度・加速度をキャプチャして、それをリアルタイムで伝送可能にする、というものでした。いいですね、21世紀ですね。MLBAMのサイトを見ると、2Dカメラの映像にキャプチャしたボールの軌跡をマッシュムーブしつつオーバーラップして表示したりしていますが、今後いろんなスポーツでこういう方向の動きは進んでいくのではないのでしょうか。逆バーチャルスタジオというかARスタジオというか。こういったヘッドアップディスプレイ的な情報提示というものは、今後重要なものになっていくのでしょうか。なんかこう、全スポーツF1化という感じ。

フォーマットでした。初期の「切手サイズ」の時代はまあそれでも物珍しさでなんとかなりましたが、ちょっと画角を大きくしようとする、表示サイズの二乗でファイル容量が増えるため、データ圧縮の必要性が大きくクローズアップされます。

結局、動画ファイルフォーマット単体ではいくつかの標準的な「codec」により圧縮・解凍できるようにしておいて、この codec を別途インストールする、という形で安定したように思います。codec によって圧縮されたデータは、デコードアルゴリズムに従ってきちんと伸張しない限り、どう見ても無意味なピクセル列のようにしか見えません。codec を可換にすることにより、多少の制約はあるものの、その枠内で任意のアルゴリズムを使えるようになったというわけです。quicktime が Apple のスカリーにより発表されたのが 1991 年、マイクロソフトが後追いで avi を発表したのが翌年ですが、それから四半世紀使ってきたわけですね。先日 quicktime は windows 版のサポートを停止したため、命運つきました。

一方で、動画ファイルフォーマットとして youtube を捉えてみましょう。もちろん youtube は「ファイル」を提供してませんので、その方面の厳密な比較は不可能なわけです。とりあえず「フォーマット」としての youtube を捉えてみると、「ブラウザで表示でき、必要なら既存のフォーマットにトランスコードできる」ということ以外、なんら制約がないことに気が付きます。youtube の内部フォーマットがどうなってるか、誰も気にしません。一方で youtube の入力フォーマット、出力フォーマットは様々で、制約にはなっていません。特にブラウザのインターフェースで考えると、ブラウザが動員できるすべての CPU パワーを援用することは不可能ではありません。その結果、「静止画と音声のカタマリ」のみから構成されるような「動画」に制限される必要がありません。

現状、動画を表示するデバイスの大部分は、結局データから「静止画と音声のカタマリ」に変換するものとなっています。こ

れはたまたまテレビが「静止画と音声のカタマリ」以外伝送できないからこうなってる、という半ば歴史的な理由によるものが大きく、なんら本質的な必要性から出てきたものではありません。例えば今後、スキンとボーン、モーションデータで構成されたアバターが 3D 空間を躍動するような伝送系が出てきた場合を考えてみましょう。3D 空間を「静止画と音声のカタマリ」としてまるまるキャプチャすることは大変難しいですが、これを予め用意したスキンとボーン、そしてリアルタイムにその場でキャプチャしたモーションデータと、背景となる舞台を加えたデータ列に置き換えると、これってそもそも「三次元テレビ」が目指したものでないでしょうか。

「そんなのモーションキャプチャの人形劇じゃん」というご指摘はごもっともですが、今の 4K カメラがとらえた演者の表情も、薄っぺらな動画テクスチャが一枚あるだけでしかないですよ？ 4K カメラがとらえた動画テクスチャを「実物を撮影したために、映像」と呼ぶのであれば、実物のモーションをキャプチャして、実物のテクスチャをキャプチャして、実物から測定したボーンから再現されたキャラクタが割りだす映像空間は、カメラの種類が違うだけで、やっぱり「映像」になるんじゃないでしょうか？ もちろん、キャプチャするデータも表情や手足から服のシワまで、ものすごく細かくスキャンしていく必要があるでしょうし、それがそこのプロシューマ機材で簡単に実現できるとも思えません。が、そうやって再現された空間を任意の視点から、任意のタイミングでリプレイできるとしたらどうでしょう。いやー、個人的には大変見てみたいですね。

さて、このようなデータチャンクを保持しようと思うと、従来の意味

での「動画ファイルフォーマット」は全く意味を持たなくなってしまいます。そもそも、撮影した素材のどこにも「静止画のカタマリ」が無いのですから。音声のカタマリはありそうですが、なくてもどうにかなるかもしれません。

codec の考え方は、処理パイプラインを固定しながら、アルゴリズムに自由度を持たせる手段でした。ここでアルゴリズムだけでなく、入出力に自由度をもたせ、さらに計算量にも自由度をもたせると、youtube 的な何かに漸近していくことでしょう。ネット前提なら単一のデータチャンクである必要はなくなり、マッシュリーパラルなストレージに部分部分が保存され、適宜読み出されるような何かであっても、全く問題になりません。

となると、今後「映像」を規定するものは何なのか、考える必要がありそうです。VR やスマホも含めて、ディスプレイデバイス一般で定義できる部分であるなら、結局はピクセルと音声の帯域幅という事にならざるを得ないかもしれません。動画テクスチャしかキャプチャできないデバイスが今後廃れるのか、それとも魔改造により別の展開を迎えるのか？ 我々は今その岐路を目の当たりにしているのかもしれない。

Hiroyasu Sakaguchi
(株) IMAGICA イメージワークス

SWE DISH

CCTスーツケース 90cmφ型 2タイプ有り
120cmφ型

衛星通信用超小型可搬アンテナ
Suitcase CCT Satellite Communications Terminal

緊急報道
ハイビジョン映像伝送
Ku-band/X-band

5分で運用開始

IATA対応収納ケース
その他にも1ケース収納型から3ケース分割型など各種ケースあり

エーティコミュニケーションズ株式会社
<http://www.bizsat.jp> TEL : 03-5772-9125

AT Communications k.k.

