

研究計画書

映像・音声の入出力・操作統一化手法の研究

慶應義塾大学環境情報学部

自署: _____

学籍番号 79950982

平成 14 年 5 月 13 日

概要

インターネットを利用した映像・音声の転送が容易となった。ユーザーやアプリケーションプログラマが映像・音声の転送を行う場合、様々なフォーマットが存在するため、映像・音声フォーマットを意識する必要がある。既存の PC における映像・音声の入出力・操作を統一化された方法で行うための機構は PC 単体での利用が前提であり、ネットワークを利用した実時間映像・音声転送は考慮されていない。そのため、フォーマット毎に転送部分において同様な開発が行われている場合が多い。本研究では映像・音声フォーマット操作方法の統一化及び、映像・音声のための入出力方法の統一化を行い、映像・音声を開発者にとって扱いやすくする環境を実現する。本研究の実現により、映像・音声転送プログラムの開発が容易になり、インターネット上で映像・音声を使ったコミュニケーションがより活発化する。

1 背景：映像・音声転送の普及

ADSL やイーサネットなど比較的広帯域データリンクの普及に伴い、エンドユーザが映像・音声をネットワークを介し利用することが容易となった。

ネットワークを利用してデータ（本文章では、情報を構成するものをデータと呼び、知覚の有無に関わらず人が受け入れるものを情報と呼ぶ）の転送を行ない、実時間性に基づいて再生を行なう事、即ちストリーミング再生における映像・音声等の連続メディアは時間経過に伴いデータの内容が変化する特徴を持つ。情報は時間軸に従い連続性を確保して受信側に到着する必要がある。従ってそれらのデータを送信する場合、欠損したデータの再送などは不必要な場合が多い。また映像・音声はデータ量が大きいという特徴をもつ。さらに、人間の視聴覚特性から情報を省略することが出来る。

これまで、用途毎に様々なフォーマットが開発されて来た。例えば、データ量が大きい情報を低帯域なネットワークで利用するために、データ量を小さくする圧縮フォーマットが開発された。特に映像・音声の情報は視聴覚特性を利用し、情報を省略してデータ量を少なくする非可逆圧縮が使われる。MPEG(Moving Picture Experts Group)フォーマットはその一例である。

情報伝達の信頼性向上のために耐エラー性を持たせたフォーマットもある。圧縮データは圧縮率が高いほど、データ欠損時の影響が大きい。そこで MPEG4 では耐エラー性を高めるための符号化が行われている。

このように、フォーマットは用途毎に開発されるが、情報を扱う方法は統一化されていない。

2 問題意識

映像・音声転送プログラムを開発する場合、生産性を低下させる 2 つの問題がある。入出力する対象毎に入出

力方法が異なることと、フォーマット毎に操作方法が異なることである。

2.1 統一化された入出力方法の欠如

映像・音声の転送を行う際、アプリケーションプログラマはファイル、デバイス、ネットワークなどに対するデータの入出力を統一的な手法で行えない。以下、本文ではファイル、デバイス、ネットワークなどのデータ入出力先をデータソースと呼ぶ。ファイルの入出力を行う場合と、ネットワーク上で実時間転送を行なう場合とでは、実際の作業は異なる。しかしこの行為の違いは、データソースの違いにすぎない。統一的にデータソースを扱える機構が存在しないため、転送部分においてアプリケーション毎に個別の実装が行われ、生産性が低下する。

この非効率な状況を回避するために、データソースに対しての映像・音声の入出力を、統一された方法で扱える機構が必要である。

2.2 統一化された操作方法の欠如

ネットワークや PC 上で映像・音声の転送、再生を行なう際、用途や状態に応じて最適なフォーマットがあり、そのフォーマットを意識する必要がある。

例えば、ある A というフォーマットは A のデータを圧縮・伸張するプログラム、即ち CODEC(COmpression DECompression) を持つプレーヤでしか再生出来ない。

映像・音声視聴用アプリケーションは AV 機器用の再生などの基本機能に対し、共通の UI(User Interface) を持つことが多い。エンドユーザがフォーマット毎にアプリケーションを変更しても操作出来る。

一方、フォーマット毎に機能呼び出すためのインタフェース、即ち API(Application Program Interface) が異なる場合、アプリケーションプログラマが開発を個別

に行う必要があり、生産性が低下する。

フォーマットに関わらず、アプリケーションプログラマが映像・音声に対して求める操作機能は共通であるため、統一された方法で操作できる機構が必要である。

また、映像・音声転送プログラムを開発する場合、状況毎に最適なフォーマットを選ぶ必要があり、アプリケーションプログラマには、ネットワーク、映像・音声フォーマットの知識に加えてその2項目間の知識が求められる。映像・音声転送プログラムの開発を簡易にするには、この要求事項の軽減が必要である。

3 関連技術

本研究はアプリケーションプログラマが映像・音声をより簡易に扱える環境を目指す。前章に挙げた問題意識に基づき、本章において以下の関連技術を3項目取り上げる。

- 連続メディアの実時間通信の為のプロトコルの例
- データの統一化された入出力・操作方法の例
- 統一化されたインタフェース上の CODEC と入出力の部品化の例

3.1 転送機構：実時間通信

TCP や UDP と言ったプロトコルは通信するために必要最低限な事項が規定されている。ネットワーク上で実時間通信を行う場合、従来の TCP や UDP を用いた転送手段では不十分である。実時間通信データの信頼性を向上させる情報を付加する必要がある。そのようなプロトコルの一例が RTP (Real-time Transport Protocol) [1] として RFC で定められている。

RTP は映像・音声といったリアルタイムのデータをマルチキャストやユニキャストで転送するアプリケーションに適したプロトコルである。タイムスタンプなどの情報を付加する。RTCP (Real-Time Control Protocol) を用い RTP のデータ転送を制御する。RTP や RTCP はコネクションの概念がないため、コネクション型、コネクションレス型に関わらず動作可能である。

3.2 入出力機構：統一化された入出力・操作方法

PC にデータを入出力するためのハードウェアが存在し、それらを制御するための様々なデバイスドライバが存在する。仮想デバイスは同じ機能をもつ異なったハードウェアを単一の方法で操作することの出来る機能の提供を行う。

FreeBSD, Linux に代表される PC-UNIX には、音声の入出力方法を統一するための仮想デバイス /dev/dsp が存在する。

図1に /dev/dsp を使った入出力の概要図を示す。

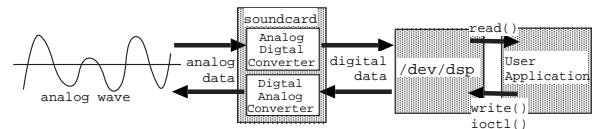


図 1: /dev/dsp の入出力の概要図

/dev/dsp を読み取ると、サウンドカードのマイク端子やラインイン端子から音声の入力が行え、書き込むとラインアウト端子から音声の出力が出来る。音質などの設定はデバイスを制御する `ioctl()` を利用して制御する。これはどのベンダーのどのサウンドカードに対しても利用可能で、統一された手法で制御し、同一フォーマットを取り出すことが出来る。

3.3 部品化：部品化と組み合わせ

映像・音声を扱う機能を機能毎に部品化し、部品の組み合わせにより、映像・音声を再生や加工することが可能となる。さらに部品化により機能毎に独立するため、機能追加時に追加部分以外に影響がない。その一例が Windows 上で動作する DirectX [2] の一部の DirectShow[3] である。

DirectShow にはフィルタと呼ばれるフォーマットの CODEC や入出力を行なう部品がある。フィルタの組み合わせで映像・音声を扱うアプリケーションを構築できる。DirectShow の開発キットには、GUI(Graphical User Interface) でフィルタを組合すことの出来る GraphEdit というツールがある。図2に GraphEdit を用いたアプリケーション構築の例を示す。DirectShow は部品の組み合わせで、映像・音声を扱うことが出来るが、PC 単体での映像・音声の扱いに特化しているため、ネットワークを介した映像・音声の入出力は自由に組み合わせることが出来ない。

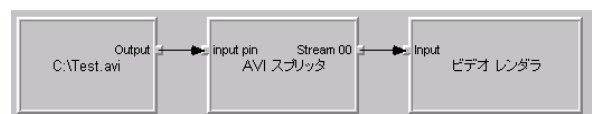


図 2: GraphEdit の例

4 研究内容：映像・音声の入出力・操作統一化手法

前章で示したように、既存の技術ではフォーマットやデータソースに非依存な操作性を提供できない。/dev/dsp は操作の統一化が行なわれているが、扱えるのは音声のみである。また、既存の機能の部品化及びその組合せ技術は、アプリケーションプログラマが機能を自由に組み合わせることが出来るが、ネットワーク入出力への対応は不

十分である．また，ネットワークの状態に対して最適なフォーマットを意識する必要がある．

以上から，本研究では以下の 3 項目を実現する．

1. 映像・音声のための統一化入出力機構の実現

映像・音声を扱う場合ならばデータソースに非依存な統一化された方法で操作可能な機構を実現する．

2. 映像・音声フォーマットの操作統一化の実現

映像・音声のフォーマットには様々な種類がある．例えば，非圧縮の方式もあれば，可逆圧縮や非可逆圧縮の方式もある．また圧縮のアルゴリズムもフォーマットによって異なる．本研究では，フォーマットの差異を吸収し，映像・音声への統一された API を実現する．

3. 状況に適したフォーマット判別機構

CODEC や，ネットワークの状態と言った条件によって実際に通信に用いるべき最適なフォーマットを判別する機構を実現する．

5 設計概要

本機構は 2 つの部分から構成される．図 3 に本機構の設計概要を示す．映像・音声のための入出力方法の統一化部分 (図 3 中 IO Wrapper) と，映像・音声のフォーマットに対する操作の統一化部分 (図 3 中 Format Wrapper) の 2 つである．

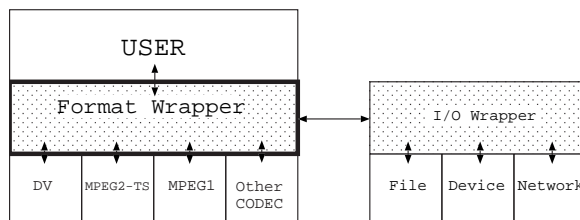


図 3: 設計概要

5.1 映像・音声の入出力方法の統一化

データソースを統一化して扱う．UNIX ではファイルディスクリプタのレベルで，ソケットとファイルの扱いは統一化されている．しかし，ファイルディスクリプタは `open()`，`close()`，`read()`，`write()` といった基本動作しか用意されていない．ネットワークやデバイスに対する実時間の映像・音声入出力はファイル入出力に比べ複雑で開発工程数が多い．例えば映像・音声転送プログラムを開発する場合，インターネット上の実時間通信用の標準プロトコルである RTP を利用すると工数が増える．映像・音声データを入出力する位置に関わらず，統一された方法で操作するにはこの工数の差を隠蔽する必要がある．本機構では IO Wrapper で独自のメソッドを提供して隠

蔽する．

アプリケーションプログラマは IO Wrapper に入出力するデータの位置を与える．実際は，データソースに適した方法で入出力を行なう．

5.2 映像・音声フォーマットに対する操作方法の統一化

映像・音声のフォーマットに対する操作の統一化は，各フォーマットを同一 API で扱えるように，フォーマット固有部分を隠蔽し，操作方法の統一化を行なう部分と，状態の通知を受け最適なフォーマットを判別する機構の 2 つの部分に分かれる．

5.2.1 操作方法の統一化

アプリケーションプログラマが特にフォーマット意識することなく映像・音声を扱う環境を構築するために，フォーマット固有部分を隠蔽し，操作方法の統一化を行なう．

利用者と実際に用いるフォーマットの間には，フォーマット固有部分の隠蔽と同一 API の提供を行なう，Format Wrapper を設ける．実際に用いるフォーマットに関わらず，アプリケーションプログラマは同一 API で操作可能になる．

RTP は扱うフォーマットごとにそれぞれ異なったペイロードタイプ [4] を持ち，場合によってはペイロードフォーマットを付加することが出来る．フォーマットの差異を意識することなく利用するには，RTP のフォーマット固有部分も隠蔽を行う．

抽象化を行うためには，映像・音声が共通にもつ情報を明確にする必要がある．

映像が共通に持つ情報の例として以下の 3 つが挙げられる．

- 単位時間あたりのフレーム数
- 画素数，解像度
- 色空間 (YCrCb, RGB)

音声が共通に持つ情報の例として以下の 3 つが挙げられる．

- サンプリング周波数
- ビットレート
- チャンネル数

5.2.2 状態によるフォーマット判別機構

ネットワークの利用可能帯域やパケット損失率，通信相手が持つ CODEC の種類などに依存して，利用可能なフォーマットが限定される．アプリケーションプログラマにフォーマットを意識させないためには，実際に転送に用いるべき最適なフォーマットを，ネットワークと通信相

手の状態に応じて、判別出来る機構が必要である。本機構では、Format Wrapper がフォーマット判別に必要な情報を受けとり、最適なフォーマットを判別する。

ここで、状態と呼ぶ情報は例えば以下の 5 つの要素である。

- CODEC
情報を伝えるためには、送信したデータをデコード出来る CODEC を通信相手が持っている必要がある。
- ネットワークの実効帯域幅
実効帯域幅により単位時間あたりに送受信可能なデータ量が決定される。実効帯域幅が狭くデータがそのままでは送信できない場合は、より高い圧縮率を持つ圧縮方法を用いるか、品質を下げる必要がある。
- ネットワークのパケット損失率
フレーム間圧縮を行なう場合は、パケット欠損時の欠損する情報量が大きいので、パケット損失率が高いネットワークには不向きである。
- ジッター成分
ジッターは信号転送や処理によって起こる時間的狂いを表す。フレーム間圧縮を行なう場合はジッターが大きいネットワークには不向きである。
- PathMTU
MTU(Maximum Transmission Unit) は、1 回の送信でネットワークデバイスが送ることの出来る最大のデータサイズである。PathMTU は通信経路上の最小の MTU 値である。PathMTU 以上のデータサイズのデータを送信すると通信経路上で分断化が起きる。分断化はジッターの原因となる。

6 これまでの活動と今後の予定

6.1 これまでの活動

学部 2 年次秋より、村井研究室に所属し、ネットワークとプログラミングに関して学んだ。学部 3 年次には PC から制御可能な無線受信機を用いて、その制御と音声の聴取を遠隔から可能とするソフトウェアを作成し、特に無線受信機の遠隔制御の抽象化と操作の統一化に注力し、成果報告として論文 [5] [6] を執筆した。

6.2 今後の予定

- D-VHS(MPEG2-TS) の転送システム
IEEE1394 を使うなど多くの点で DVTS [7] に似ている、DVTS を参考にし、映像・音声の入出力と操作の統一化に必要な共通項を洗い出す
- 本機構の設計と UNIX での実装。D-VHS と DV を対応させる。
- さらに他の CODEC も追加し、実用度を高める。

7 本研究により実現出来る世界

7.1 期待される成果物

本研究が実現すると以下のメリットがある。

- 映像・音声の為の統一化入出力手法の実現
- フォーマット依存のコードをユーザーアプリケーションから分離出来る。
- 環境変化に柔軟に対応できる。
未対応のフォーマットに対応する際に、そのフォーマット固有部分の実装を行なうだけで、本機構が提供するその他の機能を利用できる。

7.2 本機構を応用したアプリケーション

- 環境に適応可能な映像配信システム
入出力とフォーマットの操作統一化を行っている本機構を利用することで、環境に適応可能な映像配信システムが構築可能である。利用者には実際にどのフォーマットを使用しているかに関わらず、単一の操作方法で扱うことが可能であるため、ネットワークや受信者の状態に応じて実際に使用するフォーマットを変更することが容易である。
- 家電と親和性の高い映像配信システム
家電は、高機能化され、PC との親和性が高まっている。DV の普及により IEEE1394 端子が、デジタル BS 放送の開始により MPEG2-TS デコーダが多くの家電に搭載されている。これらを利用すれば、例えば DV を利用して撮影している映像を、転送経路にインターネットを、データの送受に PC を用い遠隔地の TV で鑑賞するといった利用方法が可能となる。現在、インターネットを利用して転送された映像は、PC で見るのが一般的である。本機構を利用したアプリケーションは TV などの民生用 AV 機器に出力することが出来る。さらに、新たな入出力端子が登場した場合等も対応可能である。
PC と民生用 AV 機器は利用スタイルが異なる。TV は一般家庭に広く浸透し、家庭内において家族が共有出来る位置におかれていることが多く、その映像出力機能は多くの人に利用されている。インターネットを用いた映像転送の最終出力を一般家庭に浸透している TV も利用可能となれば、今以上にインターネットを用いた映像・音声配信の実用性が高まり、映像・音声を利用したインターネット上のコミュニケーションの活発化が期待される。

7.2.1 具体例

図 4 に DVTS の動作例を示す。送信側、受信側において DV が無いと DVTS は動作しない。

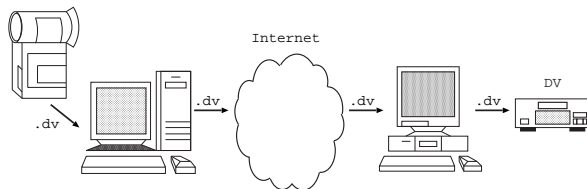


図 4: 既存の DVTS

従って DV から D-VHS へ通信を行う場合は送信側か受信側かのどちらかの PC で、本機構を利用した変換が必要となる。本機構はネットワークの状態を受け入れる事が出来る。従って、ネットワークの状態に応じ変換する位置を変更出来る。

例えば、ネットワークの実行帯域幅が狭い場合、DV フォーマットと MPEG フォーマットを比べると、MPEG フォーマットのほうが圧縮率が高く、転送に用いるフォーマットは MPEG の方が好ましい。従って送信側で変換を行う。

図 5 に帯域が狭い状況下で本機構応用アプリケーションが DV から MPEG への送信元での変換を行なう動作概念図を示す。

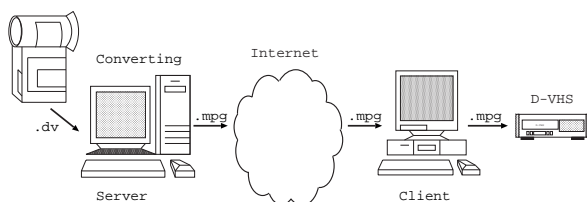


図 5: 応用アプリケーション，帯域が狭い場合

逆にネットワークの状態が無線の 802.11a のように帯域が広いが、有線の安定したネットワークに比べて、データ欠損率が高い場合、フレーム内圧縮を用いている DV フォーマットのほうが、データ欠損時におけるデータ欠損量に対する情報欠損量が少ない。従って転送に用いるフォーマットは DV フォーマットの方が好ましいので、受信側で変換を行う。

図 6 に帯域が広くデータ欠損率が高い状況下で、本機構応用アプリケーションが DV から MPEG への受信側での変換を行なう動作概念図を示す。

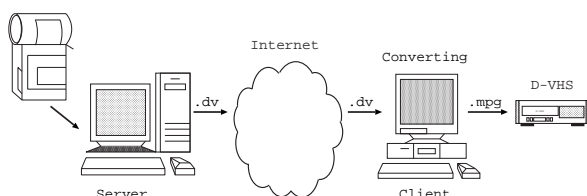


図 6: 応用アプリケーション，帯域が広くデータ欠損率も高い場合

8 政策・メディア研究科に進学を志望する理由

政策・メディア研究科では、インターネット技術に留まらず、政治や経済、教育など広範囲な研究が行われている。特に School On Internet(SOI[8]) や Digital Cinema Project のように、インターネットを中心とした新しいコミュニケーション環境に関する研究領域が開かれている。

本研究は、以上のような新しいコミュニケーション環境を構築する一要素技術であり、これらのプロジェクトに実際に適用され、フィードバックを受けることが非常に重要である。また本研究を進める上で、先進的なコンピュータ・ネットワーク環境が必要であり、コンピュータ・コミュニケーションに関する理論および実践経験を持つ研究指導者が不可欠である。

以上の理由から私は政策・メディア研究科への進学を強く志望する。

参考文献

- [1] Audio-Video Transport Working Group. RFC1889 RTP: A Transport Protocol for Real-Time Applications. <http://www.ietf.org/rfc/rfc1889.txt>, pages 1-75, Jan 1996.
- [2] Microsoft Corporation. Microsoft DirectX. <http://www.microsoft.com/windows/directx/>.
- [3] Microsoft Corporation. DirectShow for Windows XP. <http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dshow/htm/directshow.asp>.
- [4] Audio-Video Transport Working Group. RFC1890 RTP Profile for Audio and Video Conferences with Minimal Control. <http://www.ietf.org/rfc/rfc1890.txt>, pages 1-18, Jan 1996.
- [5] 入野 仁志, 久松 剛, 杉浦 一徳, 村井 純, and 中村 修. 無線受信機の遠隔操作および搬送波転送の研究. 情報処理学会第 63 回全国大会 講演論文集 (分冊 3), pages 657-658, Sep 2001.
- [6] 入野 仁志, 久松 剛, 志和木 愛子, 杉浦 一徳, 中村 修, and 村井 純. アナログ無線受信機のウェブベースの遠隔制御及び遠隔聴取の研究. マルチメディア通信と分散処理研究会, pages 19-24, Mar 2002.
- [7] Akimichi Ogawa. DVTS(Digital Video Transport System). <http://www.sfc.wide.ad.jp/DVTS/>.
- [8] SOI(School On Internet). <http://www.soi.wide.ad.jp/>.