

多様なコンテンツに対するメディアメタデータの付与とその利用

古田恒志[†] 河口信夫^{‡*} 稲垣康善[†]

[†] 名古屋大学大学院工学研究科

[‡] 名古屋大学情報連携基盤センター

* 名古屋大学統合音響情報研究拠点

furuta@inagaki.nuie.nagoya-u.ac.jp

近年、映像や音声などのデジタルコンテンツの普及に伴い、コンテンツに付与したメタデータを扱う技術に注目が集まっている。しかし、コンテンツにメタデータを付与する作業は一般に時間と労力を要する。本研究では、多様なコンテンツ中のシーンに関するメタデータを定義し、コンテンツを視聴しながら携帯端末のボタンを押す行為により手軽にメタデータを付与する手法を提案する。また、メタデータに対する分析手法や、メタデータを用いたコンテンツの高度な利用手法を提案する。また実際にメタデータを収集し、その分析を行いメタデータの有効性を確認した。

A Proposal of Media Metadata for Various Kinds of Contents, and Its Tagging, Analysis, and Use

Hisashi Furuta[†] Nobuo Kawaguchi^{‡*} Yasuyoshi Inagaki[†]

[†] Graduate School of Engineering, Nagoya University

[‡] Information Technology Center, Nagoya University

* Center for Integrated Acoustic Information Research(CIAIR), Nagoya University

Recently, metadata technology gets a lot more attention by popularization of digital multimedia contents such as video and audio. However, generally, tagging of metadata to contents requires time and energy. In this paper, we define the Media Metadata for various kinds of contents, and propose a method for easy tagging by pushing a button of the one's mobile device while watching the contents. We propose methods for analysis of metadata and advanced use of contents using metadata. We confirm the usefulness of the Media Metadata by the experimental collection and analysis.

1 はじめに

近年、DVD などのデジタル動画画像メディアの普及や、ブロードバンド・常時接続環境といったインフラの整備、HDD の大容量・低価格化などにより、マルチメディアコンテンツの利用が一般化し、頻りに視聴されるようになった。これに伴い、リッチコンテンツと言われるような、音声・動画や、インタラクティブ性を効果的に取り入れた質の高いコンテンツをユーザに提供しようとする動きが起りつつある。しかし、現状ではこのようなコンテンツが一般に広く利用されている状況にあるとはいえず、リッチコンテンツに対する需要が高まっている。

一方で、コンテンツに対しメタデータを付与し、コンテンツを効率的、また柔軟に扱う技術に注目が集まっている。メタデータとは、広義には“データに関して記述されたデータ”である。文書・映像・音楽などのデータに対して付与され、オリジナルデータの意味や構造、内容を記述するデータであり、元データの検索やブラウズ、ナビゲーション、編集作業など、幅広い目的に利用されている。メタデータを扱った研究として、オンラインオブジェクト検索のためのメタデータ表現方法⁶⁾や、特徴記述子を用いた略画による画像検索手法⁷⁾、コンテンツに対するメタデータを登録した XML DB の利用によるシーン検索⁸⁾などが挙げられる。2002 年 3 月には Moving Picture Experts Group(MPEG) が策定したメタデータの表記方法に関する標準規格・MPEG-7⁵⁾ が ISO/IEC によって国際規格として承認されており、今後メタデータを扱った研究開発がより一層盛んとなることが予想される。

本研究では、従来のようにデジタルコンテンツを単純に観る、聴くといった利用法に留まらず、コンテンツをより柔軟に、豊かに利用するための一連のフレームワークを提案することを目的とする。コンテンツ中の各シーンに対し付与されるメタデータを用いることによって、既存コンテンツの高度な利用を実現する。

一般に、メタデータを利用したアプリケーションを開発するにあたっては、オリジナルデータに対しメタデータをどのように付加するかということがひとつの問題となる。映像の色・形状や動き、シー

ンの切り替わりのような、データから抽出可能な特徴量を表すメタデータを自動入力することは可能であるが、例えば、サッカー中継映像においてその試合のハイライトシーンを示すデータや、映画の中の泣けるワンシーンを指すデータなど、意味内容記述のメタデータは計算機による自動入力が困難なため人手による入力が主流である。これは時間と労力を要する。

本稿では、一般に入力が困難である意味内容記述メタデータを手軽に付与するための手法として、携帯端末によるマーキング手法を提案する。本手法において、コンテンツ視聴者は、ユーザインタフェースとしてボタンを有する小型の端末を視聴中に携帯する。その端末は、コンテンツを提供しているシステムとネットワーク接続し、利用中のコンテンツの情報を取得する。作成者は、コンテンツを視聴しながら、任意のシーンに差し掛かった時に手元にある端末のボタンを押す(これをマーキングと呼ぶ)という単純なインタフェースによって、煩雑な操作をすることなく手軽にメタデータを付与することができる。この作業は特殊な知識や技能を持たない一般のユーザにも可能な操作であるため、この手法を実行できる環境をコンテンツを視聴するエンドユーザに提供することにより一般ユーザがメタデータ作成者となり、ひとつのコンテンツをあらゆる観点から見た複数のメタデータを取得することが期待できる。

さらに本稿では、上記の付与手法により得られる範囲のメタデータのみを用いて、ユーザの興味を引くコンテンツ内シーンの検出、ユーザ間の興味類似性の推定を可能とするメタデータ分析手法を提案し、ユーザごとの要望に応じたデジタルコンテンツの高度な視聴、メタデータの配布によるユーザ間コミュニケーション支援、コンテンツとメタデータの同期利用、などといったコンテンツやメタデータの高度利用を可能とする。

また、被験者実験を行って実際にメタデータを収集し、ユーザのマーキングの傾向などの分析を行うとともに、マーキング手法によるメタデータ利用の有効性を確認した。

本稿の概要を示す。本研究で扱うメタデータは、これら付与手法や利用法を考慮に入れた上で定義したもので、メディアメタデータと名付けた。これに関しては 2 章で述べる。3 章で、コンテンツに

表 1: メディアメタデータが保持する要素

情報		役割
メディア ID		コンテンツを特定
シーン	シーン特定子	コンテンツ中のシーンを特定
	クリエイタ ID	メディアメタデータ作成者
	シーン属性	シーンのタイプ
	コメント	注釈

は必須の情報

対しメディアメタデータを効率的に付与するマーキング手法の詳細を述べる。4 章では、メディアメタデータの分析手法、及び、メタデータを用いたマルチメディアコンテンツの高度利用や、コンテンツを介したメディアメタデータの利用を提案する。5 章でプロトタイプシステムの実装について述べる。6 章で本手法によって得られるメタデータを分析し、その考察をする。7 章では本稿をまとめ、今後の課題を挙げる。

2 メディアメタデータの定義

本章では、後の章で提案するメタデータ付与手法と利用法を考慮した上で、本研究で取り扱うメタデータを定義する。3 章で提案するマーキング手法は、コンテンツ中のワンシーンに対する情報を付与するものである。本研究では、あるコンテンツにおけるこれらの情報を 1 つもしくは複数でまとめ、これをひとつのメタデータとして定義する。このメタデータをメディアメタデータと名付ける。メディアメタデータは、あらゆるコンテンツに対し付与されることを想定し、デジタルコンテンツそのものを視聴している時だけでなく、競技場でのサッカー観戦時など、後にデジタル化されるものを生で見ている時にも付与できる仕様になっている。

メディアメタデータが保持する主な情報をまとめたものを表 1 に示す。メディアメタデータが保持する情報のうち、メディア ID とシーン特定子の二つがシーンを指し示すためのものであり、残りの情報はそのシーンに依存した諸情報である。メディアメタデータが指し示すシーンが定まらなるとシーン依存情報は意味を成さなくなるため、メディア ID とシーン特定子の二つの情報はメディアメタデータにおいて必須のものとす。

メディアメタデータ定義の詳細については参考文献¹⁾を参照されたい。

3 メディアメタデータの付与手法

本章では、メディアメタデータを実際に付与する手法について述べる。一般のコンテンツ視聴者がメタデータを付与できる形式にすることにより、コンテンツが視聴されるその都度にメタデータが蓄積されていくことを可能にする。メディアメタデータの要素の中で必須のものはメディア ID とシーン特定子であるが、シーンを特定する毎にこれらをユーザが入力するのは非常にコストがかかり現実的でない。メディアメタデータを生成する手法には、ユーザからの入力が容易であり、簡素なインターフェースが求められる。これを満たす手法として、マーキングによるメタデータ付与手法を提案する。

3.1 マーキング手法の概要

本手法は、ユーザがマルチメディアコンテンツを利用しながら、任意のシーンに差し掛かった時に手元の機器のボタンを押すこと（マーキング）により、そのシーンを示すメディアメタデータを付与する手法である。本手法を行うユーザは、細かい設定や入力を考えることなく、ボタンを押すという手軽かつ簡素な動作のみによってメディアメタデータを付与することができる。

本手法は、ユーザにコンテンツを提供する機器と、ユーザがマーキング動作を行う機器とがネットワークで接続し、コンテンツ情報やマーキング情報を互いに通信することによって行われる。ひとつのコンテンツ提供機器に対し、複数のマーキング機器が接続される場合もあり得る。観客を収容した映画館での利用などがこれにあたる。本手法の要件やプロトコルなどは参考文献¹⁾を参照されたい。

3.2 マーキング手法の拡張

マーキング手法は、メタデータにコメントを入力するのには適さない。メタデータ作成者にとって煩雑さの少ない入力方法として、以下のようなマーキング手法のインターフェース拡張が考えられる。

3.2.1 チャットによるマーキング手法

ユーザはコンテンツを視聴しながらチャットを行う。ユーザが発言をした瞬間に、その発言をコメントとし、発言時刻に対するメタデータを付与する。コメントが付随したメディアメタデータを即時付与できるのがこの手法の利点である。但し以下のような欠点も持つ。

- 発言の書き込み中はコンテンツ視聴に集中できない
- ユーザにある程度の速度のタイピング能力を要求する
- マーキング機器にテキスト入力装置を備えなければならないため、マーキング機器の携帯性が低下する。

3.2.2 マイク入力によるマーキング手法

インターフェースとしてマイクを用意し、作成者の発言を音声入力する方法が挙げられる。音声入力を発話単位ごとにメディアメタデータとして記録する。ユーザにから見れば特別な技能を必要とせず、また入力操作も単純で、手軽なメタデータ付与が行える。

4 メタデータの利用法

本章では、メディアメタデータを用いたコンテンツの高度な利用や、コンテンツを介したメディアメタデータの利用・流通を提案する。

4.1 パーソナルクリッピング

コンテンツを視聴するユーザが、以下のような要求を持つことがある。

- 映画館で映画を鑑賞している時に気になったシーンがあったので、見終わった後にそこだけもう一度見たい。
- サッカー観戦に行って、素晴らしいファインプレーを見たので、そこだけもう一度見たい。

しかし、あるワンシーンが見たくなった時、その動画コンテンツを用意してサーチ再生を繰り返し、そのシーンを探すといった一連の作業は煩雑であり、手軽には行いにくいと言える。また、このような場合に長編のコンテンツを丸々ユーザに供給するのは料金や通信にかかるコストが高い。

本研究で提案するメディアメタデータとその付与手法により、上述のような要求を満たすことが可能となる。本研究で定義するメディアメタデータによりユーザはコンテンツ中のワンシーンを記録し、後に参照できる。例えば映画館で映画を見ながら、ユーザは手元の携帯端末で気に入ったシーンを自由にマーキングする。その後、ユーザがマーキングによって得られたメディアメタデータを指定するだけで、それに記録されたメディア ID とシーン特定子をもとに、サーバから該当シーンをクリッピングしたコンテンツをダウンロードする。ユーザはマーキングしたシーンをいつでも直接的に参照できる。さらに、マーキング機器をユーザが携帯することにより、利用場所を自宅などの特定の場所に限定することなく、映画館やテーマパークなど、ユーザの行く先々での利用を可能とする。

その他のサービスとして、コンテンツ全体ではなく、マーキングしたワンシーンのみを比較的低料金で供給するサービスが考えられる。また、サッカーの競技場において試合会場に設置された複数のカメラから試合を録画しておき、それをデータ配信するシステムを用意することにより、上述の例と同様にゴールシーンやファインプレーなどユーザが気に入ったシーンを供給することができる。また同シーンを別アングルのカメラから見たシーンを供給するなどが考えられる。

本手法では、メタデータによってコンテンツ中のワンシーンを記述するため、そのシーンの動画像や音声などをそのままの大容量データとして保存する必要はなく、携帯電話など記憶容量の限られた端末においても多量のシーンを保持しておける。

4.2 シーンに対する複数ユーザの関心度

コンテンツを視聴した複数のユーザが行ったマーキングにより、コンテンツ中の各シーンに対するマーキングの付与数が獲得できる。あるシーンに対し多数のユーザがマーキングを行ってれば、そのシーンはユーザの関心を多く集めているシーンであると考えられる。

4.2.1 関心度の計算

単純には、シーンに対するユーザの関心度の尺度は、コンテンツを視聴した全ユーザのうちそのシーンにマーキングをしたユーザの割合である。ここで、ユーザごとのマーキング頻度の個人差を考える。

ユーザには、頻繁にマーキングを行うユーザと、稀にしかマーキングを行わないユーザが存在する。ほとんどマーキングしないユーザが行ったマーキングは、多くマーキングしているユーザのそれよりも重視すべきと考えられる。よって、シーンに対する関心度計算にユーザ毎のマーキング情報量を適用する。あるコンテンツに対しユーザ U_i がマーキングによるメタデータ付与を行ったとする。メタデータには k 種類の属性値が与えられている。 U_i が属性値 $A_t (1 \leq t \leq K)$ のマーキングをする事象を m_{it} 、何もマーキングをしない事象を m_{i0} とする。 $m_{ix} (0 \leq x \leq K)$ が発生する確率を $P(m_{ix})$ 、 U_i が属性値を問わず単純にマーキングをする確率を $P(m_i) = \sum_{x=1}^K P(m_{ix})$ とする。この時、 U_i がマーキングした時の情報量 $I(m_i)$ は

$$I(m_i) = -\log_2 P(m_i) \quad (\text{ビット})$$

となる。 $P(m_{ix})$ の求め方はマーキングの形式によって異なる。マーキングがポイント形式である場合は、コンテンツを時間間隔 α 毎に区間に区切り、全区間数において m_{ix} が生じた区間数の割合となり、マーキングがクリップ形式である場合、 U_i の属性 A_x のマーキングのクリップ長の総和がコンテンツの総時間に占める割合となる。

また、ユーザ U_i に対する重み $W(U_i)$ を

$$W(U_i) = \frac{I(m_i)}{\sum_j I(m_j)}$$

のように正規化した情報量と定める。あるシーンに対するユーザの関心度は、そのシーンをマーキングしたユーザの重みを合計したものとす。この関心度はあるコンテンツ内での相対的なものであり、全コンテンツを含めた絶対的なものではない。

4.2.2 関心度の利用法

コンテンツ中において関心度の高いシーンを優先的にピックアップして繋ぎ合わせることににより、ハイライトシーンを集めたダイジェスト映像を自動作成できる。これを応用して、映画の封切以前に催される試写会においてメディアメタデータ収集を行い、得られた情報に基づいて予告編映像を作成することもできよう。メタデータをその属性値によって分類して関心度を計算し、面白いシーンと泣けるシーンだけ、のように要求に応じたシーンを集めたダイジェストなども作成可能である。またメタデータをクリエイタによって分類して“20代男性の関心が高いシーン”などのようにシーン検出し、マーケティングにも利用できる。

4.3 ユーザ間のマーキング類似度

複数のメタデータ作成者が各自の主観に基づき印象に残ったシーンをマーキングしたとする。興味の似ている二人のメタデータ作成者が同一のコンテンツに対しメタデータ付与を行うと、それぞれのマーキングしたポイントの系列には似た傾向が表れることが予想できる。逆に、二つのマーキング系列の内容から、その作成者同士の興味の類似度を推測することが考えられる。

4.3.1 マーキング類似度の計算

ここでは、ユーザ間のマーキングの相互情報量、すなわち、あるユーザがマーキングした情報を知ったときに他のユーザがマーキングした情報をどれほど推定できるかに基づいてユーザ間のマーキング類似度を決定する。

2 ユーザ U_i, U_j がある同一のコンテンツに対しマーキングによるメタデータ付与を行ったとする。このとき、 U_i, U_j の行ったマーキングの相互情報量 $I(M_i; M_j)$ は、

$$I(M_i; M_j) = \sum_x \sum_y P(m_{ix}, m_{jy}) \log_2 \frac{P(m_{ix}, m_{jy})}{P(m_{ix})P(m_{jy})}$$

(ビット) となる。 $P(m_{ix})$ は 4.2.1 節と同様に求める。 U_i, U_j のマーキング類似度 $SIM(U_i, U_j)$ は

$$SIM(U_i, U_j) = \frac{I(M_i; M_j)}{H(M_i, M_j)}$$

とする。ここで $H(M_i, M_j)$ は U_i, U_j のマーキングの結合エントロピーであり、次式で求められる。

$$H(M_i, M_j) = - \sum_x \sum_y P(m_{ix}, m_{jy}) \log_2 P(m_{ix}, m_{jy})$$

$H(M_i, M_j)$ は U_i, U_j のマーキング全体の情報量を表しており、それに対する相互情報量の割合を U_i, U_j のマーキング類似度としている。図 1 に、類似度と結合エントロピー、相互情報量の関係を示す。 $SIM(U_i; U_j)$ の値はコンテンツに依存する。

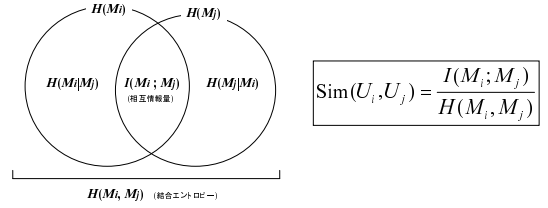


図 1: 類似度，結合エントロピー，相互情報量の関係

4.4 コンテンツとコメントの同期利用

コンテンツ製作者からの情報やコンテンツに対する視聴者の反応をコンテンツと共に供給し同時に利用する仕組みがある。シットコムやオーディオ・コメンタリーがそれである。

シットコムは Situation Comedy の略であり、登場人物と舞台を固定し、毎回生じる様々な状況を題材にしたコメディである。多くのシットコムでは撮影現場に観客を配置し、演技と共に観客のリアクション（笑い声・喝采・ひやかし・ブーイングなど）を同時収録することが特徴である。これは臨場感の向上や視聴者の感情移入に効果がある。

オーディオ・コメンタリーは映画などの音声解説の一種で、各シーンに対する監督や出演者の解説を副音声として聞くことができるものである。メイキング映像などの映像特典とは異なりシーンと共に随時解説が得られるため視聴者にとってわかりやすく、また解説に監督や俳優の個性が表れるのが魅力である。

オーディオ・コメンタリーやシットコムにおける観客のリアクションは、ワンシーンに対する製作者や視聴者のコメント（これはメタデータに他ならない）を集めて本来のコンテンツの上に重ねたものと考えられる。しかし、これらはコンテンツに包含されて供給されているため、予めコンテンツの製作者によって用意された固定のものしか利用できず、後からの追加や修正もできない。また、自分の友人が付けたコメントリーを聞きたいといった利用もできない。そこでメディアメタデータを用いたコンテンツとメタデータの同期再生を提案する。視聴者はコンテンツ利用時にそのコンテンツのメディアメタデータを取得し、視聴中にメディアメタデータの指すシーンに差し掛かるとそのコメントを再生する。再生の仕方はコメントのデータ型により、音声型ならば副音声で再生、文字列型ならば字幕のように表示するなどの形態になる。

従来と本手法それぞれのコンテンツとコメントの作成と供給の様子を図 2 に示す。マーキングによるメタデータ付与手法により、どのユーザでもあらゆるコンテンツに自由にコメントを付加することができ、インタラクティブ性が増す。一旦供給したコンテンツに後からコメントリーを追加することも可能であるし、好みの映画評論家を選んで評論を聞くこともできる。同期利用するメタデータを変更することで、同じコンテンツであっても異なる楽しみ方ができるため、一度製作したコンテンツに手を加えずに再利用することができる。これは、近年上昇する傾向にあるコンテンツ製作費を削減するのに有効である。

4.5 フィルタリング

4.5.1 メタデータのフィルタリング

メタデータの流通を考える。多数のメタデータ作成者が様々なコンテンツに対しメタデータを付与していくと、その量は膨大なものとなる。ユーザがメタデータサーバからメタデータを取得する際、膨大なメタデータから自分の利用したいものをどのように選んだらいいのだろうか。メタデータをフィルタリングする手段として以下のようなものが考えられる。

- 友人や著名人など、自分の利用したいメタデータ作成者のリストを作っておき、彼らを作成者とするメタデータを検索する。自分好みのメタデータ作成者を発見した場合はリストに

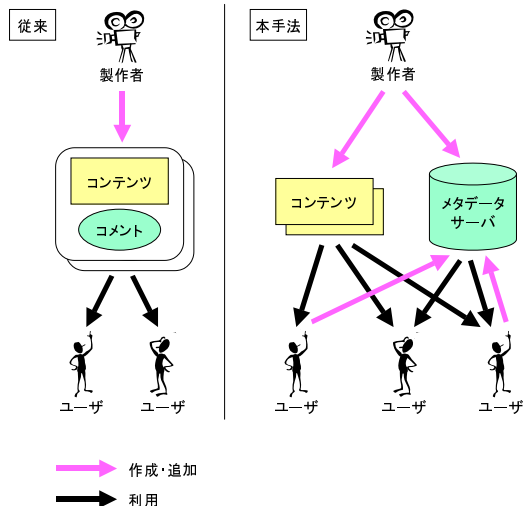


図 2: コンテンツとコメントの作成と供給

新たに追加する。4.3 節で述べたマーキング類似度は、自分と興味の近いユーザーを探す手掛かりとなる。

- メタデータ定義を拡張して評価値を持たせ、これが高いメタデータを選ぶ。メタデータは作成時には一定の評価初期値を持つが、ユーザーが利用したメタデータを評価すると評価値が更新される。多くのユーザーが評価をし更新を繰り返すことにより、信頼の高い評価値となる。

4.5.2 コンテンツのフィルタリング

膨大な数のコンテンツからユーザーに有益なコンテンツを検索する技術として、他ユーザーの評価に基づいてコンテンツを推薦する協調型フィルタリングについての研究がなされている⁹⁾¹⁰⁾¹¹⁾。4.3 節のユーザー間マーキング類似度を用いて自分と興味が似ているユーザーを発見できれば、そのユーザーが高い評価を付けているコンテンツは自分にとっても魅力のあるコンテンツである可能性が高い。ユーザーの行ったマーキングからコンテンツに対する評価を推測すれば、メディアメタデータを用いた協調型フィルタリングによって膨大なコンテンツの中からコンテンツを探す手掛かりが得られる。

5 実装

マーキングによるメディアメタデータ付与手法に基づき、パーソナルクリッピングを行うプロトタイプシステムを Java 言語 (JDK1.3, JMF2.1) を用いて実装した。図 3 はその外観である。コンテンツ提供機器として動画を再生する Player、ローカル形式のマーキング機器として Marker をソフトウェア的に実装した。Marker は 3 種類のボタンを備え、ユーザーがマーキングする時の感情に応じて押し分けられるようにした。また Player の再生・停止などを遠隔操作できる PlayerController を実装した。図 3 の左下のウィンドウはエミュレーション環境での端末間の接続状況などを表示するものである。また 2 つある Player のうち、左のものは競技場実際の試合、右のものは自宅の PC を想定している。コンテンツとして、MPEG2 形式の動画ファイルを用意し、メディア ID は、動画ファイルのファイル名とファイルサイズから決定するものとした。Player と Marker はそれぞれ通信機能を備えており、Player はコンテンツのデータベースからコンテンツを入手し再生する機能を持つ。機器間通信には cogma³⁾⁴⁾ システムを利用している。Player と Marker は互いに連携し、マーキングによってメディアメタデータ付与を行う。メディアメタデータによって記録されたシーンは Player で再生できる。ひとつの Player に対し複数の Marker が接続することも可能であり、また Marker がある Player と接続して付与したメディアメタデータを別の Player で利用することも可能である。

また、ブラウザからの要求に応じて Marker を制御するサーバを実装し、第三世代携帯電話 FOMA を用いてマーキングを行うシステムを構築した。マーキングによって記録されたシーンは、FOMA

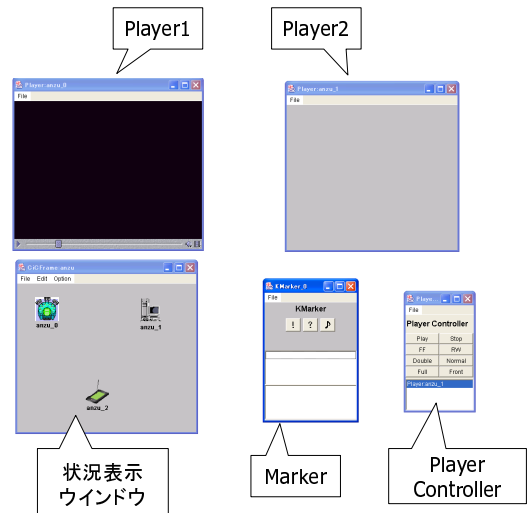


図 3: エミュレーションによるマーキングシステム

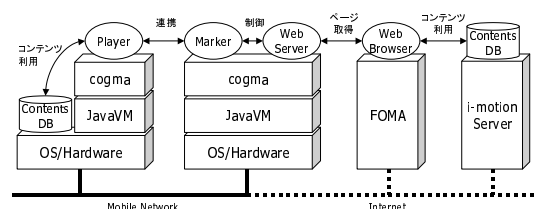


図 4: プロトタイプシステム構成図

の i モーション¹ コンテンツとして閲覧することができる。この外観を図 5 に示す。またプロトタイプシステムの構成図を図 4 に示す。

6 メタデータの収集・分析

被験者実験により実際にメディアメタデータを収集し、分析を行った。

6.1 概要

本実験の被験者は大学生 7 名、大学院生 14 名の計 21 名 (平均年齢 22.8 歳) である。実験で視聴するコンテンツとして表 2 に挙げる 3 本の映像作品^{12, 13)}を用意した。被験者を 10 人と 11 人の 2 グループに分けた上で各コンテンツを視聴させ、被験者の自由意志に基づき印象に残ったシーンにおいてマーキングさせた。マーキングには実機ではなくソフトウェア的にエミュレーションしたシステム (図 6) を利用した。マーキングシステムは GUI として 4 種類のボタンを備え、各ボタンにはそれぞれ属性値 (面白い, 泣ける, つまらない, その他) を持たせてある。被験者がマーキングを行う際、その時の印象によってボタンを押し分けることになる。「面白い」「泣ける」「つまらない」の 3 つに当てはまらない場合は「その他」を押すこととした。コンテンツ視聴後には、各コンテンツに対する評価を 1 から 5 の 5 段階で主観評価させた。また、被験者が行ったマーキングそれぞれについて、マーキングを行った位置の 10 秒前からマーキ

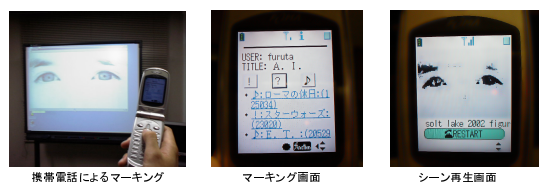


図 5: 携帯電話によるマーキングの外観

¹ 受信速度最大 384kbit のパケット通信により映像や音などのコンテンツを提供する NTT ドコモのサービス。



図 6: マーキングシステムの外観

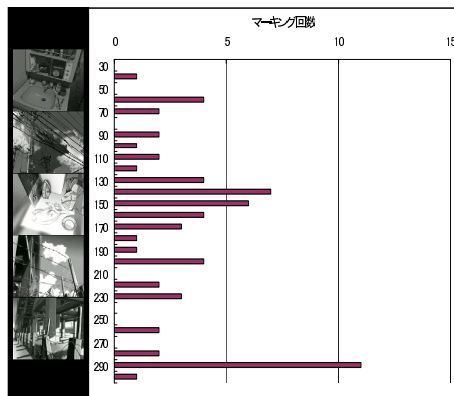


図 7: 「彼女と彼女の猫」におけるマーキングの分布

ング位置までのシーンを見せ、それによって自らがマーキングしたシーンを思い出せるかどうかを同じく 1 から 5 の 5 段階で主観評価させた。さらに、被験者が印象を抱いたシーンとそれに対するマーキングとの間に生じるタイムラグの調査のため、被験者が行ったマーキングそれぞれに、どのシーンに対してマーキングを行ったのか、そのシーンの開始位置と終了位置を指定させた。

6.2 結果と分析

6.2.1 マーキングの傾向

実験結果として、表 3 に、一部ユーザの 3 作品全体に対する平均評価値、マーキングした合計回数、終了時点までの平均時間差を示す。マーキングは被験者の自由意志に基づいていることから予測できるが、マーキングした回数は、コンテンツや被験者によってかなり幅が大きい。また、印象を持つシーンが始まってからボタンを押すまでのタイミングも被験者やシーンによってばらつきがある。中には飛び抜けてマーキング回数が多いヘビーユーザも存在していた。

図 7 は「彼女と彼女の猫」におけるマーキングの分布を、縦軸に時間、横軸にはその区間中のマーキング回数を取って表したものである。図の左には各時間区間におけるコンテンツのクリップを載せた。コンテンツ中に、マーキングの集中しているところ、ほとんどマーキングがされないところがある、コンテンツの終了間際にマーキングが集中している、などの特徴が確認できる。

6.2.2 コンテンツに対する評価とマーキング回数

図 8 は、各コンテンツにおいて、コンテンツに対する評価値で被験者を分類し、それらの被験者毎にマーキング回数を平均した値を調べたものである。一般に、ユーザがコンテンツに対し高評価を付けている時は多くマーキングをすることが裏付けられた。

6.2.3 マーキング位置からの注目シーン検出

図 9 は、実験で得られた有効なマーキングデータ 644 個について、被験者がそのマーキングを行った際に注目していたシーンの開始時点と終了時点マーキングした位置から見た相対時間で算出し、マーキングした位置から見たシーンがどのように分布しているかを調べたものである。マーキングした位置から見た相対時間を 1 秒間

表 2: 本実験で使用した映像作品

作品名	制作者	映像の全長
彼女と彼女の猫	©Makoto Shinkai/CoMix Wave	4 分 57 秒
ほしのこえ	©Makoto Shinkai/CoMix Wave	25 分 33 秒
機脈虫 零	近藤勇一	16 分 04 秒

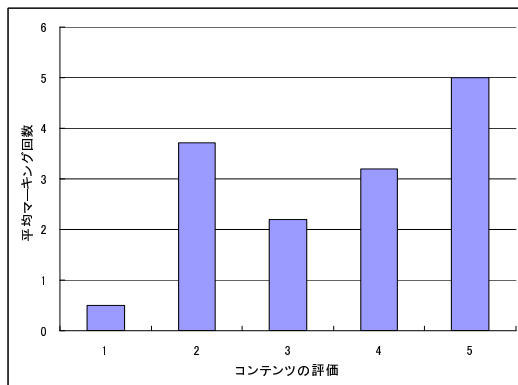


図 8: 「彼女と彼女の猫」におけるコンテンツ評価別の平均マーキング回数

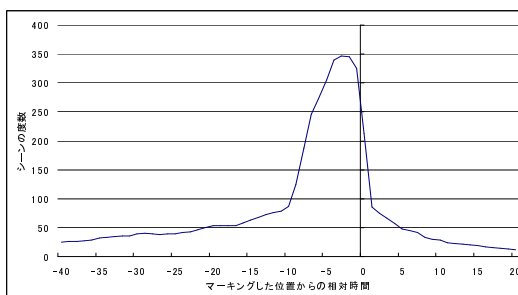


図 9: 被験者が注目していたシーンの分布

隔で区間に区切り、その区間内にマーキングされているシーンをカウントした。縦線が入っている位置は相対時間 0、すなわちマーキングしたポイントを表している。

被験者が印象を持ったシーンは、マーキングした位置から 4～3 秒前の区間をピークとし、209 秒前～77 秒後の区間まで幅広く分布していたが、ほとんどは 10 秒前～5 秒後程度の区間に集中している。パーソナルクリッピングを行う際は、マーキングした位置から見てこれらの区間をユーザに提供すればユーザの要求を満たすことができると考えられる。

図 9 のグラフを積分すると、被験者がマーキングしたシーンの時間長の総和となる。いま、ユーザのマーキングに対し、コンテンツからクリップしたシーンを提供することを考える。クリップできるシーンの長さが決まっている場合、どの区間をクリップするとユーザが意図していたシーンに近くなるだろうか。それには、その区間において図 9 のグラフを積分した値が最大になるように区間の位置を決めればよい。図 10 は、この考えに基づき、クリップできるシーンの長さが 10 秒間、20 秒間、30 秒間、40 秒間の時に、どの位置を中心にシーンをクリッピングすればユーザの要求シーンを多くカバーできるかを調査したものである。縦軸は、その位置を中心として指定秒数の区間を積分した値が、全体の積分値に占める割合を表す。例えば、クリップできるシーンの長さが 30 秒間の時は、マーキングされた位置から 9 秒前の時点を中心に（つまり 24 秒前～6 秒後の区間を）クリッピングすれば、ユーザの要求を満たす可能性が最も高いといえる。

表 3: 3 作品全体での実験結果（一部）

被験者	平均評価	回数	開始時点 (sec)	終了時点 (sec)
1	2.33	15	-11.023	-5.365
2	2.67	19	-7.761	-0.254
3	3.33	8	-41.798	-1.761
4	3.00	37	-3.926	-0.530
5	3.00	17	-11.660	3.967
平均	3.00	32.29	-16.272	-0.845

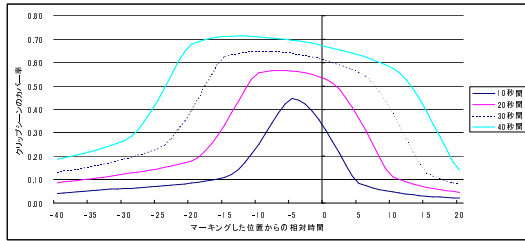


図 10: 各クリップ長毎のシーンカバー率

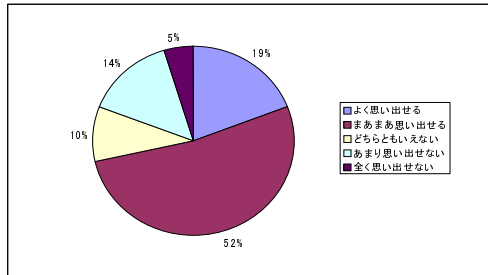


図 11: シーンクリッピングの有効性の主観評価

6.2.4 パーソナルクリッピングの有効性の評価

図 11 は、マーキングした位置からみて前の 10 秒間を被験者に見せた時、被験者が自分の行ったマーキングを思い出せたかを 5 段階評価してもらい、それを集計したものである。「まあまあ思い出せる」「よく思い出せる」を合計すると全体の 7 割を越え、多くの被験者はパーソナルクリッピングが有効と感じていることがわかる。

6.2.5 マーキング類似度の妥当性の評価

図 12 は、各コンテンツ毎に、縦軸にコンテンツ評価値の差、横軸に類似度を取って被験者のペア 210 組をプロットしたものである。最小二乗法により線形近似を行いその直線を書き入れた。どのコンテンツにおいても、類似度が高い被験者ペアほどコンテンツ評価値の差は低くなっており、マーキング類似度で興味の近いユーザを探すことの妥当性が確認できた。

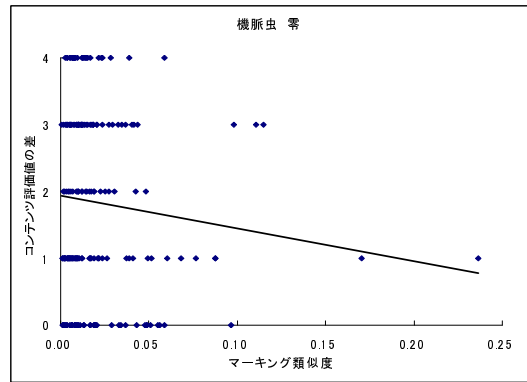


図 12: 「機脈虫 零」における類似度と評価値差

キング機器や、ボタン連打に応じて強い度合いのマーキングができる機器などが考えられる。

その他、ユーザ間のマーキング類似度はコンテンツに依存しているが、複数コンテンツに対する類似度をマージして平均の類似度を求める手法を考えている。さらに、ユーザ間のマーキング類似度をもとにコンテンツ間の類似度を考えたい。

加えて、今後は被験者やコンテンツ数を増やしたより規模の大きい実験を実施し、高精度の分析・評価を行いたいと考えている。

参考文献

- [1] 古田恒志, 河口信夫, 稲垣康善: “モバイルネットワーク上におけるメディアメタデータの付与とその活用”, マルチメディア, 分散, 協調とモバイル (DICOMO2002) シンポジウム論文集, pp.233-236(2002)
- [2] Gracenote, Gracenote CDDB Music Recognition Service, <http://www.cddb.com/>
- [3] 河口信夫, 稲垣康善: “cogma:動的ネットワーク環境における組み込み機器間の連携用ミドルウェア”, 情報処理学会コンピュータシステム・シンポジウム Nov. 2001, pp.1-8(2001)
- [4] こぐまプロジェクト, Cooperative Gadgets for Mobile Appliances, <http://www.cogma.org/>
- [5] José M. Martínez: “MPEG-7 Overview”, <http://mpeg.telecomitalialab.com/standards/mpeg-7/mpeg-7.htm>
- [6] DCMI: “Dublin Core Metadata Element Set, Version 1.1: Reference Description, DCMI Recommendation(1999)”, <http://dublincore.org/documents/dces/>
- [7] 青木 秀一, 青木 輝勝, 安田 浩: “略画による画像検索への MPEG-7 の応用”, 研究報告「オーディオビジュアル複合情報処理」No.035(2001)
- [8] 東芝: “MPEG-7 を利用した XML マルチメディア・データベース”, <http://www.toshiba.co.jp/mmlab/tech/d03.htm>
- [9] 森, 山田: “ブックマークエージェント: ブックマークの共有による情報検索の支援”, 電子情報通信学会論文誌 Vol. J83-D-I, No. 5, pp.487-494 (2000)
- [10] Upendra Shardanand: “Social Information Filtering for Music Recommendation”, MIT Media Lab, <http://agents.www.media.mit.edu/groups/agents/publications/lingo/shard-thesis/>
- [11] Ian M. Soboroff: “Collaborative Filtering with LSI: Experiments with Cranfield”, UMBC CSEE Technical Report CS-TR-98-01(1998)
- [12] 新海誠: “DVD BOOK ほしのこえ”, 徳間書店, ISBN コード 4-19-861554-3(2002)
- [13] 近藤勇一: “機脈社”, <http://www.edit.ne.jp/kondo/>

7 おわりに

7.1 まとめ

本研究では、デジタルコンテンツをより豊かに利用するための様々なフレームワークを提案した。メディアメタデータによってシーンとそれに関連した情報を記述し、マーキング動作によってこれに付与する手法を提案した。この手法では、多様なコンテンツに対するメタデータを手軽な操作によって付与することができ、一般のコンテンツ視聴者から様々なコンテンツのメタデータを多量に収集することが期待できる。

また、収集したメディアメタデータの分析手法や、メタデータを用いたコンテンツの高度利用手法、メタデータの流通など、従来よりもインタラクティブ性の高いコンテンツ利用を提案した。

さらに、実験により実際にメタデータを収集してその分析を行い、その有効性を確認した。

7.2 今後の課題

今後の課題として、コンテンツ ID の割振りが挙げられる。例えば同一の映画であっても、映画館上映時とテレビ放映時では構成が異なることが多くある。これらコンテンツの ID をどう扱うかは検討する必要がある。同じ映画を、映画館で見た時や、テレビ放映で見た時のどちらでも、同一のコンテンツに対してメタデータ付与ができるようにすることが望ましい。

また、マーキングを行うシステムのより高度なインタフェースの考案と実装が挙げられる。例えば、感情の度合いをマーキング時に同時入力できるようなシステムとして、感圧式ボタンを備えたマ