# 音・画像情報を用いた映像短縮再生法の評価実験

青柳　滋己† 　光来　健一† 　佐藤　孝治† 　高田　敏弘† 　菅原　俊治†

尾内理紀夫††

† NTT 未来ねっと研究所
〒 180–8585 東京都武蔵野市緑町 3–9–11
†† 電気通信大学電気通信学部情報工学科
〒 182–8585 東京都調布市調布ヶ丘 1–5–1
E-mail: †{aoyagi,kourai,koji,takada,sugawara}@t.onlab.ntt.co.jp, ††onai@cs.uec.ac.jp

**あらまし**　衛星放送による多チャンネル化やインターネットのブロードバンド化により映像情報が膨大になりつつあり、短時間で必要な映像情報を選び出すことが今後重要になると思われる。我々は映像中に含まれる音情報と画像情報のみを用いて映像中の重要と思われる部分を抜き出し、ユーザの希望する時間内に映像を短縮再生する方法について研究を行っている。本稿では、この短縮再生法により作成した映像の評価実験について述べる。実験では映像の種類や短縮率を変えて被験者に見てもらい、出現する事象をどのくらい把握できるか調査した。
**キーワード**　マルチメディア処理、映像短縮再生，被験者実験

# Evaluation of New Video Skimming Method Using Audio and Video Information

Shigemi AOYAGI†, Ken'ichi KOURAI†, Koji SATO†, Toshihiro TAKADA†, Toshiharu SUGAWARA†, and Rikio ONAI††

† NTT Network Innovation Laboratories
3–9–11 Midori-Cho, Musashino, Tokyo, 180–8585 Japan
†† Department of Computer Science, University of Electro-Communications
1–5–1 Chofugaoka, Chofu-shi, Tokyo, 182-8585 Japan
E-mail: †{aoyagi,kourai,koji,takada,sugawara}@t.onlab.ntt.co.jp, ††onai@cs.uec.ac.jp

**Abstract**　There are now various kinds of video contents available via cable TV and satellite, so it is important to select important video contents from among the huge amounts available in a reasonable time. We have developed and implemented a new video skimming algorithm that can generate a shortened video within a user-specified time. In this paper, we describe the results of evaluation experiments. In the experiments, we investigate that the testee can grasp the events appeared in the skimmed video.
**Key words**　multimedia, video skimming, experiment by testee

## 1. はじめに

近年のインターネットのブロードバンド化による動画配信の広がりや、CATV や衛星を用いた TV 放送の多チャンネル化により、我々が種々多様な映像情報に接する機会が急速に増加しつつある。また、TV チューナを備えた PC やハードディスクレコーディングビデオなどの普及で、デジタル化された映像も身近なものとなってきており、視聴者の映像視聴環境が変わりつつある。このような環境下で、膨大な映像情報の中から自分に必要な情報を効率良く選び出していくことが今後ますます重要になっていくと思われる。

しかしながら映像情報の場合、テキスト情報における速読のように、同じ情報量を短時間で得ることは難しい。テキストの速読に近いものとして、映像の早送りが一般に使われる。画像情報は、注視すれば通常の数倍の速さで早送り再生してもある程度内容を把握することが可能であるが、人の耳は高速で再生された音の内容を把握できるようには出来ていない。特に音声の場合は通常の 2 倍や 3 倍といった速度で再生すると内容を把

握するのが困難である。したがって、市販のビデオデッキや動画再生プレイヤ等では、映像の早送りの際は音声を再生しないもの、音声の一部だけを途切れ途切れに再生するものが多い。一部のビデオデッキは、無音区間を飛ばし、音があるところは聞き取れる範囲で高速再生するといった早送り機能を備えているが、この場合は無音区間を詰めて再生するため、音と画像の同期が失われ、映像理解を難しくしてしまっている。今後、映像情報がますます増加していくことを考えると、映像情報をできるだけ内容を損なわずに短時間で見る方法を確立することが重要になる。

人間の耳の性能から音声の高速再生が困難である以上、限られた時間の中で映像情報を視聴するためには、映像を要約して必要な部分の映像だけを再生するという方法が考えられ、研究もさかんに行われている。映像要約で重要なのは、ユーザにとって欲しい情報が含まれているかということと、総再生時間である。ある映像を見たときに、どの部分が重要だと思うかは見るユーザに依存するので非常に難しい問題である。また、時間が限られたユーザにとって総再生時間が指定できるのは必要な機能であるが、再生時間を考慮した要約や短縮再生を行うシステムは非常に少ない。 総再生時間が指定可能なら、例えば、「先週 1 時間のドラマを録画したが、その続きがあと 30 分で放映が始まる。それが始まる前に先週分を見てしまいたい」など、時間を効率的に使うことができる。

本稿では、映像中の音情報と画像情報を用いて、ユーザが総再生時間指定可能な映像短縮法 [2] について説明し、さらにプロトタイプシステムを用いて作成した短縮映像を使った、被験者による評価実験の結果を報告する。この被験者実験では、単純なアンケート等による主観評価ではなく、映像中の事象の出現の有無をチェックしてもらう方法により、被験者の映像の理解度を調べている点に特徴がある。

以下、本編では使う誤解を招きそうな言葉の定義を明確にしておく。映像とは本稿がターゲットにしている音と画像が含まれる動画をさす。映像の大元は TV 放送でもインターネットで配信される動画像でもかまわないが、解析をするのでデジタル化されたものとして話を進める。映像は画像情報と音情報からなる。画像情報は静止画像の羅列である。音情報は BGM、効果音、人の声、ノイズ、といったものが合成されたものである。音声は音情報に含まれる人の声をさす。

第 2 章では関連研究を、第 3 章では本研究の映像短縮法の説明を、第 4 章では評価実験について述べ、第 5 章でまとめる。

## 2. 関 連 研 究

映像に関する研究は、画像の解析や検索、ブラウジング等、多岐にわたって行われている。映像の特徴抽出を行う手段として、映像全体の構造を認識するためのカット点の検出やフェイドイン、ワイプといった特殊効果を検出する方法が古くから検討されている [15] [16]。Informedia システム [6] [11] ではシーン検出などの他、画像情報中に出現する人の顔を認識してデータベース構築に用いている。画像情報だけでなく、音の情報を用いる研究も行われており、音を映像のインデクシングに用いる [10] では、音声パートの認識や音楽のスペクトル特徴を利用して、音楽、男性の声、女性の声を分別している。Saracenoら [3] は音情報を、無音区間・スピーチ・音楽・ノイズの 4 つに分類し映像のカット点検出の精度をあげる方法について述べている。また、Informedia システムのように、音声認識をデータベース構築のために用いているものもある。森山ら [13] は TV ドラマに対象を絞り、映像や音を要約に使う方法について検討している。映像はショット毎に分離され、音も効果音・BGM・セリフに分類されているという設定のもとで、それぞれを映像要約のための情報に使う方式について詳しく調べている。将来、MPEG-4 など情報が構造化されたデータが出現した場合、ドラマに対しては有効と思われる。しかし、これらのシステムでは、映像情報をシーンなど意味のある単位に分割することに主眼が置かれ、必要なシーンをつなげてユーザの希望する時間内で再生する、などといったことは考えられていない。

要約した映像の再生時間について注目したものに Li ら [5] のシステムがある。無音区間の削除と話速変換技術を用いていて、再生プレイヤは無音区間を削除し話速変換を使用した場合の総再生時間を表示する機能を持つ。しかし、話速変換による高速化は難しく、話者によっては 150%程度の高速化しただけで聞き取れなくなってしまう場合もある [1]。また、TV 放送の場合には無音となる区間はほとんどないため、この方法ではすべての映像に対しての大幅な高速化は望めない。

要約した映像情報をユーザに見せるインタフェースの研究もなされている。分割した映像をシーン単位でまとめておき、そのシーンを代表する静止画像を表示し、その静止画像をクリックすることでそのシーンの動画が見られる、というインタフェースを採用したものが多い [16]。シーンを代表する静止画像を画面に表示する際に、割り付けの大きさや場所を考えて表示する [7] という研究もある。また、静止画像でなく動画像を表示する [6] ものもある。これら代表フレームや動画を画面上に表示する方式は、一度見た映像に対して、後でもう一度確認するといった用途で用いる場合には有効に働くが、初めて見る映像の全体内容を把握する用途には余り役に立たない。また、シーンを代表する静止画像や動画像を見てしまうことにより、楽しみにしていたドラマの結末がわかってしまう、といった弊害が起きる可能性もある。Drucker ら [4] は Smart Skip というデジタル映像ブラウジングインタフェースを実装し、評価実験を行っている。

映像や音以外の情報を用いた研究もある。柿沼ら [9] は映像だけでなくシナリオがあることを前提に、ドラマ映像を対象にしたデータベースの構築を行っている。長尾ら [8] は、映像のトランスクリプトを XML 形式を用いて表し、そのトランスクリプトを要約し、その要約されたトランスクリプトに対応する映像を再生する、という方式で映像要約を実現している。トランスクリプトを作成するために音声認識を用い、さらに XML 形式を作成するためのツールも提供している。これらの方式では、映像以外の情報を用いるため、きめこまかな要約が可能であるが、シナリオの作成や、要約のための XML 形式のタグ埋め込みに手間がかかる点が問題となる。

## 3. 映像短縮法の詳細

本章では短縮再生アルゴリズムについて述べる。このアルゴリズムでは、映像中に含まれる音情報と画像情報のみを用いて重要そうな部分を識別している点と、ユーザが指定した時間以内に総再生時間が設定可能な点に特徴がある。

### 3.1 アルゴリズム

映像情報は、音情報と画像情報に分けられ、音情報の解析をベースに画像情報で補正をかけて再生する部分を決定する。音情報、画像情報から次の特徴量を計算して短縮に用いる。

（1） 音の動的特徴量

動画中の音データを一定周期のフレームごとにケプストラム分析を行い、その各ケプストラムの各次において求めた係数の 2 乗和である Δ ケプストラムの値は、スペクトルのゆるやかな動きに比例する。音声はこのゆるやかな動きが生じるので Δ ケプストラムの値は大きくなり、定常的な雑音などはスペクトル変化がないため、Δ ケプストラムの値は小さくなる。この Δ ケプストラムの値を調べることにより、無音区間や定常雑音の区間と音声区間を判別することが可能である [12]。ただし、楽器等で演奏された音楽も音声と同じくスペクトルの緩やかな変化がありΔ ケプストラムの値は大きくなるため、音声と音楽を区別することはできない。動的特徴量は 400ms を単位として計算を行うが、200ms でシフトして計算するため、結果は 200ms 単位で求められる。

（2） 音のパワー

重要なところでは人の話す声や効果音、BGM などが大きくなるという経験則に基づいて、200ms 単位で音のパワーを計算しておく。音の波形データの値を $S_i$ とすると、音のパワー P は $P = 10log_{10}(1 + \sum S_i^2)$ で計算できる。

（3） ZCR(Zero Crossing Rate)

ZCR は一定時間内に音の波形データが 0 を交差する回数である。音声部分では、この ZCR が高くなる傾向がある。ZCR の値 $Z_{cr}$ は以下で与えられる。

$$Z_{cr} = \Sigma neg(S_i \cdot S_{i+1})$$

ただし、

$$neg(x) = \begin{cases} 0, & (x \geq 0) \\ 1, & (x < 0) \end{cases}$$

（4） 画像からのカット点検出

画像情報を調べ、画像が大きく切り替わる場所であるカット点の検出を行う。カット点でない部分の画像は、変更がゆるやかであるということだから、カット点検出により似通った映像のまとまり (ショット) が抽出できる。要約システムでは、複数のショットの類似性などからショットより大きなまとまりであるシーンを出すものも多いが、本システムではカット点検出のみしか行っていない。

カット点検出のアルゴリズムとしては長坂ら [14] の分割 $\chi^2$ 検定法を用いた。分割 $\chi^2$ 検定法では、まず 1 枚の画像を 4x4 の 16 矩形領域に分割し、それぞれの矩形において 64 色種のヒストグラムを調べる。比較する 2 枚の画像を $f_1, f_2$ としたとき、矩形 $r$ の色 $i$ の色濃度を $H(f_1, r, i)$ で表すと、 各矩形における $\chi^2$ 検定の値は

$$\sum_{i=0}^{63} \frac{(H(f_1, r, i) - H(f_2, r, i))^2}{H(f_1, r, i)}$$

となる。分割 $\chi^2$ 検定では、$r = 0..15$ の全 16 個の値のうち小さい方の 8 個の和を評価値とする。

本稿のシステムは、録画された TV 放送や配信される (された) 映像などの再生において使用されることを想定しており、リアルタイム性を追求される場面は少ないと思われる。そこで、プロトタイプシステムでは、上の 4 つの特徴量はあらかじめ計算を行っておく。再生時には、ユーザからの総再生時間の指定や閾値の設定の変更に応じて再生する区間の絞り込みを行う。

流れは図 1 のようになる。

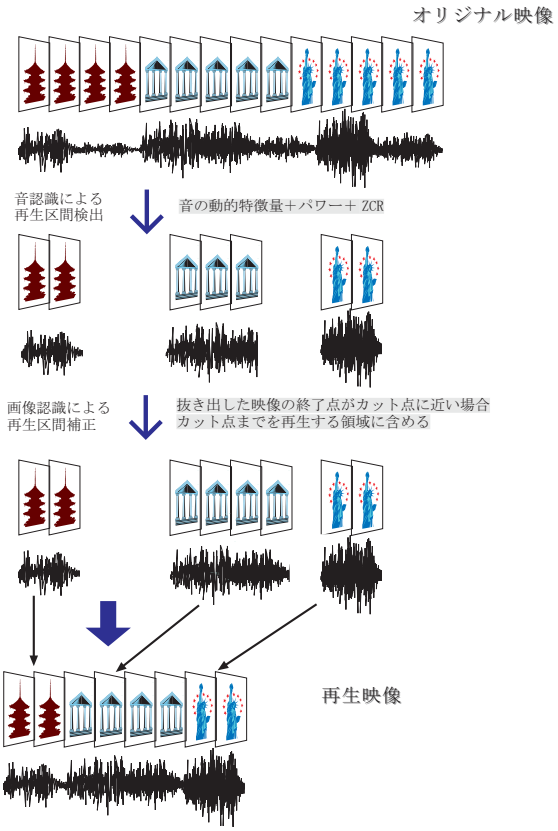


図 1 アルゴリズムの流れ

(1) 音の動的特徴量を用いて、音声らしき音声区間の抽出を行う。動的特徴量は 200ms 単位で結果がでるので、その 200ms の動的特徴量が閾値を超えた場合にその区間を音声だとする。ただし、ある 200ms だけが閾値を越えるがその前後は閾値を越えない単独の 200ms 区間は削除する。200ms しかない部分では音声の可能性は低く、もし音声であっても意味のある言葉を話している可能性が低いからである。

(2) (1) で求めた区間に対して、さらに音のパワーと ZCR に対して閾値をそれぞれ設定し、合致する範囲を絞り込む。

(3) (2) により、音データ上で区切れがある区間が抽出される。実際の映像では登場人物のセリフや解説などが終わって

から画像上でシーンが変わって次のシーンに入る場合が多い。その場合、音声が終わった直後で映像を切るのではなく、シーンが変わるまで再生したほうがユーザの映像への理解が得られやすい。そこで、音声が終わってから一定時間内にカット点がある場合、そのカット点まで再生する範囲を延長する。

動的特徴量、パワー、ZCR の各特徴量に対して閾値を設定し、それぞれの閾値を変更することで適合する区間が増減し、それによって総再生時間を変更することができる。閾値は各特徴量の平均や最大値といった値から映像に応じて計算しており、最大に短縮した場合にオリジナルの 50%以下になるように調節している。

### 3.2 プロトタイプの作成

Microsoft Windows 上で動作するプロトタイプを作成した。本プロトタイプは、AVI 動画ファイルを読み込み、各種計算を行った上で短縮再生を行う。3.1 で述べた前処理は、動画像ファイルの初回読み込み時に行い、別ファイルに記録している。2 回目以降は前処理をせずに別ファイルに記録されたデータを使用する。

前処理にかかる時間であるが、Pentium III(1GHz)、搭載メモリ 512MB の Win98 マシンを用い、22 分ほどの映像データ (映像: 320x240dot,IndeoVideo32 圧縮、音声:16bit,11KHz, モノラル) を処理するのに、12 分ほどかかった。音の特徴量計算だけなら 10 秒から 15 秒程度で終わるので、時間の大半はカット点検出にかかっていることになる。 現在は確実に動作することを優先にカット点検出のアルゴリズムを作成しており、高速化についてはまったく考慮していない。また、画面サイズ全部使う必要があるとも思われない。アルゴリズムを見直したり、解像度を落とすといった方法で前処理の時間を短縮することは可能である。

本プロトタイプのユーザインタフェースを図 2 に示す。ユーザはダイアログ中のスライドバーを使用して短縮再生時間を指定することができる。図 2 のダイアログ中、上のスライドバーを右に動かすと総再生時間を短くできる。また、下の 3 本のスライドバーを動かすことで、動的特徴量、パワー、ZCR それぞれに対する閾値の値を直接変更することができる。

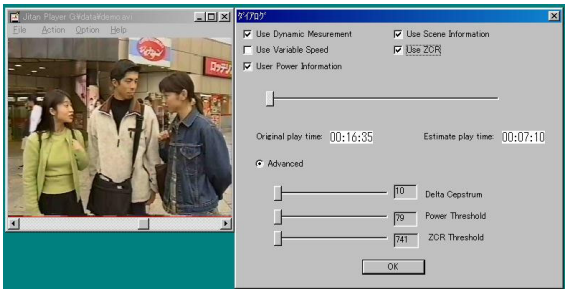


図 2　プロトタイプ実行画面

## 4. 評 価 実 験

プロトタイムシステムを用いて短縮映像を作成し、その映像を被験者に見てもらい内容がどの程度理解できるか実験を行った。

短縮映像の評価方法として、例えば、複数人の被験者にオリジナル映像と短縮映像を見比べてもらい、その差の感想等をアンケート記入により答えてもらう方法などが考えられる。

しかし、単純なアンケート記入法では、アンケートの設問に左右される部分が多いため、設問設定が難しく、また、フリー記述の感想にしてしまうと思ったことを書いてもらえない場合も考えられる。また、被験者が映像内容をどの程度理解できたか調べるために、見た映像の内容を記述してもらう方法も考えられる。しかし、映像内容の書き出しは一人の被験者に対する作業量が多すぎるし、映像が長時間になればなるほど作業の難易度も上がってしまう。また、書いてもらった内容記述から、被験者がどの程度理解してるかを判定することが難しいという問題がある。

そこで、本稿では被験者に映像内容の事象が書かれたリストを配り、映像を見てもらいながら映像中に出現した事象にチェックをつけてもらう、という方式を採用した。この方式では、チェックがついた事象の数により、被験者の映像理解度を比較することが可能である。

実験は事象リスト作成実験と短縮映像映像評価実験の 2 段階に分けて行った。

### 4.1 事象リスト作成実験

事象リストを実験作業者が作成してしまうと、実験側の意図が入ってしまう可能性があるため、事象の書き出し作業も被験者にやってもらった。

事象リスト作成のための実験手順は次の通りである。

（1） 被験者に短縮してないオリジナルの映像を見てもらい、映像中に現れた事象についてできるだけ細かく記述してもらった。ただし、後で事象の整合性をとるために、オリジナル映像上にタイムコードを表示させ、事象はその事象が起きた時のタイムコード＋事象の内容を記述してもらった。

（2） 実際に作業に入る前に、準備実験で練習してもらった。本実験で作業してもらう映像とはまったく別の 10 分程度の長さの練習映像を用意し、練習映像の最初の 5 分については事象のサンプルをあらかじめ作成しておいた。準備実験では、被験者に事象リストのサンプルを渡し、練習映像を 5 分ほど見せて、事象リストをどのように書けばよいか理解してもらった。その後、事象サンプル書いてない練習映像の後半 5 分について、実際に事象を書く練習をさせた。書き終わった後、事象がおおまかすぎる場合等、書き方に問題がある場合は指導を行った。

（3） 準備実験の終了後、すぐに本実験を行った。なお、本実験では、テープの早送りや巻き戻しにつては制限を設けず、不明な場合は何度も見てもらって事象リストを作成してもらった。

（4） 4 人〜5 人の被験者に別々に事象リストを作成してもらった後、その結果をあわせて一つの映像について一つの事象リストを作成した。この時、事象は最も細かくなるように、各被験者の事象リストの和集合の事象リストを作成した。

### 4.2 短縮映像評価実験

事象リスト作成実験で作成したリストを用い、次の手順で評価実験を行った。

（1） まず、事象リストにダミー事象を挿入した。ダミー事

表 1　実験に使用した映像の種類

| | ジャンル | オリジナルの長さ |
|---|---|---|
| 映像 A | アニメ | 21 分 56 秒 |
| 映像 B | ドラマ (和) | 44 分 5 秒 |
| 映像 C | ドキュメンタリ | 28 分 3 秒 |
| 映像 D | ドラマ (洋) | 36 分 48 秒 |

象とはその映像内では絶対に起こりえない、選択されるはずがないような事象である。挿入場所は、シーンが移り変わる直前で、そのシーンで登場していた人物があり得ない行動をした、という事象を加えた場合が最も多い。

（2） 短縮映像を作成。短縮映像は各映像について、オリジナルの再生時間を 100%としたとき、100%、80%、70%、60%、50%、40%の 6 種類の長さの映像を作成した。

（3） 事象リストを被験者に配布し、映像を見てもらい、出現したと思われる事象にチェックをつけてもらった。ただし、今回はダミー事象が含まれていることは被験者には一切説明しなかった。

（4） この実験は準備実験なしで行った。また、映像を見てもらう際、早送り・巻き戻しは使用禁止にした。早送り・巻き戻しを禁止した理由は、この短縮再生法は従来の早送りの替わりに用いられ、短時間で多くの内容を見ている人間が理解することを目標・想定しているからである。ただし、チェックをつける時間が必要なので、画像の一時停止だけは許可した。

（5） 実験中は、被験者の様子をビデオカメラで撮影した。これは、被験者が実験の手抜きをしないように心理的圧力を加えるために設置した。

（6） 被験者には、映像 A〜映像 D それぞれについて 6 種類の長さのうちの 1 本を選んで見てもらった。ただし、同じ種類の映像は一度しか見せないようにした。また、たとえば、最初に映像 B の 80%、次は映像 C の 50%･･･というように、短縮度もなるべく異なるように順番を指定して連続で見てもらった。

実験に使用した映像はアニメ、ドラマ、ドキュメンタリの 3 種類 4 本である (表 1 参照)。同じ映像は 2 度は見せないようにしたため、一人の被験者が見た映像の本数は最大 4 本である。映像 A〜D を見せる順番は、被験者によって変えるようにした。また、見せる映像の短縮率もなるべく変えるようにした。被験者は 18 才〜25 才程度の男女で、事象リスト作成実験に 21 人、短縮映像評価実験は 93 人に対して行った。ただし、事象リスト作成実験を行ったうちの 12 人は、短縮映像評価実験も行っている。ただし、その際、短縮映像評価実験は事象リスト作成実験とは別の日に行った。

各映像に対する総事象数とダミー事象数を表 2 に示す。ここでの総事象数は事象リスト作成時の事象数である。つまり、被験者に渡したリストには総事象数＋ダミー事象数の事象が書かれることになる。データとは、被験者が映像を見て対応する事象リストにチェックをつけた結果のリストのことであり、一人の被験者がチェックをつけた映像 1 本分の事象リスト 1 つにつき、1 と数える。正当データ数とは、映像 A と映像 B についてはダミー事象についたチェックが 1 つ以下のデータの数、映像 C と映像 D はダミー事象が 1 つだけなので、ダミー事象にチェックをつけなかったデータの数を意味する。誤答データ数とは映像 A と B は 2 つ以上、映像 C と D は 1 つダミー事象にチェックをつけたデータの数である。映像 A と B については事象数もダミー事象数も多く誤ってダミーにつけてしまった可能性も考慮して、ダミー事象に 1 つまでチェックをつけたものについては正当なデータとしている。

なお、表 2 において、正当データ数と誤答データ数の和は、その映像を見た被験者の総数になるが、その値は映像によって異なっている。その理由は、短縮映像評価実験では時間の許す限り多くの種類の映像を連続で被験者に見てもらったため、映像を見るのに要する時間は被験者によって異なるため、すべての被験者に全種類の映像を見ることができなかったからである。

実験結果を図 3 に示す。図 3 のグラフはそれぞれ 1 つの映像に対応している。グラフの横軸は被験者が見た短縮映像のオリジナルに対する長さ、縦軸は事象の出現率である。ここで、出現率とは事象に被験者がチェックをつけた事象数が全体の事象数の何%かを表す。

映像短縮評価実験では、あらかじめ配布された事象リストに現れたと思われる事象にチェックをつける方式をとった。短縮再生法では画像の早送りや話速変換といったものを一切使用していないので、図 3 は被験者が見て、短縮映像がどの程度の事象が短縮映像に残っているかを示していると思われる。図 3 のグラフの線の傾きが緩やかであればあるほど、本短縮の効果が大きいということになる。4 つのグラフを見ても、いずれも傾きにやばらつきがあるものの、オリジナルの 50%程度までは事象出現率は緩やかな減少を見せており、短縮映像に事象が残っていることを示している。

本実験により得られるのは、被験者が映像中の事象の出現をどのくらい理解できたかということであり、それは必ずしも被験者の映像内容の理解度を示すものではない。しかし、被験者が映像内容をどのくらい理解しているのかを客観的に調べるのは困難であるため、事象理解度を数値化できる本実験は、短縮映像の評価法としては有効だと考えている。

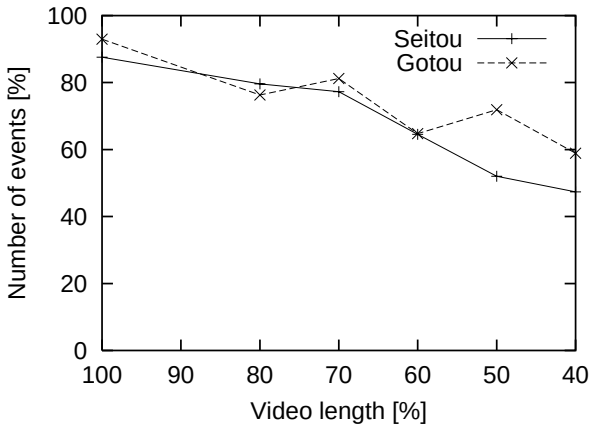


図 4　映像 A における正答と誤答の平均出現率の遷移

図 4 は映像 A において、正答データにおける事象の平均出現率と、ダミー事象にチェックがついた誤答データにおける事象

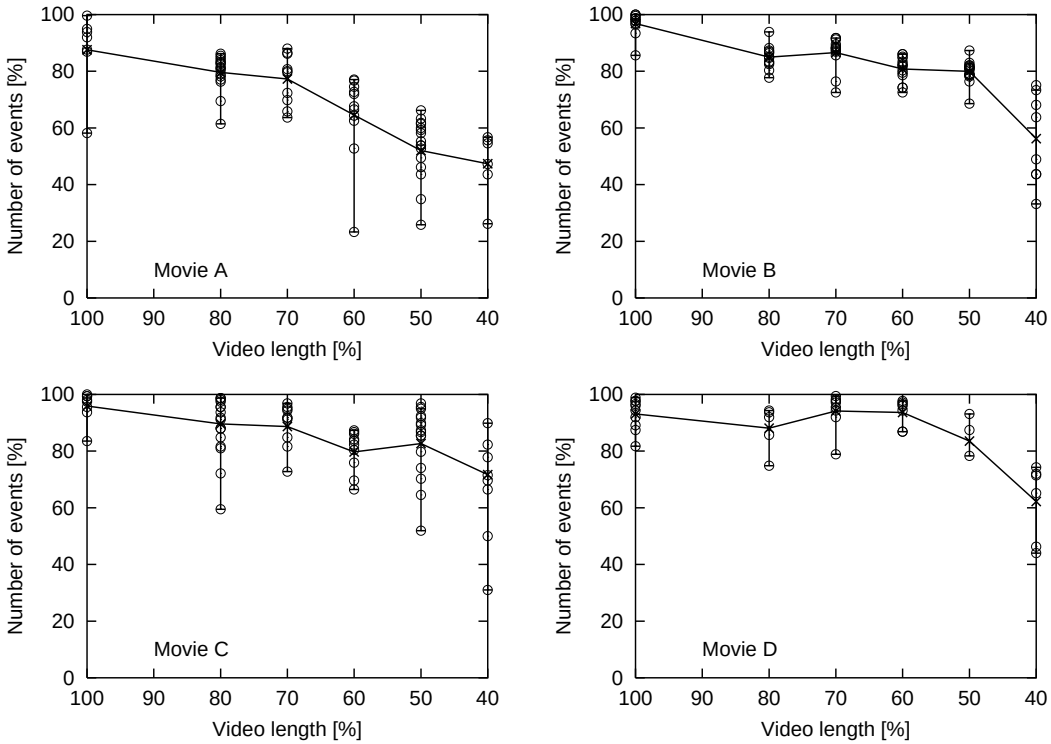


図 3　実 験 結 果

表 2　映像の種類と事象数、正答・誤答データ数

| | 総事象数 | ダミー事象数 | 正答データ数 | 誤答データ数 |
|---|---|---|---|---|
| 映像 A | 275 | 11 | 66 | 20 |
| 映像 B | 229 | 14 | 75 | 9 |
| 映像 C | 158 | 1 | 79 | 4 |
| 映像 D | 175 | 1 | 42 | 14 |

の平均出現率を表したものである。この図から、映像 A では一部逆転している部分もあるが、ほとんどの場合において、誤答データの方が平均出現率が高めに出ていることがわかる。そもそも誤答とは映像に出現しないダミー事象にチェックをつけたものであるから、ダミー事象以外にも過ちやその他の理由で余分にチェックされている可能性が高いと思われる。よって、ダミー事象挿入により正答データだけを選別することは、事象出現率を本来の出現率以上に上がってしまうのを防ぐ効果があると思われる。

### 4.3　慣れによる事象出現率の変化

本短縮法による映像を初めて見た場合と二度目に見た場合では、二度目の方が慣れて映像内容がよく理解できるという可能性も考えられる。そこで、本短縮法を最初に見た場合と二度目以降での事象出現率の平均値の差を比較した。表 3 は本短縮方式で映像 D を一度目に見た場合と二度目以降に見た場合の事象出現率の平均である。40%以外の値では予想通り二度目以降の値が一度目に比べて大きくなっている。しかし、現状ではサンプル数が少なすぎるため、より多くの被験者で実験する必要がある。

### 4.4　集中度テストによる疲労度の測定

実験の開始直前と終了直後に、集中度テストによる疲労度測定を行った。集中度テストは、1 行 12 文字、全 29 行のカタカナ文字のリスト中から、カタカナ 6 文字を選び出し斜線で消すという作業を 2 分間行うものである。また、実験の終了時にも同じ集中度テストを行った。ただし、終了時のテストはカタカナではなく平仮名文字にした。このテストでは、疲労にともない、総消去数が減少し、消去ミス数 (消去しなければならなかったのに消去しなかった) が増加すると考えられる。

表 3　映像 D における慣れによる事象出現率の変化 [%]

| | 100% | 60% | 50% | 40% |
|---|---|---|---|---|
| 一度目 | 88.57 | 86.86 | 78.29 | 73.15 |
| 二度目以降 | 96.07 | 96.34 | 90.30 | 56.72 |

被験者 59 人に対してこのテストを行ったところ、実験終了直後の方が実験開始直前に比べ、一人あたり平均で 2.6 文字分の消去総数の減少、2.0 文字分の消去ミスの増加が観察された。また、6 個以上のダミー事象にチェックをつけた 6 人の平均値を調べると、消去総数は 7.5 文字分の減少、消去ミスは 16.8 文字であり、平均よりもかなりの数の増加が観察された。したがって、ダミー事象にチェックをつけた一因として疲労による集中力の低下が考えられる。実際、ダミーに多くチェックをつけた被験者の集中度テストの結果を調べたところ、いずれも平均以上の集中度低下が見られた。しかし、最後に行ったアンケートにおいて本短縮法を「是非使いたい」「使ってみてもよい」と答えている場合もあるため、本短縮法がその被験者に合わなかったと特定することはできなかった。

## 5.　お わ り に

本稿では、映像情報中の画像情報と音情報を用い、総再生時

間が指定可能な映像短縮法のアルゴリズムついて述べ、本短縮法により作成した映像の評価実験について述べた。評価実験では、事象リストを作成し、それにチェックをつけさせる方式を用いたことにより、被験者の事象理解度をより詳しく調べることが可能であり、また、ダミー事象を挿入することで被験者の手抜きがあった結果の一部の排除も可能になった。

しかし、今回の実験では総勢 100 人を超える被験者で実験したにもかかわらず、有効なデータ数は予想を遥かに下回ってしまった。映像内容が無効なデータ数に影響を及ぼしている可能性もあるため、今後は、有効データ数が少なかった原因を調査し、より多くの種類の映像に対して実験を行い、本短縮法に適した映像の種類の調査を試みる予定である。

また、今回行った疲労度テストを短縮映像ではなくオリジナル映像のみを見せた場合にも同様に行い比較することで、短縮映像を見たことによる疲労度への影響がわかる可能性もある。これも今後の課題である。

**謝辞**

本研究について、ご支援下さった小柳恵一未来ねっと研究所ネットワークインテリジェンス研究部部長、被験者実験に関して議論していただいた NTT CS 基礎研究所の大野健彦研究員、ならびに NTT 未来ねっと研究所ネットワーク情報処理研究グループの各氏に感謝いたします。

**文　　献**

[1] 青柳、高田、佐藤、菅原：“音情報と画像情報を用いた動画高速閲覧のための一考察,” 情報処理学会研究報告, 2000-DPS-99, Vol.2000, No.88, September 2000.

[2] Shigemi Aoyagi, Toshihiro Takada, Koji Sato, Toshiharu Sugawara, and Rikio Onai: “Video Skimming Method for Flexible Play Time,” In Proceedings of the Sixth IASTED International Conference on Internet and Multimedia Systems and Applications, pp.330–335, 2002.

[3] Caterina Saraceno and Riccardo Leonardi: “Audio as a Support to Scene Change Detection and Characterization of Video Sequences,” Proceedings of IEEE ICASSP, MDSP3L.1, Vol.4, pp.2597–2600, 1997.

[4] Steven M. Drucker, Asta Glatzer, Steven De Mar, and Curtis Wong: “SmartSkip: Consumer level browsing and skipping of digital video content,” Proceedings of ACM CHI 2002, pp.219–225, April 2002.

[5] Francis C. Li, Anoop Gupta, Elizabeth Sanocki, Li-Wei He, and Yong Rui: “Browsing Digital Video,” Proceedings of ACM CHI 2000, pp.169–176, April 2000.

[6] Informedia, “http://www.informedia.cs.cmu.edu/”

[7] John Boreczky, Andreas Girgensohn, Gene Golovchinsky, and Shingo Uchihashi: “An interactive Comic Book Presentation for Exploring Video,” Proceedings of ACM CHI 2000, pp.185–192, April 2000.

[8] Katashi Nagao, Shingo Hosoya, and Kevin Squire:“Semantic Transcoding: Making the World Wide Web More Understandable and Usable with External Annotations,” WIT2000-S1-3, 2000.

[9] 柳沼、坂内: “DP マッチングを用いたドラマ映像・音声・シナリオ文書の対応付け手法の一提案,” 電子情報通信学会論文誌, D-II, Vol.J79–D–11, No.5, ppp.747–755, 1996.

[10] Kenichi Minami, Akihito Akutsu, Hiroshi Hamada, and Yoshinobu Tonomura: “Video Handling with Music and Speech Detection,” IEEE Multimedia Magazine, Vol.5, No.3, pp.17–25, 1998.

[11] Michael A. Smith and Takeo Kanade:“ Video Skimming and Characterization through the Combination of Image and Language Understanding Techniques,” Proceedings of Computer Vision and Pattern Recognition, pp.775-781, June 1997.

[12] 水野、高橋、嵯峨山: “スペクトルの動的および静的特徴量を用いた言語音声の検出,” 日本音響学会講演論文集 (秋), 3-2-1, 1995.

[13] 森山、坂内：“ドラマ映像のトラック構造に基づくダイジェスト生成,” 信学技報 PRMU2000-29, pp.43–50, 2000.

[14] 長坂、田中: “カラービデオ映像における自動索引付け法と物体探索法,” 情報処理学会論文誌, Vol.33, No.4, pp.543–550, April 1992.

[15] 田村、池田 (編): “知能情報メディア”, 総研出版,1995.

[16] Yong Rui, Sean X. Zhou, and Thomas S. Huang: “Efficient Access To Video Content in a Unified Framework,” IEEE International Conference on Multimedia Computing and Systems, Vol.2, pp.735–740, 1999.