映像短縮再生システムの教育映像への適用評価

青柳滋己[†] 佐藤孝治[††] 高田敏弘[†]
菅原俊治[†] 尾内理紀夫[†††]

インターネットの普及やブロードバンド化により，インターネットの利用が e-Learning 等の教育の分野にも広がりつつある．また，講義や講演等をビデオ録画し，後からいつでも見られるシステムが実用化され，実際に使われている．しかし，それらのシステムでの動画に対する操作は，再生・早送り・巻き戻し等の従来の VCR の機能や，スライドバーにより任意の位置から再生する等の機能しかない．今後，動画像の利用が増加することを考えると，動画像をより短時間で，しかも意味ある情報をなるべく欠落させずに見る機能が必要になる．我々は，映像中の音情報と画像情報を用いて，重要と思われるシーンを抜き出し映像を短縮する方法について研究を進めている．我々は作成したプロトタイプシステムを教育用映像に利用し，短縮化した画像を初期教育用，さらには復習等の目的に適用することを考えている．本稿では，教育用映像を本システムを用いて短縮した場合の内容の理解度に与える影響について，被験者実験を試みたので，その実験と解析結果について報告する．

Evaluation of Video Skimming Method to Educational Purpose Movies

Shigemi Aoyagi,[†] Koji Sato,[††] Toshihiro Takada,[†]
Toshiharu Sugawara[†] and Rikio Onai[†††]

Much video content is now available via cable TV and satellite. Educational programs through the Internet began popular. Fast-forwarding through items of video content is a natural way to check whether or not they are of interest. However, the sound is not synchronized with the images and the lack of comprehensible audio data means that we must concentrate on the images to understand the content. We have studied and developed the video skimming algorithm that uses only audio and image information of video data to extract the important part scenes. In this paper, we show the result of evaluation experiments that investigates the effectiveness of our video skimming algorithm that applies to educational movies.

1. はじめに

インターネットの普及やブロードバンド化により，e-Learning と呼ばれるインターネットを利用した各種教育が広がりつつある．e-Learning の特長は，いつでも好きなときに学習が可能であり，そこで使われる講義等の映像は何度も見ることができるという機能を備えている点にある．しかし，映像は基本的に通常速度で再生されるため，たとえば 20 分の映像を見るには 20 分必要となる．今後，ますます映像情報が膨大になることを考えると，映像視聴に要する時間が少しでも短くなるのが好ましい．

† NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories
†† NTT サイバースペース研究所
NTT Cyber Space Laboratories
††† 電気通信大学電気通信学部情報工学科
Department of Computer Science, University of Electro-Communications

しかし映像情報の場合，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だが，高速で再生された音の内容を把握することは難しい．特に音声の場合は通常の 2 倍以上の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では，映像の早送りの際は音声を再生しないか，音声の一部だけを途切れ途切れに再生している．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，飛ばした無音区間部分の画像は再生されるため音と画像の同期が失われ，映像理解を難しくしている．

従来の映像要約研究では，映像を構造化することに主眼がおかれ，動画に対する操作は，再生・早送り・巻

き戻しといった従来の VCR の機能や，スライドバーにより任意の位置から再生する機能に限られることが多い．たとえば，早送りの場合は音が再生されない場合が多く，再生される場合でも 1.2 倍速程度が限界でそれ以上速度をあげると特殊な訓練をしないと聞き取れない．また，スライドバーを使えば任意位置から再生可能であるが，初めて見る映像に対して用いてもかえって理解の妨げとなる可能性がある．初めて見る・初めて勉強する，あるいは理解していない内容の映像を見る場合には，従来の早送りやスライドバー等の機能では，重要な場面を飛ばすか，分かり難くしてしまうこともあり，逆に学習の妨げとなる可能性もある．

短縮映像を構造化してサムネイル表示する研究システムがあるが，初めて見る映像のサムネイルを見てもそれほど効果的とは思えない．短縮化のための短縮率の変更が可能なシステムもあるが，総再生時間等の指定ができないものが多い．そのようなシステムでは，例えば，今日の本放送が始まるあと 20 分の間に先週分の放送を見ておきたい，といった要求を実現するための短縮率の設定に手間がかかってしまう．

我々は，映像中の音情報と画像情報を用いて，重要と思われるシーンを抜き出し映像を指定した総再生時間内に短縮する方法について研究を行い[1]，さらにプロトタイプシステムを使って，ドラマ・アニメ映像に対し，短縮率と重要な内容の把握度の関係を実験により調査した[2]．この結果によれば，多くの場合 60–50%程度の短縮率までは，内容把握を大きく妨げる現象は見られなかったが，50%程度を境に認知率が大きく下がることが明らかになった．本稿のテーマである教育目的用には，上記の実験結果から 70%程度の短縮映像が最も適していると判断した．本稿では，我々のシステムを用いて 70%程度短縮した映像を教育用に利用した場合について実際に被験者実験を試みたので，その結果について報告する．

以下，2 章では関連研究について述べ，3 章では短縮アルゴリズムの概要を，4 章では教育を目的とした被験者実験の方法とその結果について述べる．

2. 関 連 研 究

教育分野においてはインターネットを使用した e-Learning のシステムが普及しはじめており，システム中で映像等のマルチメディアデータが多く用いられている．また，衛星通信やインターネットを用いた遠隔教育システムの研究や評価もさかんに行われている[3]．He ら[4] は，講義やプレゼンテーションの映像のデータベースから要約を作成する研究を行っている．音のピッチや無音部分の情報，スライド切替えの情報，ユーザのアクセス情報を用いて自動的に要約作成を行う．横井ら[5] は，黒板やホワイトボードを用いた映像の構造化がしにくい講義映像について，複数カメラを用い，ユーザが見たい映像にアクセスしやすくする研究を行っている．これらの研究は，講義映像へのアクセシビリティ向上を目的としており，特に復習用途向けとしての効果が期待できる．しかし，予習目的で映像を見る場合は，最初から全部見なければ理解できない場合が多いため，要約の効果を発揮できない．

一方，映像に関する研究は，画像の解析や検索，ブラウジング等，多岐にわたる．映像の特徴抽出の手段として，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[6)~8)]．しかし，これら映像要約の研究では，評価はほとんどされていない．Informedia システム[9),10)] のように，シーン検出等のほか，画像情報中に出現する人の顔を認識したり，音声認識の結果をデータベース構築に用いていたりするものもある．しかし Informedia では任意の短縮率を設定可能だが，総再生時間を考慮していない点，短縮映像の評価をしていない点で我々の研究とは異なる．

また，短縮映像の評価を行っていても，ある映像に対して，予想した映像の構造が意図どおり抽出できているかを調べ，達成度合いを求めているものが多い．たとえば森山ら[11] は，映像を効果音，BGM，セリフ等に分類し，それがうまく分離できているかを調べて評価を行っている．これと同様に，シーンチェンジについても，人間が自分で映像を見て正解を定め，自分で定めたシーンチェンジが検出された割合で評価としているシステムが多い．

Drucker ら[12] は SmartSkip 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．この評価実験では一般のアンケートによる調査のほか，Skip（10 秒または 30 秒の瞬時早送り），FF（早送り），SmartSkip の 3 つを使い映像中の天気予報を捜す等の課題を与え，時間や使用回数等の比較を行っている．この実験は映像の評価ではなく，インタフェースの評価実験であるが，被験者を用いて評価を行っている点で我々の研究に近い．しかし，映像中の特定場所を探す所要時間で評価を行っており，その時間が短いことは重要ではあるが，我々の行っている短縮映像そのものの評価とは異なる．

日高ら[13] は，作成した要約映像の評価を被験者実験を通して行っている数少ない例である．ただし，要約映像と対比する映像がランダム抽出により作成した

ものであり，題材が会話映像のみであること，被験者全員にプレビュを視聴させたあと実験を行っている点等，短縮映像の評価の観点からは十分とはいえない．

一方，本稿では，短縮映像は，次章に示すアルゴリズムにより生成されたものを利用する．特に，教育用の映像を対象に，通常映像と短縮映像とを見せたグループに分け，内容の把握度を試験するという客観的手法を用いているのが特徴である．

3. 映像短縮法の概要

本章では我々が用いた映像短縮法のアルゴリズムの概略を述べる．このアルゴリズムでは映像情報に含まれる音情報と画像情報のみを用いて短縮化を行う．詳細は文献 2) を参照されたい．

3.1 アルゴリズムの概略

映像情報は，音情報と画像情報に分け，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 4 つの特徴量を計算して短縮に用いる．なおこれらの特徴量はすでに他の短縮法でも使われている．

(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で求めた係数の重み付け 2 乗和を動的尺度と呼ぶ．この動的尺度の値は，スペクトルのゆるやかな動きに比例する．音声ではこのゆるやかな動きが生じるので動的尺度の値は大きくなり，定常的な雑音等は小さくなる．この動的尺度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できる[14]．ただし，楽器等で演奏された音楽も動的尺度の値は大きく，音声と音楽を区別できない．動的尺度は 400 ms を単位ごとに計算するが，200 ms でシフトして計算するため，結果は 200 ms 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 等が大きくなるという経験則に基づいて，200 ms 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 $P = 10\log_{10}(1 + \sum S_i^2)$ 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ので，カット点検出により似た映像のまとまり（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からショットより大きなまとまりであるシーンを抽出するものも多いが，本システムではカット点検出のみ行う．
長坂ら[15] の分割 χ^2 検定法をカット点検出アルゴリズムとして採用した．分割 χ^2 検定法は，まず 1 枚の画像を 4×4 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，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 χ^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 χ^2 検定では，$r = 0, \ldots, 15$ 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，録画された映像の TV 放送や配信時の再生で使用することを想定しており，リアルタイム性を追求されないと考える．プロトタイプシステムでは，これら 4 つの特徴量をあらかじめ計算しておく．再生時にはそのデータとユーザからの閾値の入力や総再生時間の入力により，再生箇所を選択する．その計算のアルゴリズムの流れを**図 1** に示す．

(1) 音の動的特徴量を用いて，音声と判断される区間の抽出を行う．動的特徴量は 200 ms 単位で計算されるが，ある 200 ms は条件に合うがその前後は条件に合わないような単独の区間は削除している．これは長さ 200 ms の部分が音声である可能性は低く，仮に音声であっても意味のある言葉を話してはいないからである．
この動的特徴量のみを用いて音声区間の抽出を行った場合，無音区間を削除したものと非常によく似た結果が得られる．具体的には，人が話すときの息継ぎの間や，意図的にあけた文と文の間の間，話の内容が変わるときの間，といっ

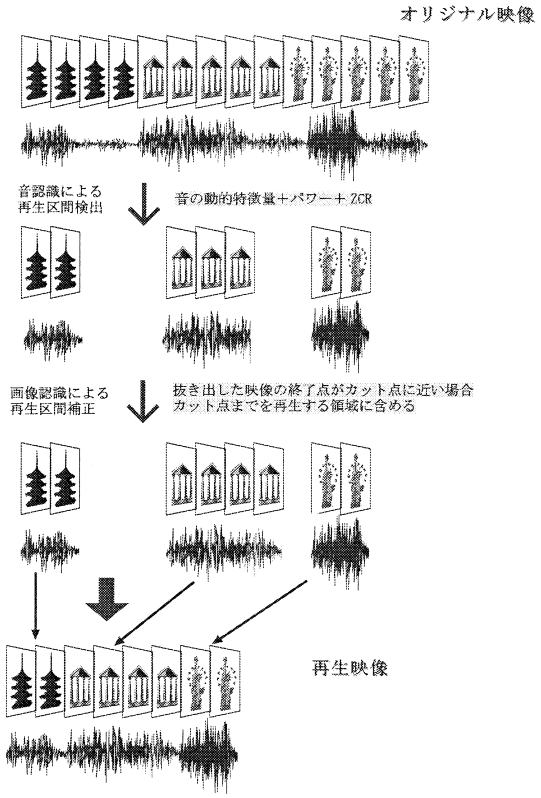

図 1 アルゴリズムの流れ
Fig. 1 Flow of the algorithm.

たものが削除されてしまう．しかし，200 ms 以下の細かい間が失われることはないため，文中の単語等が不明瞭になるといったことはない．ただし，閾値をあまりにもあげすぎると，話の初めの部分や終わりの部分が削除されてしまうため，何を言っているのかが分からなくなってしまう．

無音区間の検出には一般に音のパワーが用いられるが，パワーを使用した場合，エアコンの雑音等の定常ノイズのレベルが高い場合には音声区間の検出が難しいが，動的特長量を用いた場合はノイズレベルがかなり高くても音声区間の検出が可能であるという利点がある．

(2) (1) で求めた音声区間に対して，さらに音のパワーと ZCR を用いて，設定に合致する範囲を絞り込む．このとき，(1) で求めた動的特長量が一定時間続く区間全部を対象にする．つまりパワーと ZCR の値を区間内の 200 ms 単位で調べ，閾値を超える 200 ms 単位が 1 カ所でも存在すれば，その区間全部が再生されることになる．したがって，この (2) の操作によって，連続して話している文の語尾だけが消える，といったことは生じない．

一般に無音区間やノイズ等を取り除いてもオリジナルの 8 割程度の長さにしか短縮できない．そこで，無音区間や話の間を取り除いた音声区間のうち，パワーの低い部分やを音声ではないと思われる区間を削除することで，さらなる短縮化を行っている．

(1) で得られた結果と，(1)+(2) を行った結果を比較すると，(1)+(2) では，ボソボソと小さい声で話した言葉や文等が丸ごと削除される．ただし，パワーの小さい語尾だけがなくなって何を言っているのか分からなくなる，といったことにはならない．

(3) (1), (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等が終わってから画像上で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らずに，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．この理由から，音声が終わってから短い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．

この (3) では，カット点の場所に音声の切れ目＋少量時間の無音区間が追加される．通常，カット点はシーンチェンジの場面であることも多く，シーンが変わる際には音声も無音となり少し長めの間がとられる場合が多い．(1)+(2) の結果と (1)+(2)+(3) の結果を比べると，(3) まで行っているほうがシーンチェンジの音声の間も一部復活するため，(1)+(2) の結果より音声パートも聞きやすくなる．

なお，カット点検出の閾値は固定で 320×240 の画像で 10,000 に設定してある．シーンチェンジが発生している画面では一般に 100,000 以上の値になる場合が多く，逆にシーンチェンジがない場合にはせいぜい数百程度の値しかでないため，固定値で十分だからである．

各特徴量に対して閾値を導入し，各閾値を変更することで適合する区間が増減し，総再生時間を変更する．音の 3 つの特徴量については，それぞれ閾値を平均や最大値といった値から機械的にそれぞれ 10 個程度の特徴量値を算出し，これらの特徴量を組み合わせて再生時間を計算し，その中から希望再生時間に最も近い値を出すという方式を採用している．ただし，3 つのパラメータのうち動的特徴量の閾値をあげすぎると意

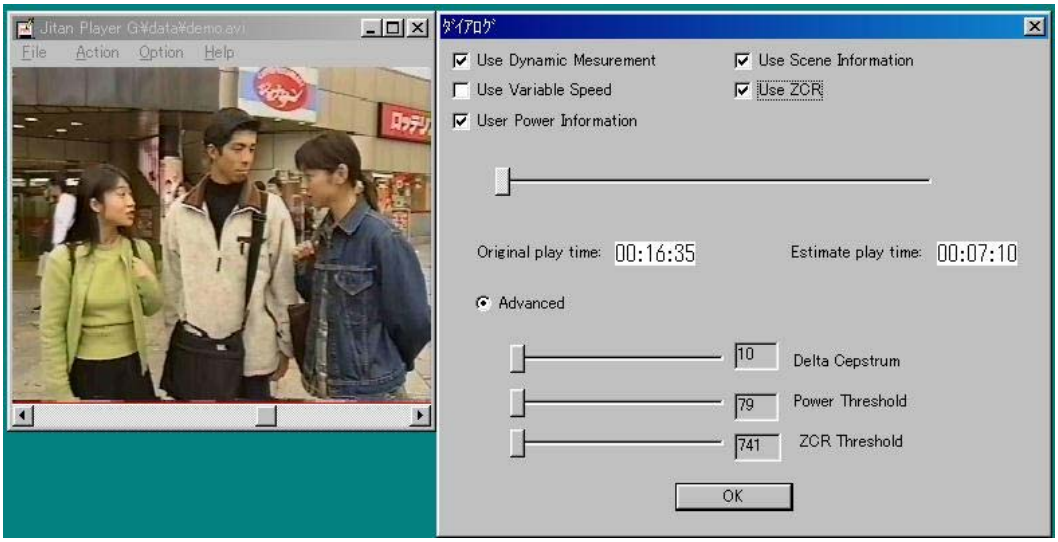

図 2 プロトタイプ実行画面
Fig. 2 A screenshot of the prototype.

味がとれなくなる可能性が高いため，動的特徴量の閾値があまり高くならないようにパワーの閾値と ZCR の閾値で調整している．この方式では，希望再生時間どおりになるとは限らず，1，2 分程度のずれが生じることもある．プロトタイプとして実験用映像作成のためには以上の機能で十分だが，将来改善する予定である．

3.2 プロトタイプの作成

プロトタイプは Microsoft Windows 上で作成した．本プロトタイプは，AVI 動画ファイルを読み込み，各種計算を行ったうえで短縮再生を行う．3.1 節で述べた前処理は，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．

前処理に要する時間は，Pentium III（1 GHz），搭載メモリ 512 MB の Win98 マシンを用い，約 22 分ほどの映像データ（映像：320×240 dot，IndeoVideo32 圧縮，音声：16 bit，11 KHz, モノラル）を処理するのに，12 分ほどである．音の特徴量計算のみの場合 10 秒から 15 秒程度で終了するので，大半はカット点検出処理にかかっていると推測される．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，高速化については考慮していない．解像度を落とす等の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．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，音の特徴量に対する閾値の値を直接変更することができる．

4. 評価実験

4.1 実験方法

我々の短縮再生法を教育映像に用いた場合の効果を調べるため，被験者実験を行った．被験者は 18〜25 歳の男女で，30 名以上入るほぼ同じ大きさ・形状の部屋を 2 つ用意した．各部屋には同じ大きさのスクリーン・プロジェクタ・ビデオ・机・イスが設置されており，同じ環境でビデオ視聴や被験者作業を行った．実験に用いた映像は，長さ 30 分程度の高校生用の地学と生物の教育用ビデオで，それぞれについて，オリジナル映像と本プロトタイプシステムにより 70%の時間に短縮した映像を記録したテープを用意した．ここで短縮時間を 70%としたのは，文献 2) より我々の映像短縮法による効果は 70%程度のときが最も効果的であるという実験結果による．また，映像の視聴効果を調べるため，テスト問題を地学・生物ともに 2 種類作成した．この際，2 種類の問題は同一の問題を含まず，なるべく同じレベルになるように留意したが，さらに差をなくすためにグループを 2 つに分け，2 種類のテストを前後入れ替えて利用した．テスト問題はいずれも 100 点満点で，問題はすべて複数候補からの選択式にした．70%の映像を作った際の閾値パラメータの値を**表 1** に示す．

表 1 短縮映像作成の閾値の値
Table 1 Threshold values for creating skimmed movies.

映像	動的特徴量			パワー			ZCR		
の種類	平均	最大	閾値	平均	最大	閾値	平均	最大	閾値
地学	0.72	4.74	0.36	99.2	116	103	382.1	1501	349
生物	0.71	3.86	0.37	95.6	111	100	367.5	1577	140

表 2 被験者数と映像の種類と事前・事後テストの種類
Table 2 Number of testers of each group.

グループ	人数	事前テスト	1 度目映像	事後テスト	事前テスト	2 度目映像	事後テスト
A	14(15)	地 1	地学 100%	地 2	生 2	生物 70%	生 1
B	14(13)	地 2	地学 100%	地 1	生 1	生物 70%	生 2
C	14	地 1	地学 70%	地 2	生 2	生物 100%	生 1
D	14	地 2	地学 70%	地 1	生 1	生物 100%	生 2
E	15	生 2	生物 100%	生 1	地 1	地学 70%	地 2
F	12	生 1	生物 100%	生 2	地 2	地学 70%	地 1
G	13	生 1	生物 70%	生 2	地 2	地学 100%	地 1
H	13	生 2	生物 70%	生 1	地 1	地学 100%	地 2

() は 2 回目の生物の実験の際の人数.
地 1 · · · 地学問題 1，地 2 · · · 地学問題 2，生 1 · · · 生物問題 1，生 2 · · · 生物問題 2.

閾値は動的特徴量の値を中心に設定している．ZCR はあまり精度が高くないため，あまり閾値をあげないようにし，動的特徴量とパワーを主に用いている．

70%のビデオ映像は，画面はパラパラとスキップし，音声も間がないところもあり，また，場合によっては一部音声が飛んだと思われる箇所もあった．しかし映像と音の同期がずれている場合に感じる違和感は感じない．また，集中して見聞きすれば内容が十分理解できるものと感じた．逆に 70%の映像を見てしまうと，オリジナルの映像はかなり冗長に感じられた．

実験の手順を次に示す．実験は 2 日に分け，同じ実験を繰り返した．

(1) 被験者をランダムに 2 つの部屋に入れる．
(2) 各部屋において，被験者をさらに 2 つのグループに分ける．片方の部屋の 2 つのグループをそれぞれ A と B，もう片方の部屋の 2 つのグループを C，D とする（2 日目のグループは E，F，G，H とする）．
(2) 最初に事前テストを行う．A と C，B と D はそれぞれ同じ問題だが，A と B，C と D には異なる問題を解かせた．
(3) A，B のグループの部屋では 100%の映像を，C，D のグループには 70%の映像をプロジェクタ投影して被験者に見せる．
(4) 映像を見終わった直後に事後テストを行う．事後テストの問題は事前テストと入れ替えて行う．たとえば，A のグループは B のグループで利用した事前テストを解かせる．
(5) 休憩を挟んだ後，生物の問題・映像を用いて (1)〜(4) をもう一度行う．ただし，最初に 70%を見たグループには 100%の映像を，他方のグループには 70%の映像を見せる．
(6) 2 日目も同様に (1)〜(5) を行う．ただし，2 日目は最初に生物の問題・映像を見せ，次に地学の映像を見せる．

なお，被験者は 1 回のみの参加とした．つまり，1 日目と 2 日目の両方に参加した被験者はいない．

表 2 に被験者の各グループの人数，見せた映像やテスト問題について示す．被験者人数は 1 日目の A〜D グループあわせて 56 名，2 日目の E〜H グループあわせて 53 名の全 109 名である．

4.2 実験結果

表 3 に，各グループにおけるテストの平均点を示す．なお，本表 3 における（事後–事前）の平均とは，事前テストの得点と事後テストの得点の差の平均である．地学と生物の映像内容を比較した場合，生物の方が映像中に出現する専門用語の数も多く，テスト問題も生物の方が難易度が高い．それは表 3 の平均点にも表れており，事後テストの得点も事前テストからそれほど増えていないのが分かる．

最初に，事前テストと事後テストの結果を調べ，映像を見たことによる効果があるかどうかを調べた．映像による学習効果があれば事後テストの平均値が事前テストに比べて上昇すると考えられる．そこで，グループ A〜H に対して行った地学の事前テストと事後テストについて，平均値に差がないと仮定して，対応のあるデータに対する有意水準 5%の t 検定を行った．同様に生物について平均値に差がないと仮定して，対

表 3 実験結果
Table 3 Result of experiments.

	一回目				二回目			
	事前平均	映像長さ	事後平均	(事後–事前)の平均	事前平均	映像長さ	事後平均	(事後–事前)の平均
A	33.9(地 1)	100%	70.4(地 2)	36.5	22.7(生 2)	70%	30.7 (生 1)	8.0
B	44.3(地 2)	100%	81.8(地 1)	37.5	31.5(生 1)	70%	41.2 (生 2)	9.6
C	33.9(地 1)	70%	74.3(地 2)	40.4	30.9(生 2)	100%	41.1(生 1)	11.1
D	43.9(地 2)	70%	77.1(地 1)	33.2	30.0(生 1)	100%	50.7(生 2)	17.5
E	27.3(生 2)	100%	34.7(生 1)	7.3	45.8(地 1)	70%	77.1(地 2)	31.3
F	29.6(生 1)	100%	44.2(生 2)	14.6	44.3(地 2)	70%	81.8(地 1)	37.5
G	32.7(生 1)	70%	48.5(生 2)	15.8	29.6(地 2)	100%	72.7(地 1)	43.1
H	24.6(生 2)	70%	28.8(生 1)	4.2	41.9(地 1)	100%	70.0(地 2)	28.1

表 4 t 検定の結果
Table 4 Result of t test.

	一回目			二回目		
	テスト	映像	$P(T <= t)$ 両側	テスト	映像	$P(T <= t)$ 両側
A	地 1 →地 2	地学 100%	4.40E-7	生 2 →生 1	生物 70%	0.0473
B	地 2 →地 1	地学 100%	1.91E-6	生 1 →生 2	生物 70%	0.0576
C	地 1 →地 2	地学 70%	4.48E-7	生 2 →生 1	生物 100%	0.0395
D	地 2 →地 1	地学 70%	1.13E-4	生 1 →生 2	生物 100%	5.12E-3
E	生 2 →生 1	生物 100%	5.45E-3	地 1 →地 2	地学 70%	5.24E-6
F	生 1 →生 2	生物 100%	0.0763	地 2 →地 1	地学 70%	7.76E-4
G	生 1 →生 2	生物 70%	2.38E-3	地 2 →地 1	地学 100%	1.60E-6
H	生 2 →生 1	生物 70%	0.0938	地 1 →地 2	地学 100%	1.49E-5

応のあるデータに対する有意水準 5%の t 検定を行った．t 検定の結果の $P(T <= t)$ 両側の値が有意水準の 5%より小さい場合には仮定が棄却され，事前テストと事後テストでは有意差があることになる．結果を**表 4** に示す．

表 4 から，地学についてはすべてにおいて有意差が見られたが，B，F，H の生物については仮定が棄却できず，有意差があることが示せなかった．

次に 100%と 70%の映像を用いた場合の同等性について調べる．ただし，比較の条件を同じにするため，地学と生物の映像を見た順番が同じで，かつ事前・事後テストの順番も同じペアについてのみ調べた．生物については，有意差が示せた C と A の組だけ調べた．一般の t 検定では有意差がないと仮定し，その仮定を棄却することで有意差があることを示せるが，「有意差がない」ことが「同等」であるとはいえない[16]．我々は「70%に短縮した映像を用いてもその効果は 100%に比べて遜色ない」あるいは「100%に比べてそれほど劣っていない」ということが示したいので，医学等で使われている臨床的（生物学的）同等性検証を用いる．これは，新薬の効果等の審査にも用いられており，「新薬は従来薬と同等の効果がある」場合や「新薬は従来薬に比べてメリットがあり，たとえ有効性が多少劣っていてもメリットがあるため認可したい」場合等に使われる．

我々は「70%の短縮映像が 100%のオリジナル映像に比べてそれほど劣っていない」ことを示したい．これは非劣性検証と呼ばれる．細かな検定方法については文献 16) にあるので，ここでは計算法だけ述べる．

医学的に意味のある最小の差 Δ を導入する．比較すべき 2 群の正規標本の母平均を μ_A，μ_B とすると，帰無仮説 H_0 と対立仮説 H_1 は

$$H_0 : \mu_A <= \mu_B - \Delta$$
$$H_1 : \mu_a > \mu_b - \Delta$$

となる．このとき検定統計量 T はサンプル数を n_A，n_B とすると，自由度 $\nu = n_A + n_B - 2$，サンプリング誤差 SE($\hat{\delta}$)，母平均の推定値 $\bar{X}_A$，$\bar{X}_B$ としたとき $T = (\bar{X}_A - (\bar{X}_B - \Delta))/SE(\hat{\delta})$ で与えられ，$T > t_\alpha(\nu)$ となれば，有意水準 α で非劣性と判定できる．

地学について，比較は事前・事後とも同じテストを解いた A と C，B と D，G と F，H と E についてそれぞれ上式にあてはめ計算を行った．結果を**表 5** に示す．Δ の値についてはいろいろ議論はあるが[16]，ここでは新薬検査等で一般的に使われている 0.1 を用いた．また，α の値も同じ理由で 0.05 とした．

この結果から，A と C，G と F については有意水準 5%で同等であることがいえる．つまり，100%の映像と 70%の映像の効果は同等で，差はテストの差分平

表 5 非劣性検定の結果
Table 5 Result of non-inferiority test.

グループ	人数	(事後–事前) 平均	標準偏差	SE($\hat{\delta}$)	T	$t_\alpha(\nu)$
A(地 100%)	14	36.4	14.7	2.22	3.44	$t_{26}(0.05) = 1.706$
C(地 70%)	14	40.4	16.4			
B(地 100%)	14	37.5	17.3	2.87	−0.19	$t_{26}(0.05) = 1.706$
D(地 70%)	14	33.2	22.8			
G(地 100%)	15	28.0	16.8	2.63	6.80	$t_{26}(0.05) = 1.706$
F(地 70%)	13	43.1	20.1			
H(地 100%)	12	31.3	11.7	3.24	−0.014	$t_{23}(0.05) = 1.714$
E(地 70%)	13	28.1	22.7			
C(生 100%)	14	11.1	18.1	2.22	−0.88	$t_{27}(0.05) = 1.703$
A(生 70%)	15	8.0	14.2			

均の 10%以内であることが示せた．

しかし B と D，H と E について同等であることは示せなかった．また，生物についても同じく同等であることは示せなかった．同等といえた A と C，G と F の組と B と D，H と E の組の差は，同等といえた方は事前テストに（地学 1），事後テストに（地学 2）を使っており，同等といえなかった組は事前テストに（地学 2），事後テストに（地学 1）を使った点であるが，このことがこの結果に影響しているのかどうかは現在のところはまだ不明であり，今後の課題である．

5. おわりに

本稿では，我々が開発した映像短縮アルゴリズムを教育映像に適用し，被験者実験を試みた結果について報告した．今回の実験では，用意した 2 種類の映像のうち，1 種類は教育映像の難易度が高すぎて教育効果が十分に現れなかった．また，もう 1 種類の内容が比較的容易な映像については，行った 4 つの実験のうち 2 つについては我々のシステムで作成したオリジナルの 70%の長さの短縮映像を用いた場合でも，オリジナル映像を見た場合とほぼ同等の教育効果が得られることが分かった．この結果から映像の短縮が教育に有用な場合がある可能性が明らかになった．今後は難易度の異なる映像や別の種類の教育映像でも実験を試みたい．

また，今回のテスト問題はすべて複数候補からの選択式を採用し，さらに被験者には分からない場合は空欄でよいと指示したにもかかわらず，勘に頼って欄を埋めたと思われる回答がいくつか見受けられた．このようなデータをいかに排除するか等，実験法についてもさらに改良を進める予定である．

謝辞 本研究について，貴重なご助言をくださった NTT サイバーソリューション研究所の大野健彦研究員，東京工業大学理工学研究科の光来健一氏に感謝いたします．

参考文献

1) 青柳滋己，高田敏弘，佐藤孝治，菅原俊治：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 (2000).
2) Aoyagi, S., Takada, T., Sato, K., Sugawara, T. and Onai, R.: Video Skimming Method for Flexible Play Time, *Proc. 6th IASTED International Conference on Internet and Multimedia Systems and Applications*, pp.330–335 (2002).
3) 村上正行，八木啓介，角所 考，美濃導彦：受講経験・日米受講習慣の影響に注目した遠隔講義システムの評価要因分析，電子情報通信学会論文誌 D-I，Vol.J84-D-I, pp.1421–1430 (2001).
4) He, L., Sanocki, E., Gupta, A. and Grudin, J.: Auto-Summarization of Audio-Video Presentations, *Proc. ACM Multimedia 1999*, pp.489–498 (1999).
5) 横井隆雄，遠山聖司，藤吉弘亘：イメージモザイキングによる講義のディジタルアーカイブと再生，インタラクション 2004 (2004).
6) 田村秀行，池田克夫（編）：知能情報メディア，総研出版 (1995).
7) Rui, Y., Zhou, S.X. and Huang, T.S.: Efficient Access To Video Content in a Unified Framework, *IEEE International Conference on Multimedia Computing and Systems*, pp.735–740 (1999).
8) Lienhart, R., Pfeiffer, S. and Effelsberg, W.: Video Abstracting, *Comm. ACM*, Vol.40, No.9, pp.56–62 (1997).
9) Informedia. http://www.informedia.cs.cmu.edu/
10) Smith, M. A. and Kanade, T.: Video Skimming and Characterization through the Combination of Image and Language Understanding Techniques, *Proc. Computer Vision and Pattern Recognition*, pp.775–781 (1997).
11) 森山 剛，坂内正夫：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).

12) Drucker, S.M., Glatzer, A., Mar, S.D. and Wong, C.: SmartSkip: Consumer level browsing and skipping of digital video content, *Proc. ACM CHI 2002*, pp.219–225 (2002).
13) 日高浩太，野口恵美，竹内順二，水野 理，中嶌信弥：音声の感性情報に着目したマルチメディアコンテンツ要約技術，インタラクション 2003, pp.17–24 (2003).
14) 水野 理，高橋 敏，嵯峨山茂樹：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).
15) 長坂晃朗，田中 譲：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 (1992).
16) 丹後俊郎：医学データ—デザインから統計モデルまで，共立出版 (2002).

(平成 16 年 7 月 8 日受付)
(平成 17 年 3 月 1 日採録)

青柳 滋己（正会員）

1965 年生．1988 年東京工業大学理学部情報科学科卒業．1990 年 3 月同大学大学院理工学研究科情報科学専攻修士課程修了．同年日本電信電話株式会社入社．現在，同社コミュニケーション科学基礎研究所に所属．複合メディア情報処理の研究に従事．電子情報通信学会，日本ソフトウェア科学会各会員．

佐藤 孝治（正会員）

1967 年生．1989 年慶應義塾大学理工学部数理科学科卒業．1991 年同大学大学院理工学研究科計算機科学専攻修士課程修了．同年日本電信電話株式会社入社．現在，同社サイバースペース研究所に所属．分散システム，マルチメディアシステム，オペレーティングシステム等の研究に従事．日本ソフトウェア科学会各会員．

高田 敏弘（正会員）

1962 年生．1986 年東京工業大学理学部情報科学科卒業．1988 年同大学大学院理工学研究科情報科学専攻修士課程修了．同年日本電信電話株式会社入社．基礎研究所，未来ねっと研究所等を経て，現在，コミュニケーション科学基礎研究所主任研究員．1994 年スタンフォード大学客員研究員．並列オブジェクト指向計算，分散システム，Web における情報発信/多言語化/マルチメディア処理，実空間コンピューティング等の研究に従事．ACM，日本ソフトウェア科学会各会員．

菅原 俊治（正会員）

1982 年早稲田大学大学院理工学研究科（数学専攻）修士課程修了．同年日本電信電話公社入社（武蔵野電気通信研究所基礎研究部）．以来，知識表現，学習，分散人工知能，マルチエージェントシステム，インターネット等の研究に従事．1992〜1993 年，マサチューセッツ大学アムハースト校客員研究員．現在，NTT コミュニケーション科学基礎研究所主幹研究員．博士（工学）．日本ソフトウエア科学会，電子情報通信学会，ISOC，IEEE，ACM 各会員．

尾内理紀夫（正会員）

1950 年生．1973 年東京大学理学部物理学科卒業．1975 年同大学理学系大学院物理学専攻修士課程修了．同年日本電信電話公社（現 NTT）に入社．1982〜1985 年に ICOT プロジェクトに参画，1997〜1998 年に RWC プロジェクトに参画．2000 年より電気通信大学情報工学科教授．著書に『コンピュータの仕組み』（朝倉書店），編書に『オブジェクト指向コンピューティング III』（近代科学社）『インタラクティブシステムとソフトウェア V』（近代科学社）等．マルチメディア情報処理，情報検索，セマンティックコンピューティング等に興味を持つ．1996 年情報処理学会プログラミングシンポジウム山内奨励賞受賞．工学博士（東京大学）．日本ソフトウェア科学会，人工知能学会，ACM 各会員．