

## 短縮再生と映像切替えを用いた二重映像視聴手法

高田 格<sup>†1</sup> 津村 弘 輔<sup>†1</sup>  
重野 寛<sup>†2</sup> 岡田 謙 一<sup>†2</sup>

本研究は、新しい映像視聴スタイルを提唱することを目的としている。具体的にはリアルタイムに送られてくる映像を一時的に蓄積し、映像の短縮再生を用いたコンテンツの圧縮による再生時間の短縮を行う。ここでいう短縮再生とは映像を早送りしながら再生することで本来の再生時間より短い時間で再生することと定義する。その短縮再生を用いて時間圧縮された2つのコンテンツに対し、自動で交互切替えを行うことにより、2つの映像を1つの映像時間分でリアルタイムに視聴する二重映像視聴スタイルを実現する。しかし、映像をスイッチングする度に発話途中で映像が切り替わってしまい、会話内容の把握が困難になる恐れがある。そこでその問題を解決するために、映像切替え時にオーバーラップによる映像再開支援手法を導入した。そして我々は提案手法に基づきプロトタイプシステムを実装し、評価実験を行った。その結果、提案手法を用いることでスムーズに映像に復帰することができ、内容把握にも支障を出さずに2つの映像を短時間で視聴することが可能であることを確認した。

### Twofold Video Watching Method with Shortened Video Playing and Video Switching

ITARU TAKATA,<sup>†1</sup> KOUSUKE TSUMURA,<sup>†1</sup> HIROSHI SHIGENO<sup>†2</sup>  
and KEN-ICHI OKADA<sup>†2</sup>

In this research, we propose a new video watching style. In particular, the real-time audio/video streams are stored temporarily and these streams are compressed by the shortened video playing. Here, we define the shortened video playing as the playing of the video for a shorter time than the original video using fast-forward playing. Two videos contents compressed by the shortened video playing are switched automatically at a constant interval. By doing so, we realize the twofold video watching style that users can watch two videos for the time of one original video. However, the grasp of conversational content becomes difficult because the video is switched at the middle of a conversation. To solve this problem, we introduced the video restart supporting method with overlap in video switching. We implemented the prototype system based on our proposal and carried out evaluation experiments. From these results, we confirmed that users can return smoothly to a video and watch two videos for a shorter time without losing the conversational content.

#### 1. はじめに

近年、放送の多チャンネル化や通信インフラの高度化等により、映像データや音声データのような時系列に連続するコンテンツが多数提供されている。これらのコンテンツにはスポーツ番組や会議中継等のリアルタイムな視聴が好ましいものも多い。そこで、コンテンツの増加に対応するためにプロの編集者が作るような

質の高い映像を自動的に生成する研究が多数行われている。例として、料理番組<sup>1)</sup>、プレゼンテーション<sup>2)</sup>、スポーツ<sup>3),4)</sup>、講義<sup>5)</sup>や会議<sup>6)</sup>における適した撮影領域や撮影対象を自動的に決定するカメラワークの研究がある。我々はこれまでにタイムシフトを用いた複数カメラのスイッチング手法<sup>7)</sup>の研究を進めてきた。しかしこれらの研究は、同時に提供される複数のコンテンツをリアルタイムに視聴したいという視聴者の要望を満たしていない。

これらの要望を満たすため、近年はディスプレイに複数のコンテンツを表示する技術が提供されている。しかしながらこの技術では、視聴者は同時並行的に複数のコンテンツを視聴しなければならず、内容を理解

†1 慶應義塾大学大学院理工学研究科

Graduate School of Science and Technology, Keio University

†2 慶應義塾大学理工学部

Faculty of Science and Technology, Keio University

することが非常に困難である。また、倍速再生技術を利用し、複数のコンテンツについて再生時間を短縮して視聴する方法も考えられる。しかしながらこの方法は、記憶媒体に記録済みで自在に読み出し可能なコンテンツに対してのみ適用できる方法であり、リアルタイムの視聴には適用することができなかった。

また、遠隔会議システムから会議の状況がリアルタイムに配信されるコンテンツや、コマーシャルが入らない番組等、コンテンツの内容によってはコンテンツ全体にわたって満遍なく視聴したいという要望がある。これらのコンテンツの全体を把握するために切り替えながら視聴してしまうと、コンテンツが途切れ途切れにしか視聴できないといった問題や、ユーザは切替え後のコンテンツを視聴している間に切替え前のコンテンツの内容について記憶が曖昧になってしまうといった問題が発生する。

そこで、本研究では上記問題点を解決し、圧縮された2つの映像コンテンツを交互に切り替えて再生することにより、リアルタイムに近い感覚で2つの映像を視聴可能な二重映像視聴手法を目指す。

以下、2章では提案手法について、3章では予備実験について、4章では映像切替え時の問題点について述べ、5章では本システムの実装について述べる。6章にて評価・考察を行い、7章を本論文のまとめとする。

## 2. 短縮再生と映像切替えを用いた二重映像視聴手法

本研究では、内容把握にも支障を出さずに2つの映像を短時間で視聴する新しい映像視聴スタイルを目指し、変速再生と映像切替えを用いた二重映像視聴手法を提案する。

### 2.1 映像視聴スタイル

映像視聴スタイルとして、1つの映像を視聴後にもう1つの映像を視聴するスタイルがある。これを本論文ではシリアル型視聴と定義する。このシリアル型視聴では、映像の早回しを行わないため、内容理解は容易であるが2つの映像の時間分拘束されてしまうため、効率良く映像を視聴することができない。仮に早回しで視聴することにより効率を上げようとしても、2つの映像すべてを録画した後でしか早回しが行えず、リアルタイムに視聴することができないという問題点がある。

別の映像視聴スタイルとして、2つの映像を同時に視聴するものがある。これを本論文ではマルチ型視聴と定義する。このマルチ型視聴では、2つのリアルタイムのコンテンツを視聴する場合にも対応できるとい

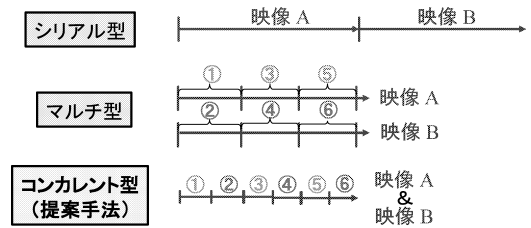


図1 映像視聴のスタイル

Fig.1 Style of video watching.

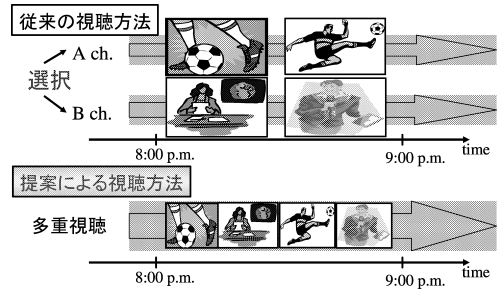


図2 二重テレビ視聴のイメージ

Fig.2 Image of twofold TV watching.

う利点がある。しかしこの視聴スタイルでは、2つのコンテンツの音声がお互いの音声を打ち消し合い、内容把握に支障が生じることが我々の研究により分かっている<sup>8)</sup>。

そこで我々は、映像の早送りによる短縮再生を行い、かつ、2つの映像を自動で交互に切り替えて視聴を行うことにより1つの映像分の時間で2つの映像視聴ができ、リアルタイムのコンテンツに対応可能となる映像視聴スタイルを提唱する。これをコンカレント型視聴と定義する。

シリアル型視聴、マルチ型視聴、コンカレント型視聴の概念図を図1に示す。このとき、図1のコンカレント型に示されている1と2の映像部分、すなわち映像Aを見始める時間から次の映像Aの部分を見始めるまでの時間を1周期と定義し、本論文で表記される1周期とはすべてこの時間のことと定義する。

コンカレント型視聴は2つのリアルタイムのコンテンツの視聴も可能にするため、テレビ番組の視聴にも対応できる。図2に具体的なイメージを示す。チャンネルAではサッカー番組の中継が行われており、チャンネルBではニュース番組の中継が行われている。従来の映像視聴方法では、視聴者自らチャンネルを変えることで2つの映像を視聴していた。そのため、すべての映像を視聴することは不可能であり、飛び飛びの映像しか視聴することができなかった。しかし、本提案により両映像を短く交互に短縮再生することで映像すべ



図3 映像密度圧縮のイメージ

Fig.3 Image of time compressed video.

てを視聴することができるようになると考えられる。

## 2.2 提案手法

本提案手法の実現を考えたとき、元々蓄積された映像コンテンツに適用するなら問題はない。しかし、本提案手法においてはリアルタイムの映像コンテンツに対応させることを目標としている。そこで、まず2つのリアルタイムのコンテンツを1周期分蓄積する。このとき映像は何も表示されない。そして蓄積が終わると、その蓄積されたコンテンツに対して短縮再生で映像を視聴する。視聴している際に、実際のリアルタイムの映像コンテンツは蓄積され続け、またその蓄積された映像を視聴することとなる。このように最初に映像を蓄積することで、つねに短縮再生で映像を視聴することが可能となる。

次に短縮再生による映像密度の圧縮について考える。このとき、映像密度を以下の式で定義する。

$$\text{映像密度} = \frac{\text{実際の映像時間}}{\text{再生時間}}$$

映像は一般的に通常速度で再生されるため、映像時間分の再生時間が必要となる。上記の式に従うと、映像視聴時間が短くなる方が映像密度が上がり、好ましい<sup>9)</sup>。そこで短縮再生をすることで再生時間の値が小さくなり、映像全体の密度が高まる。その結果、図3に示すように映像が圧縮され、余剰時間が創出される。そしてその余剰時間にもう1つの映像を視聴することが可能となる。

また、何も処理を行わずに再生速度を上げると音声の音程が高くなり、聞き取りが困難になる。そこで、TSM (Time Scale Modification) と呼ばれる音声補正技術<sup>10)</sup>を施すことで、再生速度を変化させることによって音声の聴講に支障をきたさないようにする。

ここで、1つの映像のみを視聴していると、もう1つの映像は蓄積され続けてしまい、リアルタイム性が損なわれてしまう。そこで映像の自動交互切替えを行い、二重映像視聴の実現を図る。映像の自動交互切替えとは、ある一定時間映像を視聴するともう1つの映像に自動的に切り替わるというものである。この手法を導入することで、1つの映像を短縮再生で視聴している間にもう1つの映像が蓄積され、そして次にその蓄積されていた映像を短縮再生で視聴することが可能

となり、この繰返しによりリアルタイム性を保持したまま2つのコンテンツを視聴することができる。

本提案手法を実現するために、短縮再生を行う際の適切な映像の再生速度、また適切な映像切替え間隔を調べる必要がある。そこで我々は予備実験を行った。次章にその詳細を述べる。

## 3. 予備実験

短縮再生した映像を用いて様々な実験を行うことで、短縮再生技術が人の聴覚情報処理能力に与える影響と、内容理解を損なわない適切な再生速度を明らかにする。初めに、様々な再生速度に変化させた映像を用いて最も適した再生速度を求め、次に、その求めた再生速度の映像を用いて内容理解を損なわないかを調査する。さらに、2つの映像を切り替える最も適した間隔を求めるために実験を行う。以下にその詳細を示す。

### 3.1 再生速度の主観評価実験

本実験は内容理解を損なわない適切な再生速度がどの程度であるかを主観評価によって測り、同時に人間の聴覚情報処理能力の限界を調査することを目的とする。

我々は本実験を行う前に、再生速度の適正值は大まかにどの範囲にあるかを調べることを目的とした予備実験を少人数で行った。その結果、2.0倍速での視聴は可能そうだが、3.0倍速を超えると不可能であるということを確認した。そのことから本実験では2.0~3.0倍速の間に適正值があると考え、再生速度を2.0倍速、2.2倍速、2.4倍速、2.6倍速、2.8倍速、3.0倍速と6種類の再生速度で16名の被験者に映像コンテンツを視聴してもらうことで適正值を測ることとした。

本実験では、16名の被験者のうち8名には最初に2.0倍速の映像を視聴し、そこから速度を上げた映像を順に視聴してもらった。また、残りの8名には最初に3.0倍速の映像を視聴し、そこから速度を下げた映像を順に視聴してもらった。なお本実験において、研究室の新人配属に関してディスカッションを行っている会議映像を、内容は異なるが話の難易度は同程度のものを6種類用意した。そうすることで、被験者には各再生速度で内容の異なる映像コンテンツを視聴してもらった。

評価実験用のコンテンツとして会議映像を用いた理由は、会議映像は話の流れを把握する必要があり、また、短時間の映像でもキーワードを複数含ませたものを用意することができ、評価実験に適しているためである。また、会議映像は音声情報が多く含まれるため、再生速度の上昇にともない視聴が非常に困難になりや

表 1 適正速度の評価

Table 1 Evaluation of appropriateness in video playing speed.

評価速度	2.0 倍	2.2 倍	2.4 倍	2.6 倍	2.8 倍	3.0 倍
得点	2.3	3.4	4.1	3.1	1.6	0.4

すい．そこで，視聴の難易度が高いと思われる会議映像コンテンツにおいて，その内容を理解できれば多くのコンテンツが視聴可能であると考えた．以降の評価実験で会議映像を用いた理由も同様である．

評価方法として，実験被験者に最も再生速度が速く，内容理解ができた再生速度から順位付けをしてもらうこととした．具体的には，内容理解がしっかりできて再生速度が最も速いものに最高点数である 5 点とした．そして，再生速度が速くても内容理解ができない，また内容理解ができて再生速度が遅いと思うものは主観評価で 4 点～0 点とした．そして各再生速度において平均値をとったものを評価とする．

表 1 に結果を示す．このことから，2.4 倍速が最も評価が高いことが分かる．また，この結果を基に Dunnett の方法で多重検定を行った．コントロール群を最も高い評価となった 2.4 倍速とし，他を対照群とした．また，コントロール群の平均値を  $u_c$ ，対照群の平均値を  $u_i$  とし，対立仮説を  $u_c \neq u_i$  とした．検定を行ったところ，有意水準  $\alpha = 0.05$  に対し 2.0 倍速，2.8 倍速，3.0 倍速のときは  $P < 0.05$  と帰無仮説が棄却され，有意差があることが分かった．逆に，2.2 倍速と 2.6 倍速のときは  $P > 0.05$  と帰無仮説が保留される結果になった．このことから 2.2 倍速～2.6 倍速の間に適正值があることが考えられる．

また同時に，それぞれの再生速度で映像を視聴してもらった際に，話の流れを把握することができたかどうかを質問することで人間の聴覚能力の限界値を調べた．評価項目は，話の流れを理解できたかという質問に「はい」と答えた人数の割合である．その結果を表 2 に示す．このことから，再生速度が上がるにつれて話の内容を理解できた被験者が減っていていることが分かる．

表 1 の結果から，2.4 倍速が最も高いポイントを得ていることが分かる．しかし，表 2 の結果を見ると 2.4 倍速では話の内容を理解できている人が 81% なのに対し，2.2 倍速では 94% と 13% 値が上昇した．このことから，ほとんどの人がコンテンツを問題なく視聴できるようにするために，本研究では再生速度を 2.2 倍速とし，これを適正值として設定する．検定の結果からも 2.2 倍速を設定しても妥当であることが分かっている．

表 2 聴覚能力の限界速度

Table 2 Limit speed of auditory ability.

再生速度	2.0 倍	2.2 倍	2.4 倍	2.6 倍	2.8 倍	3.0 倍
割合 (%)	100	94	81	63	25	0

表 3 短縮再生の適正評価

Table 3 Evaluation of usability of fast-forward playing.

再生速度	安楽死	旅行計画
1 倍速	65.1%	82.5%
2.2 倍速	63%	79.1%

### 3.2 1 倍速と短縮再生における理解度比較実験

本実験では，前節で求めた再生速度で映像を視聴した際に，1 倍速で映像を視聴した場合と比較して理解度にどれほど差が出るのかを調査することを目的とし，短縮再生の適正評価を行う．

本実験は短縮再生速度として 2.2 倍速と 1 倍速再生との理解度比較を行い，それにより再生された映像をどの程度理解することができたかを測る実験となっている．ここでいう理解度とは，会議中に出てくるキーワードをどの程度聞き取ることができたかと定義する．本実験においても会議映像を映像コンテンツとして使用した．本実験では約 2 分の会議映像コンテンツを 2 種類用意した．その会議映像の議題はそれぞれ「安楽死」と「旅行計画」に関してである．本実験では問題正解率を評価項目として設定した．具体的な問題としては安楽死の何が問題となっていたか，旅行の行き先はどこになったか，といった具体的な語句を問う問題となっている．以上の実験を 16 名の被験者に行ってもらった．このとき，実験結果を映像の内容や問題の内容によらない結果にするために，被験者 16 名を 8 名ずつの 2 グループに分け，それぞれのグループで映像の再生速度と会議映像の組合せを変えて実験を行った．具体的には，1 つ目のグループが 1 倍速で「安楽死」に関する議題の会議映像を，2.2 倍速で「旅行計画」に関する議題の会議映像を視聴した．そしてもう 1 つのグループが視聴する会議映像はそれとは逆にした．実験結果を表 3 に示す．

まず両速度の正解率を見たところ 2.2 倍速，1 倍速ともに正解率にほとんど変化がないことが分かる．また 2.2 倍速において，旅行計画に対して安楽死の会議内容のほうが正解率が下がっているが，1 倍速においても同様の結果が見られる．このことから正解率は会議の再生速度に依存するのではなく，会議内容に依存することが分かる．よって短縮再生による理解度に対する影響はないものと考えられ，短縮再生を映像視聴に適用しても問題はない．

### 3.3 映像切替え間隔適正実験

本実験では、内容理解を損なわないスムーズな2つの映像の交互視聴が可能となる適切な切替え間隔を測ることを目的とし、映像切替え間隔の適正実験を行う。

このとき映像コンテンツの組合せについて考えると、コンテンツの種類は多数あり、さらにそれらのうち2つを同時に視聴するとなると組合せは相当数存在することとなる。それらすべてにおいて適切な切替え間隔を求めることは困難であるため、3.1節でも述べたように今回は最も視聴の難易度が高いと考えられる会議映像コンテンツを2つ用いて実験を行い、その結果を適正值として本研究で用いることとする。もし本提案手法に適用するコンテンツが会議映像以外に限定される場合、そのコンテンツに応じた映像によって本節の実験を行うことでより適切な切替え間隔を求めることが可能である。

実験は、5秒、10秒、20秒、40秒、60秒、60秒以上と6種類の切替え間隔を設定し、それぞれの映像切替え間隔で2つの映像コンテンツを交互視聴してもらう。ただしこのとき再生する映像は3.1節の実験により得られた再生速度である2.2倍速で再生する。また実験には、「旅行計画」に関して議論している会議映像と「新人教育」という議題に関して議論している会議映像の2つを映像コンテンツとして用いた。この2つの会議映像を上記切替え間隔で交互に視聴し、3.1節と同様の評価方法で順位付けしてもらい、それぞれの映像切替え間隔において平均点を算出した。その結果を表4に示す。

表4より、40秒間隔が最も評価が高くなっている。また、この結果を基にDunnettの方法で多重検定を行った。コントロール群を最も高い評価となった40秒とし、他を対照群とした。コントロール群の平均値を $u_c$ 、対照群の平均値を $u_i$ とし、対立仮説を $u_c \neq u_i$ とした。そして検定を行った結果、有意水準 $\alpha = 0.05$ に対し5秒、10秒、20秒、60秒以上のときは $P < 0.05$ と帰無仮説が棄却され、有意差があることが確認された。また、60秒のときは $P > 0.05$ と帰無仮説が保留される結果になり、40秒から60秒の間に適正值があると考えられる。しかし今回は離散的な値を用いて実験を行ったため、結果の値の大きい40秒を本提案における切替え間隔として設定する。ただし、映像コ

表4 映像切替え間隔の評価  
Table 4 Evaluation of video switching interval.

切替え間隔	5秒	10秒	20秒	40秒	60秒	60秒以上
得点	0	1	2.6	4.7	4.1	2.5

ンテンツは2.2倍速で再生されているため、実際のコンテンツは2.2倍速×40秒=88秒分、つまり約1分半程度の内容を持つ映像を視聴し終えた後に次の映像に切り替えることとなる。

### 4. 映像切替え時の問題点

映像を交互に切り替える際に、映像切替え間隔分の時間が経過後に映像を自動で切り替えてしまうと、会話の途中で映像が切れてしまうため、再びその直後の映像から開始されると内容に関する理解が困難になってしまうという問題点があげられる。具体的には、図4のように映像Aと映像Bを交互視聴している際に、会話内容を考慮することなく40秒という時間だけで映像を切り替えてしまうと、再び映像A、映像Bに戻ったときに、発話が途中で途切れてしまっているため内容理解が困難となり、スムーズに映像に復帰することが不可能となる。

#### 4.1 映像切替え時のオーバーラップによる映像再開支援

上記問題点を解決するために、映像切替え時に $\Delta t$ 秒間という微小時間の映像を重複させて再開させることで、映像の内容を容易に思い出すことができ、スムーズな映像復帰が可能となる。具体的なイメージを図5に示す。図5の絶対時間を実世界の時間、相対時間を映像コンテンツの持つ時間と定義する。

まず最初に映像Aを $(40 + \Delta t)$ 秒間視聴する。その視聴を終えた後に今度は映像Bを $(40 + \Delta t)$ 秒間視聴する。再び映像Aを視聴する際に今度は $\Delta t$ 秒前の状態、つまり40秒経過した地点から映像を再開することにする。また映像Bを再び視聴する際にも同様の処理を行う。

この映像重複による再開支援手法は、テレビ放送においてコマーシャルの後にコマーシャルに入る少し前の映像から放送する等、現状でも有用性が認められ、使用されている。また、我々は映像重複を行ったものを行わないものとの内容理解度が変化するかを調査する予備実験を少人数で行った。その結果、映像重複による再開支援を行うことで明らかに内容理解度が上昇したことを確認した。よって、この映像重複による再

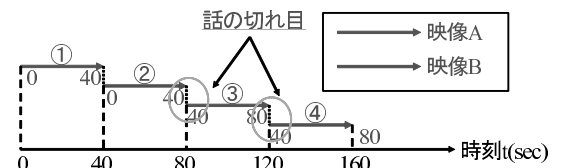


図4 映像切替え時の問題点  
Fig. 4 Issues in video switching.

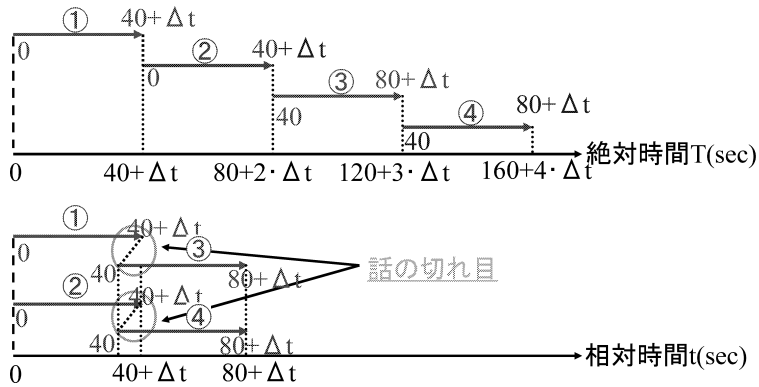


図 5 映像重複による内容理解の再開支援  
Fig. 5 Restarting support of content comprehension by video overlapping.

開支援手法を導入することで、発話が途中で途切れていてもその少し前から映像が再開されるためその発話の理解が容易になり、かつ、話の流れを思い出すことができ、スムーズな映像復帰が可能となる。

この再開支援手法を我々の提案に導入するにあたり、どのくらいの時間分の映像を重複させるのが適切かということは非常に興味深い問題ではあるが、コンテンツの種類や映像の内容によって重複させるべき時間は変化し、その映像を2つ組み合わせて視聴となると非常に多数のコンテンツの組合せに関して最適な重複時間を求めなければならない。そこで、今回我々は映像重複時間が長い方が理解度が向上するであろうという仮説を立てて話を単純化し、再生速度等の条件からできる限り長い重複時間を適用できるように算出によって重複時間を求めることとした。次節に重複時間  $\Delta t$  の算出方法を記す。

4.2 重複時間の算出

ここでは2つの映像を交互に視聴しても1つの映像を1倍速で視聴したときと同じ時間で視聴可能にすることを目的として、 $\Delta t$  の算出方法について述べる。映像の重複時間を  $\Delta t$  (sec)、短縮再生速度を  $k$ 、映像時間を  $T$  (sec) (簡単にするために両映像とも同じ時間の映像とし計算する)、映像切替え間隔は 40 (sec) と設定する。

ここで映像切替え回数を図 6 に示すように相対時間の時系列で考えると、1セットにつき2回の映像重複が発生することになる。その考えに基づき総重複時間は以下の式で表すことができる。

$$\text{総重複時間} = \text{切替え回数} \times 2\Delta t$$

ここで映像時間  $T$  を  $k$  倍速で短縮再生すると、1つの映像を視聴し終えるのに要する時間は  $\frac{T}{k}$  (sec) となる。そこで1つの映像に対する切替え回数は  $(\frac{T}{40 \cdot k} - 1)$  回発生することになり、総重複時間は

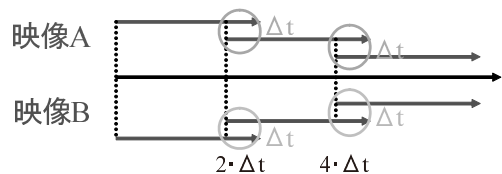


図 6 総重複時間の算出  
Fig. 6 Calculation of total overlap time.

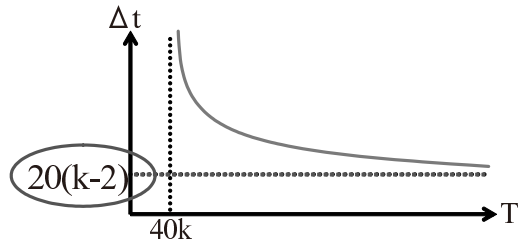


図 7 重複時間と映像時間の関係  
Fig. 7 Relation between overlap time and time of video.

$2\Delta t \times (\frac{T}{40 \cdot k} - 1)$  (sec) 発生することになる。以上のことより、1つの映像を1倍速で聞いたときと同じ時間で2つの映像を視聴し終えるためには、以下の式を満たす必要がある。

$$2\Delta t \times (\frac{T}{40 \cdot k} - 1) + \frac{2T}{k} \leq \frac{T}{1}$$

これを式変形すると、以下のようになる。

$$\Delta t \leq \frac{20(k-2)}{1 - \frac{40 \cdot k}{T}}$$

よって  $\Delta t$  の式より、映像重複時間と映像時間のグラフを求めると図 7 のようになる。

本研究では  $k = 2.2$  (映像を 2.2 倍速再生で行う) と設定した。そのため、 $40k = 88$  となる。ここで映像時間  $T$  は 88 (sec) に比べて圧倒的に大きなオーダであるため、 $T$  を無限に発散させた状態の  $\Delta t$  を算出しても間

題ない。その結果算出された  $\Delta t$  の値は  $20(k-2)$  (sec) となり、 $k = 2.2$  を代入すると  $\Delta t \leq 4$  (sec) となる。この条件を基に映像重複時間  $\Delta t$  を設定すれば、1つの映像コンテンツを1倍速で視聴したときと同じ時間で2つのコンテンツを見終えることが可能となる。

### 5. 実装

提案手法に基づいて、短縮再生された2つの映像コンテンツを自動切替えを行いながら視聴可能なプロトタイプシステムを構築した。以下にその詳細を述べる。

#### 5.1 映像出力アルゴリズム

本提案によるシステム構成図を図8に、以下にそれぞれの役割を示す。

まず2つの映像A, Bを一時的に(1)短期映像蓄積装置に蓄える。交互に映像をスイッチングするまでの時間分蓄積し終えると、その蓄積された映像は(2)可変圧縮装置へと入力される。その後、時間圧縮された映像・音声は(3)圧縮映像蓄積部へと遷移し、一時的に蓄積される。そしてそれらの蓄積映像に対して(4)映像選択部を通すことにより、出力されるべき映像が選択される。その後、実際に(5)映像出力部を通して適切な映像切替え間隔によって各々の映像・音声が(6)出力される。また、各映像の1巡目の際にはオーバラップは働かないが、2巡目以降では圧縮映像蓄積部で蓄えられた映像に対して(7)オーバラップ処理部が働くことになるので、映像の2巡目以降は(5)から(3)へ重複分の映像を送るというループが行われる。これにより2巡目以降は、切替え時の映像オーバラップを施した映像・音声交互に出力されることになる。さらに、(8)ユーザ操作部では映像の切替え間隔時間や倍速再生速度の決定等の変更が行える。そこで変更が行われれば、変更された値が(9)制御部を通じてシステムに組み込まれることになる。

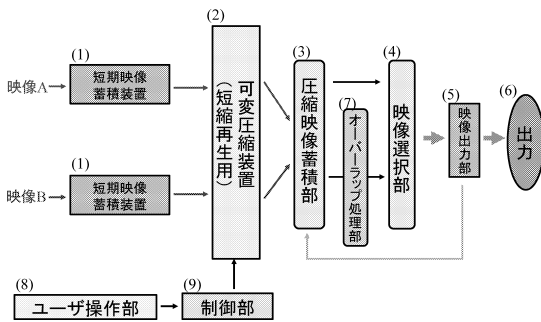


図8 システム構成図  
Fig. 8 System configuration diagram.

以上は働かないが、2巡目以降では圧縮映像蓄積部で蓄えられた映像に対して(7)オーバラップ処理部が働くことになるので、映像の2巡目以降は(5)から(3)へ重複分の映像を送るというループが行われる。これにより2巡目以降は、切替え時の映像オーバラップを施した映像・音声交互に出力されることになる。さらに、(8)ユーザ操作部では映像の切替え間隔時間や倍速再生速度の決定等の変更が行える。そこで変更が行われれば、変更された値が(9)制御部を通じてシステムに組み込まれることになる。

以上のシステム構成により入力映像に対して処理が施され、映像が出力される。映像A, Bに対する入出力関係を図9に示す。

まず、映像Aと映像Bを入力映像とした1周期分の映像を蓄積し終えるとそれぞれの蓄積映像に対して圧縮をかけ、その圧縮映像を出力し始める。ここで1周期とは、切替えが起こるまで1つの映像を見ている時間と定義する。

ここで2つの映像  $A_n, B_n$  を圧縮した映像を、それぞれ  $A'_n, B'_n$  と定義する。また、 $A_n, B_n$  の重複させる映像部分を  $\Delta A_n, \Delta B_n$  と定義し、さらに  $\Delta A_n, \Delta B_n$  を圧縮した映像を  $\Delta A'_n, \Delta B'_n$  とそれぞれ定義する。同時に映像Aに関して、1周期の定義は1回目の映像切替え間隔のみ  $A'_1 + \Delta A'_1$  とし、2回目以降からは  $\Delta A'_1 + A'_2 + \Delta A'_2, \Delta A'_2 + A'_3 + \Delta A'_3$  と続いていき、 $\Delta A'_n + A'_{n+1} + \Delta A'_{n+1}$  ( $n \geq 2$ ) という一般式で定義する。映像Bについても同様の定義を行う。

また、映像切替え間隔を40秒、再生速度2.2倍と設定すると、重複時間  $\Delta t$  は4秒と算出され、1周期は44秒となる。圧縮映像における  $A'_1, B'_1$  はそれぞれ40秒となり、 $\Delta A'_1, \Delta B'_1$  はそれぞれ4秒と設定される。そして各圧縮映像  $A'_1 + \Delta A'_1, B'_1 + \Delta B'_1$  を視聴し終えた後に再び映像Aを視聴することになる。ここで、再開時における映像Aは続きの  $A'_2$  部分が

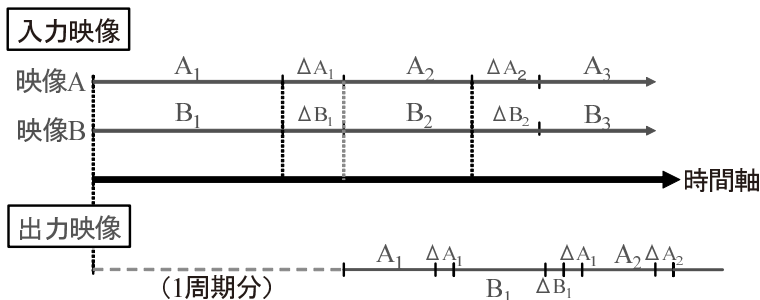


図9 映像の入出力関係  
Fig. 9 Relation between video input and output.



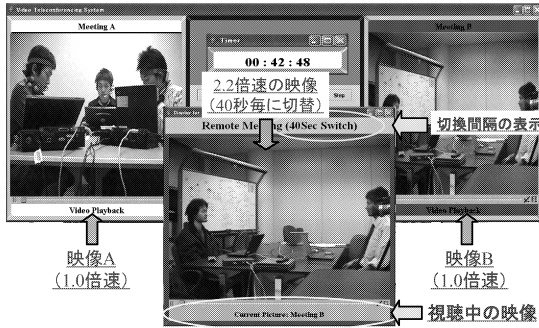


図 10 実装画面

Fig. 10 Example of implementation window.

ら開始されるのではなく、重複部分を持たせて再開させるため、映像は  $\Delta A'_1$  の部分から再開されることになる。

ここで、1 周期を 44 秒 (映像切替え間隔 40 秒 + 重複時間 4 秒) と設定しているため、 $\Delta A'_1 + A'_2 + \Delta A'_2 = 44$  秒となり、これより  $A'_2 = 36$  秒となる。よってこれ以降の  $A'_n, B'_n$  はすべて 36 秒として計算される。以上のように出力映像が生成される。

5.2 実装画面

以上の要素を取り入れ、プロトタイプシステムを実装した。実装画面を図 10 に示す。本プロトタイプシステムにおいて、奥の 2 画面に 1.0 倍速で再生された 2 つの映像コンテンツが表示される。そして中央にその 2 つの映像コンテンツから交互に映像が選択され、2.2 倍速の映像が表示される。このとき音声は、2.2 倍速で再生されている映像のもののみ出力される。

6. 評価・考察

本提案による新しい映像視聴スタイルの有用性を調査することを目的とし、評価実験を行った。以下にその詳細を述べる。

6.1 評価手法

本手法の有用性を検証するために、図 1 で示したシリアル型視聴、マルチ型視聴、そして提案手法であるコンカレント型視聴により 2 つの映像コンテンツを視聴してもらう。被験者は全部で 15 名となっており、それらの被験者を 5 名 3 グループに分け、上記 3 通りの視聴法で実際に映像コンテンツを視聴してもらった。本実験では問題正解率を評価項目として設定した。このとき、「未来の携帯電話」と「卒業旅行計画」に関して議論している映像コンテンツを、会議参加者を変えることで内容を変え、それを 3 セット用意した。各会議映像コンテンツの参加者はそれぞれ 4 名で、各議題につき 5 分間の内容となっている。このように、内

表 5 会議内容の理解度評価

Table 5 Evaluation of level of comprehending the contents of conference.

会議スタイル	シリアル型	マルチ型	コンカレント型
正解率	75%	45%	66%

容の異なる映像コンテンツを 3 セット用意し、グループごとに視聴方法と映像コンテンツの組合せを変えて実験を行うことで、動画内容・問題内容によらない結果を出した。問題は、A さんは新しい機能として何が欲しいか、1 日目の昼御飯は何を食べることになったか、といった具体的な語句を問う問題で構成される。次節に実験結果を示す。

6.2 実験結果と考察

表 5 に、シリアル型視聴、マルチ型視聴、本提案手法であるコンカレント型視聴の結果をそれぞれ示す。

今回の実験結果において、視聴方法による問題の正解の仕方に差が見られた。具体的には、シリアル型視聴においては語句の聞き逃し等はほとんどなかったため、75%と高い正解率を得ている。しかし、シリアル型視聴の問題点として、他の 2 つの映像視聴スタイルに比べて倍の時間が必要となるため、映像密度の低下ということがあげられる。またこのことが要因となり、前半部の会議映像の最初の方の内容を忘れてしまい、問題に答えられない人が多くいた。その結果、前半の映像の方が後半の映像より正答率が下がるという傾向が確認された。

マルチ型視聴は同時に 2 つの映像を視聴するという特性から、2 つの映像コンテンツの音声がお互いの音声を打ち消し合ってしまう、正解率 45%と他の 2 つの視聴方法に比べ低い値となっている。このマルチ型視聴においては個人の能力差による正答率への影響が大きく、最も正答率の高い人と最も正答率の低い人の差が非常に大きかったことが確認された。

提案手法であるコンカレント型視聴においては、シリアル型視聴で多く見られた前半部分の内容を覚えておらず問題に解答できないという傾向は大幅に改善された。また、マルチ型で見られた 2 つの映像の音声が打ち消し合うという問題も解消され、個人の能力差による正答率の違いも改善された。しかし、映像の再生速度を速くするという特性から、音声の補正はかけたものの、どうしても映像における発話者の元々の声が高い人の発話は聞き取りにくく、その人が発言した語句を聞き逃してしまい、正答率が下がるという傾向が見られた。このことが、提案手法であるコンカレント型は正解率 66%とシリアル型に比べて正解率が 9%低下したが、マルチ型に比べると正解率が 21%上昇し、



シリアル型の正解率に大幅に近づけているという結果につながったと考えられる。今後の課題として、ある程度元々の声が高い人に関しては、聞き取りやすいと感じるまで音声を低くする等の工夫が必要であると考えられる。しかし、これらのことを考慮しても今回我々の提案手法であるコンカレント型視聴は高い正答率、マルチ型視聴と同様に2つの映像を1つの映像分の時間で視聴することができており、リアルタイムに対応可能であるという点から新しい映像視聴スタイルとして有用性があり、本提案手法の実現は可能であるといえる。

## 7. おわりに

近年、放送の多チャンネル化や通信インフラの高度化等にともない、映像データや音声データのような時系列に連続するコンテンツが多数提供されている。これらのコンテンツにはスポーツ番組や会議中継等のリアルタイムな視聴が好ましいものも多い。そこで本研究では、映像の早送りによる短縮再生と映像自動交互切替えを用いて視聴を行うことにより、1つの映像を視聴するのに要する時間で2つの映像をほぼリアルタイムに視聴可能な新しい視聴スタイルを提案した。その新しい視聴スタイルを実現するために予備実験を行い、適切な映像再生速度、映像切替え間隔を調査し、導入した。また、映像切替え時の発話が途中で途切れてしまうために起こる、内容理解が困難になってしまうという問題、映像が切り替わってもその映像の流れを思い出すことが困難であるという問題を解決するために、映像切替え時のオーバーラップによる映像再開支援手法を導入した。さらに我々はプロトタイプシステムを実装し、提案手法の有用性を調査するための評価実験を行った。その結果、本提案によるコンカレント型視聴では正解率66%とマルチ型に比べると正解率が21%上昇し、シリアル型の正解率に大幅に近づけ、かつ、マルチ型と同様にリアルタイムのコンテンツに適用できることが分かった。このことから本提案によるコンカレント型視聴は映像の内容理解を損なわずに映像密度を向上させているという点で、新しい映像視聴スタイルとして有用性があり、本提案手法の実現は可能であることを確認した。

謝辞 本研究の一部はグローバルCOEプログラム「アクセス空間支援基盤技術の高度国際連携」により行われました。

## 参考文献

- 1) Pinhanez, C.S. and Bobick, A.F.: Approximate world models: Incorporating qualitative and linguistic information into vision systems, *Proc. AAAI'96*, pp.1116-1123 (Aug. 1996).
- 2) 尾関基行, 伊藤雅嗣, 中村裕一, 大田友一: 複合コミュニティ空間における注目の共有—人物動作理解による物体への注釈付け, *VRSJ 第6回大会論文集* (Sep. 2001).
- 3) 井口泰典, 土居元紀, 真鍋佳嗣, 千原国宏: スポーツ映像放送のための実時間映像解析によるマルチカメラの自動制御と自動スイッチング, *映像学誌*, Vol.56, No.2, pp.271-279 (Feb. 2002).
- 4) 松本圭介, 須藤 智, 斎藤英雄, 小沢慎治: サッカー放送における視点選択のための多視点画像の統合によるボール追跡, *電学論*, Vol.121-C, No.10, pp.1530-1539 (2001).
- 5) 村上昌史, 大西正輝, 福永邦雄: 状況理解と映像評価を考慮した講義の知的自動撮影, *情報処理学会研究報告*, CVIM-125-5 (Jan. 2001).
- 6) 井上智雄, 岡田謙一, 松下 温: テレビ番組のカメラワークの知識に基づいたtv会議システム, *情報処理学会論文誌*, Vol.37, No.11, pp.2095-2104 (1996).
- 7) 加藤淳也, 住谷哲夫, 井上亮文, 重野 寛, 岡田謙一: タイムシフトを用いた会議中継カメラのスイッチング手法, *情報処理学会論文誌*, Vol.47, No.3 (2006).
- 8) 安西 悠, 江木啓訓, 西川真由佳, 湯澤秀人, 松永義文, 岡田謙一: 遠隔会議への同時多重参加を目的とした理解度評価, *ヒューマンインタフェース学会論文誌*, Vol.8, No.2, pp.61-68 (2006).
- 9) 青柳滋己, 佐藤孝治, 高田敏弘, 菅原俊治, 尾内理: 紀夫映像短縮再生システムの教育映像への適用評価, *情報処理学会論文誌*, Vol.46, No.5, pp.1297-1305 (2005).
- 10) Wong, J.W.C., Au, O.C. and Wong, P.H.W.: Fast Time Scale Modification Using Envelope-Matching Technique (EM-TSM), *Proc. 1998 IEEE International Symposium*, Vol.5, pp.550-553 (June 1998).

(平成19年4月13日受付)

(平成19年10月2日採録)

高田 格 (学生会員)

2006年慶應義塾大学理工学部情報工学科卒業。現在、同大学院前期博士課程在学中。会議支援、多重ワーク支援の研究に従事。





津村 弘輔 (学生会員)

2005年慶應義塾大学理工学部情報工学科卒業。2007年同大学院前期博士課程修了。現在、IBM ビジネスコンサルティングサービス株式会社に勤務。在学中、会議支援、多重ワーク支援の研究に従事。



重野 寛 (正会員)

1990年慶應義塾大学理工学部計測工学科卒業。1997年同大学院理工学研究科博士課程修了。1998年同大学理工学部情報工学科助手(有期)。現在、同大学理工学部情報工学科准教授。博士(工学)。計算機ネットワーク・プロトコル、モバイル・コンピューティング、ネットワーク・セキュリティ、マルチメディア・アプリケーション等の研究に従事。著書『～ネットワーク・ユーザのための～無線LAN技術講座』(ソフト・リサーチ・センター)、『コンピュータネットワーク』(オーム社)等。電子情報通信学会、IEEE、ACM 各会員。



岡田 謙一 (フェロー)

慶應義塾大学理工学部情報工学科教授、工学博士。専門は、CSCW、グループウェア。情報処理学会誌編集主査、論文誌編集主査、GW研究会主査等を歴任。現在、情報処理学会 MBL 研究会運営委員、BCC 研究グループ主査、日本 VR 学会理事、CS 研究会委員長。情報処理学会論文賞(1996, 2001年)、情報処理学会 40 周年記念論文賞、日本 VR 学会サイバースペース研究賞、IEEE SAINT'04 最優秀論文賞を受賞。情報処理学会フェロー、IEEE、ACM、電子情報通信学会、人工知能学会各会員。