

## 多重ワーク時における複数カメラによる 遠隔会議映像の自動スイッチング手法

津村 弘輔<sup>†</sup> 住谷 哲夫<sup>†</sup> 高田 格<sup>‡</sup> 重野 寛<sup>‡</sup> 岡田 謙一<sup>‡</sup>

### 概要:

本研究では、デスクワークと遠隔会議閲覧の多重ワーク支援を目的とし、多重ワークにおける遠隔会議中継カメラの自動切替手法を提案した。まず、複数台のカメラが撮影した複数の映像・音声、センサーを用いて取得した参加者の視線データをメモリ上に  $\Delta t$  秒間蓄積する。次に、その  $\Delta t$  秒の間に音声・視線データから発話情報、ノンバーバル情報を取得する。最後に、蓄積された発話情報、ノンバーバル情報を基に聞き手の反応を推定し、その反応を基に映像切替の抑制、適切な発話者への切替を行う。評価実験を通じて、本手法を用いて視聴者への負担が少ない映像生成が可能であること、特に負荷がかかった状態で話の内容を理解しやすい映像生成が可能であることを確認した。

## Automatic Switching Technique of Remote Conference Video Image for Multitasking Workers with Multiple Cameras

Kousuke Tsumura<sup>†</sup>, Tetsuo Sumiya<sup>†</sup>, Itaru Takata<sup>‡</sup>, Hiroshi Shigeno<sup>‡</sup>, Ken-ichi Okada<sup>‡</sup>

### Abstract:

In this paper, we aimed at the multiple work support between deskwork and the remote conference. We proposed the automatic switch technique of the remote conference relay camera in the multiple work. First of all, we stored multiple data of video and audio shot by multiple cameras and glance data of each participant by sensors for a length of  $\Delta t$  in the memory. Next, utterance information and nonverbal information will be acquired from the voice and the glance data in the  $\Delta t$ . Finally, the listener's reaction is presumed based on accumulated utterance information and nonverbal information, the video image shot switch is controlled based on the reaction, and it switches to appropriate those who utter it. Through the evaluation experiment, We confirmed that this technique generated the video image shot with a little load to the viewer, especially, when the load hung to the listener, it was confirmed that the effect was high.

### 1. はじめに

近年、情報社会の進展によりオフィスワーカーの仕事量は増加している。これに対し、高生産性を確保するために同一時間帯に複数の仕事を遂行する多重ワークという新しいワークスタイルが注目されている<sup>1)</sup>。多重ワークを支援する要求事項として、複数作業の状況把握を容易にし作業の切り替えを支援することが挙げられる。このような多重ワーク支援を目的とする研究として、ユーザーの仕事の時間管理を目的としたもの<sup>2),3)</sup>、複数ワークを頻りに切替え人間の視覚的短期記憶の性質を検討しているもの<sup>4)</sup>、メインワーク以外の情報を周辺提示するもの<sup>5)~7)</sup>がある。一方、実際のオフィスにおける仕事に注目すると、多くのオフィスワーカーは会議の出席や資料作成等のデスクワークに対して非常に多くの時間を費やしている。こうした事実から、オフィスワーカーは個人で複数のタスクを抱え、さらに対面コミュニケーションによってタスクの切り替えや

中断を余儀なくされていることが明らかとなってきた<sup>8)</sup>。しかし、現在の IT 機器の多くは書類作成や E-mail などの個別のタスクを支援することを目的としてデザインされており、複数タスクにおける関係を考慮していない<sup>9)</sup>。我々はこのことをふまえ多重ワーク支援という研究に注目した。多重ワーク支援は同時に複数業務の遂行を支援することを目的としているが、本研究では特に 2 つのタスクの関係に焦点を当てた。オフィスワークにおけるタスクをクラス分けしたものを図 1 に示す。図 1 において、我々は相手がいて中断や干渉の制御がしにくいリアルタイムコミュニケーションを「対話ワーク」と定義した。また、E-mail や書類作成などリアルタイム性がなくタイミングが個人に依存するものを「非対話ワーク」と定義した。オフィスワークは多様化しつつあるが、大まかな枠でとらえれば図 1 のように分類できる。オフィスワークをコントロール性という面でもとらえた場合、対話ワークの処理を考慮することが大切だと考える。そこで、本研究では資料作成等のデスクワークと遠隔会議閲覧の多重ワークの支援を目的とし、具体的には多重ワークにおいても視聴者に負担が少なく、内容理解度が高い映像の生成を目指す。

映像の自動生成を目的とした研究として、カメラのズー

<sup>†</sup> 慶應義塾大学大学院理工学研究科

Graduate School of Science and Technology, Keio University

<sup>‡</sup> 慶應義塾大学理工学部

Faculty of Science and Technology, Keio University

|        |        | サブワーク                                      |                              |
|--------|--------|--|------------------------------|
|        |        | 対話ワーク                                      | 非対話ワーク                       |
| メインワーク | 対話ワーク  | 例)<br>二つの会議に<br>同時に出席する                    | 例)<br>会議に出席しながら<br>メールの処理を行う |
|        | 非対話ワーク | 例)<br>デスクで文書作成を<br>しながら後ろに集まって<br>いる人の話を聞く | 例)<br>特許を閲覧しながら<br>メールの処理を行う |

図 1 多重ワークの分類

ム機能・首振り機能により撮影領域や撮影対象を自動決定するカメラワークを実現する研究<sup>10)~13)</sup>，複数のカメラ映像から現在のシーンを表現するのに適した一つの映像を選択する研究<sup>11)~13)</sup>が行われている。しかし，視聴者の環境を考慮して映像を選択する研究は行われておらず，多重ワーク環境の視聴者に既存手法を用いて生成した映像を見せることは視聴者への負担，映像の内容理解度の点で問題が生じる。

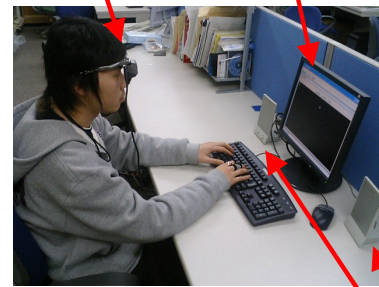
そこで本研究では，デスクワークと遠隔会議閲覧の多重ワーク支援を目的とし，多重ワークにおける遠隔会議中継カメラの自動切替手法を提案する。本手法では，複数台のカメラが撮影した複数の映像・音声，センサーを用いて取得した参加者の視線データをメモリ上に  $\Delta t$  秒間蓄積する。次に，その  $\Delta t$  秒の間に音声・視線データから発話情報，ノンバーバル情報を取得する。最後に，蓄積された発話情報，ノンバーバル情報を基に聞き手の反応を推定し，その反応を基に映像切替の抑制，適切な発話者への切替を行う。実際に提案手法に基づき映像切替を行うプロトタイプシステムを実装し，評価実験により本手法の有効性を確認する。

以下，2章では関連研究であるタイムシフトを用いた映像切替手法について，3章では提案手法について，4章では実装について，5章では評価実験と結果と考察について述べ，6章を本論文のまとめとする。

## 2. タイムシフトを用いた映像切替手法

本項では，我々がこれまで行ってきたタイムシフトを用いた映像切替手法<sup>14)</sup>について述べる。この手法では，撮影中の映像を故意に一定時間遅らせて送出したものをタイムシフトを用いたストリーミングと定義している。タイムシフトを利用した切替では，映像を送出前に蓄積するためシーンを撮影した時刻に対して，視聴者に送出される映像が蓄積時間だけ遅れている。この蓄積時間の間にセンサーを用いて発話等のイベント情報を蓄積しておく。最後に，蓄積されたイベント情報を基にシーンの状況を判断し，一本の映像・音声に編集する。意図的に映像と音声を遅らせて送出することで，数秒先の出来事を完全に把握したのと同様の映像編集が可能となる。これにより，従来のリアルタ

## 単眼HMD Desktop Display



## Speaker

図 2 HMD とデスクトップディスプレイを用いた多重ワーク支援環境

イムの切替では不可能であった発話者の存在感を強調するために発話を行う数秒前から発話者の映像に切替える「ずり上げ切替」など，録画番組に用いられる切替による演出が可能である。本研究でも，タイムシフトの概念を利用した映像切替手法を提案する。

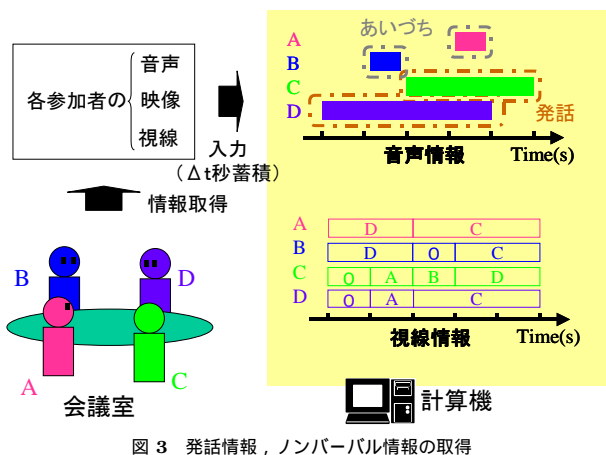
## 3. 提 案

### 3.1 作業環境設定

本研究では，図 2 に示すような HMD とデスクトップディスプレイを用いた作業環境を設定し，デスクワークと遠隔会議閲覧の多重ワークの支援を目指す。ディスプレイを多層化させて設置させることにより，多重ワーク支援の要求事項である複数作業の同時状況把握を容易にし，複数作業の切替を支援できる。図の作業者が行っているようにデスクトップ画面でメインワークであるデスクワークを行い，単眼 HMD 画面でサブワークである遠隔会議閲覧を行う。

### 3.2 本手法のアプローチ

前述の通り，本研究で想定する視聴者の環境は多重ワーク環境である。このような環境では，視聴者の負荷を減らすために全体を通して映像の切替をできるだけ抑えること，内容を把握させるために適切な発話者へと映像を切替えることが必要である。そこで，本手法では発話者に対する聞き手の反応に着目する。我々が聞き手の反応に着目する理由は以下の通りである。通常の対話において，聞き手は参加者の発話に対し何らかの反応を示す。例えば，発話に対するあいづちは自分がその発話を聞いているという相手への意思表示となり，発話者とのコミュニケーションを円滑にする。また，発話者に対し視線を向けることもその発話に注目しており，その発話を聞いているという意思表示となる。我々はこのことに着目し，聞き手が相手の発話を聞いているという意思表示が多い，つまり聞き手の反応が大きい程その発話は重要な情報を含んでいると考えた。聞き



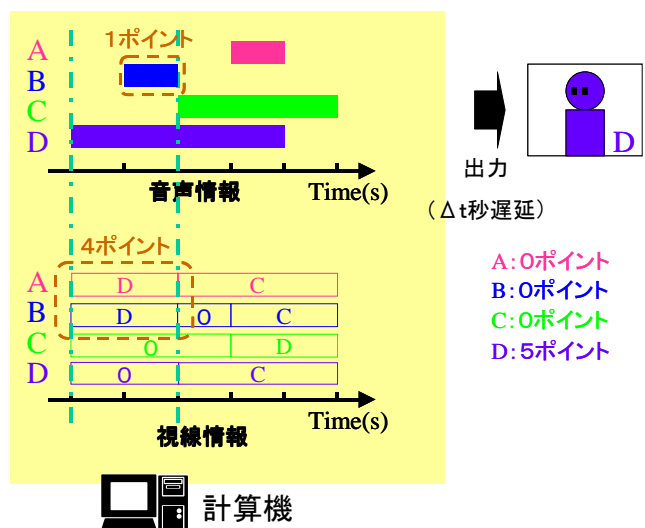
手の反応に合わせ, 反応が少ない発話に対してはその発話者の映像を映さず, 反応が大きい発話に対してはその発話者の映像を映すことで, 多重ワーク環境での視聴に適した視聴者に負担の少なく, 内容理解度の高い映像を生成できると考えられる. このような考えに基づき, 多重ワーク中の視聴者を考慮した聞き手の反応に基づく映像切替手法を提案する.

### 3.3 提案手法

本研究では, 撮影対象として会議シーンを対象としている. 本手法ではまず, 複数台のカメラが撮影した複数の映像・音声, センサーを用いて取得した参加者の視線データをメモリ上に  $\Delta t$  秒間蓄積する. 次に, その  $\Delta t$  秒の間に音声・視線データから発話情報, ノンバーバル情報を取得する. 最後に, 蓄積された発話情報, ノンバーバル情報を基に聞き手の反応を推定し, その反応を基に映像切替の抑制, 適切な発話者への切替を行い, 視聴者への負担が少なく内容理解度の高い映像を生成する. 以下, 発話情報とノンバーバル情報の取得方法, 聞き手の反応を考慮した映像切替手法について述べる.

### 3.4 発話情報, ノンバーバル情報の取得

図 3 に発話情報, ノンバーバル情報の取得の流れを示す. 会議室における複数カメラ, マイク, センサーから取得された各参加者の映像・音声・視線に関するデータはメモリ上に  $\Delta t$  秒間蓄積される. まず, 計算機は取得した音声データを基に,  $\Delta t$  秒間の音声のタイムテーブルを作成する. そして, そのタイムテーブルを基に 2 秒以上の音声を発話と認識し, 1 秒以下の音声をノンバーバル情報であるあいづちと認識する. 次に, 取得した視線データと会議のレイアウト情報から, 誰が誰を見ているかという視線に関するノンバーバル情報の  $\Delta t$  秒間のタイムテーブルを作成する. そして, 1 秒以内に視線の対象が移る場合は, その対象者に注目をしていないと判断しフィルタリングを行う. このようにして得られた発話情報, ノンバーバル情報から聞き手の反応を推定する.



### 3.5 聞き手の反応に基づく映像切替

映像切替は以下の手順で行う. (1) 発話時の参加者のノンバーバル情報を調べポイント換算する (あいづち 1 回:1 ポイント, 視線 1 秒:1 ポイント), (2) 各参加者についてポイントの算出を行う, (3) ポイントに応じて映像切替を行う. 映像切替の種類と動作条件を表 1 に示す.

新たな発話があった場合, ノンバーバル情報を調べる範囲は発話開始時から発話が終了するまで, または他の参加者によって発話が重複されるまでの範囲である (図 4). 図 4 の例では, D の発話中に B があいづちを入れたため D に 1 ポイント加算される. また, A と B が D に視線を送っているため 1 秒につき 1 ポイントで 4 ポイント加算され, 結果 5 ポイント取得している. このように新たな発話があった時, 発話者のポイントが聞き手のポイントを上回った場合はその発話者の映像に切替わる. しかし, 発話者のポイントが聞き手のポイントを下回った場合はその話者の映像へは切替えず, 前の映像を持続させる. このように聞き手の反応が少ない発言の映像を抑制することで切替数を減らし, 視聴者の負担を減らすことができる.

発話が重複した場合, ノンバーバル情報を調べる範囲は発話が重複されている範囲である (図 5). 発話が重複している時にあいづちが生じた場合は, あいづちが生じた時にあいづちを打った人物が向いている参加者に 1 ポイント加算される. 図 5 の例では, C が 6 ポイント, D が 1 ポイント取得している. このように話者同士のポイントと比較し, 2 ポイント以上高い場合はその発話者の映像に切替わり, 低いポイントの話者へは切替わらず抑制される. また, 話者同士のポイントと比較しその差が 1 ポイント以下であった場合は, どちらかの話者に切替るのではなく全体を映すシーンカメラの映像が選択される.

表 1 映像切替の種類と動作条件

| 発話状況        | 映像切替の動作条件           | 映像切替の種類           |
|-------------|---------------------|-------------------|
| 新たな発話が生じた場合 | 発話者のポイントが聞き手のポイント以上 | 発話者の映像に切替         |
|             | 発話者のポイントが聞き手のポイント以下 | 直前の映像を継続          |
| 発話が重複した場合   | 他発話者より 2 ポイント以上高い   | ポイントが高い発話者の映像に切替  |
|             | 他発話者との差が 1 ポイント以下   | 全体を映すシーンカメラの映像に切替 |

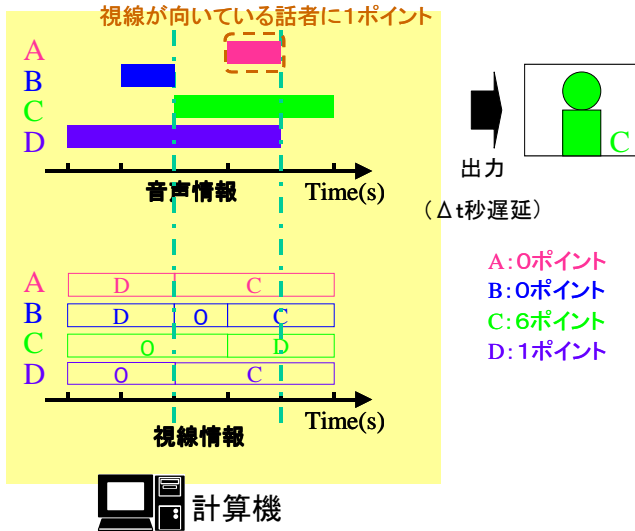


図 5 聞き手の反応に基づく映像切替 (発話が重複した場合)

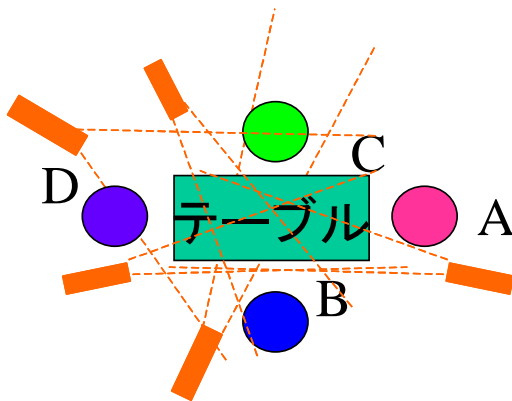


図 6 会議空間のレイアウト

#### 4. 実装

提案手法に基づいて、会議映像を自動切替するプロトタイプシステムを構築した。

##### 4.1 会議空間のレイアウト

撮影時のレイアウトを図 6 に示す。縦 5m × 横 10m 程度のスペースに 5 台の固定カメラを配置した。カメラ 1~4 は各話者を映すカメラとして、カメラ 5 は会議空間全体の撮影を行うシーンカメラとして使用した。

##### 4.2 映像切替アルゴリズム

映像切替アルゴリズムの構成図を図 7 に示す。各カメラにより撮影された映像は、アルゴリズムに基づいて切替が行われ、選択された映像が選出される。

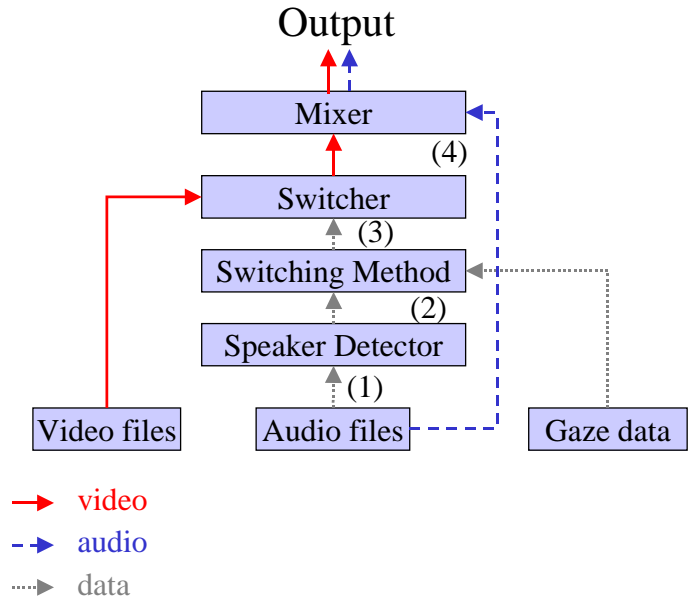


図 7 映像切替アルゴリズムの構成図

本研究では映像切替手法の妥当性の検証を行うため、視線情報は目視で取得した理想的なデータを用意した。また、会議の映像・音声もあらかじめ記録したデータを用意し、これらをアルゴリズムにかけ映像切替を行う。

- (1) 各マイクで取得された音声データ (Audio files) は、再生される際  $\Delta t$  秒間蓄積される。この  $\Delta t$  秒間に発話情報とノンバーバル情報 (あいづち) の取得を行う。
- (2) (1) で得られた発話情報、ノンバーバル情報と  $\Delta t$  秒間蓄積された視線情報 (Gaze data) を基に聞き手の反応を推定し、次のどの映像を選択するか決定する (Switching Method)。
- (3) 遅延された映像データ (Video files) はスイッチャー (Switcher) へ入力される。このスイッチャーを制御し (3) で決定したカメラの映像を選択し出力する。
- (4) この出力された映像は先ほどの遅延された音声とミックスされ出力される (Mixer)。

##### 4.3 実装環境

ソフトウェア環境に関しては、音声認識部、スイッチング方法決定部、スイッチャー、ミキサーの各モジュールは J2SDK1.4 と JMF2.1.1e API の Java 言語で実装した。ハードウェア環境に関しては、カメラは Canon 社製 VC-CI, Sony 社製 DCR-VX2000 を、マイクは Ele-

com 社製 Multimedia Earphone with Microphone(MS-HS59SC) を使用した。各ソフトウェアを実行する計算機として DELL 社製 DIMENSION8300(CPU:Pentium4 1.5GHz OS:WindowsXP Professional) の PC を使用した。

## 5. 評価

本手法を用いて生成した映像が視聴者に負担の少ない映像であるか、多重ワーク環境下で閲覧した場合視聴者にどのような影響を与えるか、を確かめるために評価を行った。

### 5.1 切替の適切度

切替の適切度に関しては抑制すべき発話を本手法により抑制できた箇所の一致率、そして映像を通して切替の生じた回数を評価した。抑制すべき発話は以下の 3 項目<sup>15)</sup> を満たす発話と定義した。

- 誰かの発話に対する返答
- 相手が返答する必要のない言葉
- 完全な会話へ発展しない言葉

また、抑制すべき箇所の一致率  $P$  を以下の様に定義した。

$$P = \frac{N_m}{N_d} \times 100(\%)$$

( $N_m$ : 抑制すべき発話箇所を本手法により抑制できた箇所の合計、

$N_d$ : 抑制すべき発話箇所の合計)

評価には、参加者 4 名で 5 分間行ったフリーディスカッションの映像を用いた。抑制すべき発話箇所は 39カ所であった。

#### 5.1.1 比較手法

本手法の有用性を検証するために、以下の比較手法を用意した。

- 発話自動切替手法  
発話時に話者の映像に自動的に切替える。
- 状態遷移自動切替手法  
過去の発話履歴から確率的に次の話者と発話時刻を予想し切替える。
- タイムシフトを用いた切替手法  
本手法と同様に  $\Delta t$  秒発話情報を蓄積し、その情報を基に切替える。

また、本研究では映像切替の妥当性について検証するため、誰が誰を見ているという視線情報には目視で取得した理想的なデータを用意した。しかし、実際にシステムに組み込む際には多少のエラーが出るが自動検知を用いた方が実用性が上がる。そこで本手法のエラーに対する耐性を評価するため提案手法で用いる視線情報に 3,5,8 % のノイズを加えたものを用意した。ここで言うノイズとは、各参加者の視線情報に 3,5,8 % の確率でランダムな値を混入するものである。

#### 5.1.2 実験結果

表 2 に、既存手法と比較した一致率、切替数の結果を示

表 2 一致率と切替数の評価結果 (既存手法との比較)

| 評価項目    | 提案手法 | 発話切替  | 状態遷移  | タイムシフト |
|---------|------|-------|-------|--------|
| 一致率 $P$ | 73 % | 3 %   | 12 %  | 31 %   |
| 切替数     | 38 回 | 126 回 | 126 回 | 77 回   |

表 3 一致率と切替数の評価結果 (ノイズ混入時との比較)

| 評価項目    | 提案手法 | ノイズ 3 % | ノイズ 5 % | ノイズ 8 % |
|---------|------|---------|---------|---------|
| 一致率 $P$ | 73 % | 73 %    | 71 %    | 67 %    |
| 切替数     | 38 回 | 38 回    | 38 回    | 40 回    |

す。本手法では一致率 73 %、切替数 38 回であった。一方、発話自動切替手法では一致率 3 %、切替数 126 回、状態遷移自動切替手法では一致率 12 %、切替数 126 回、タイムシフトを用いた切替手法では 31 %、77 回となった。つまり、提案手法は既存手法に比べ抑制すべき場所を多く抑制しており、切替数も少なく済んでいた。このことから、本手法は既存手法よりも視聴者に負担の少ない映像を生成できていることがわかる。

表 3 に、提案手法で用いる視線情報にノイズを混ぜた場合と比較した一致率、切替数の結果を示す。ノイズを混入しない場合では一致率 73 %、切替数 38 回であった。一方、ノイズを 3 % 混ぜた場合では一致率 73 %、切替数 38 回、ノイズを 5 % 混ぜた場合では一致率 71 %、切替数 38 回、ノイズを 8 % 混ぜた場合では一致率 67 %、切替数 40 回であった。つまり、ノイズを混ぜた場合でも結果はあまり変わらないということがわかる。このことから、本手法は視線情報のデータのエラーにも耐性があり自動検知を用いた場合でも有効であると言える。

### 5.2 多重ワークにおける映像評価

本手法を用いて生成した映像が、多重ワーク環境の視聴者にどのような影響を与えるかを確かめるために実際に多重ワークを行い、その結果から映像の評価を行った。

被験者の作業環境は図 2 に示す通りである。ワークの内容は、デスクトップ画面でメインワークであるタイプライクを行い、HMD 画面でサブワークである会議の閲覧を行う。タイプライクは 3~6 秒のランダムな間隔で表示された 2~5 文字のローマ字をタイプするものであり、HMD 画面の映像が終了すると同時に表示も終了する。HMD 画面に表示される会議は 1 分間の映像であり、音声はデスクトップ脇のスピーカーから出力される。実験は、(1) 被験者は会議映像を見ながら、表示される文字をタイプする、(2) 映像終了後に会議の内容に関する問題を解く、という流れで行う。

#### 5.2.1 評価項目と比較手法・環境

実験の評価項目として問題の正解率、タイプの精度を設定し、本手法の有用性を示すための比較手法として発話自動切替手法を用意した。また、多重ワーク環境ではない視聴者に対する影響も調べるため、2 つのワークを個別に遂行する個別ワーク環境でも実験を行った。以上の実験を計算機の使用に慣れている大学生 20 名で行った。



| 作業環境  | 提案手法   | 発話切替   |
|-------|--------|--------|
| 多重ワーク | 66.9 % | 53.9 % |
| 個別ワーク | 76.9 % | 75.0 % |

| 作業環境  | 提案手法   | 発話切替   |
|-------|--------|--------|
| 多重ワーク | 96.7 % | 95.0 % |
| 個別ワーク | 98.5 % |        |

### 5.2.2 実験結果

表 4 に問題の正解率の結果，表 5 にタイプ精度の結果を示す．メインタスクであるタイプ精度に関しては，多重ワーク環境でも個別ワーク環境でも結果にあまり差はなかった．しかし，サブワークである問題の正解率に関しては個別ワーク環境の方が多重ワーク環境よりも良い結果となった．このことから，多重ワークではメインワークへの影響は比較的少ないが，サブワークへの影響は多く出ることがわかる．また，個別ワーク環境における問題の正解率は提案手法によって生成された映像を見た場合と，発話切替手法によって生成された映像を見た場合であまり差は見られなかった．しかし，多重ワーク環境では提案手法によって生成された映像を見た場合の正解率の方が高い．このことから，提案手法は多重ワークのような視聴者に負荷がかかった環境でより効果を発揮することがわかる．

## 6. おわりに

本研究では，デスクワークと遠隔会議閲覧の多重ワーク支援を目的とし，多重ワークにおける遠隔会議中継カメラの自動切替手法を提案した．まず，複数台のカメラが撮影した複数の映像・音声，センサーを用いて取得した参加者の視線データをメモリ上に  $\Delta t$  秒間蓄積する．次に，その  $\Delta t$  秒の間に音声・視線データから発話情報，ノンバーバル情報を取得する．最後に，蓄積された発話情報，ノンバーバル情報を基に聞き手の反応を推定し，その反応を基に映像切替の抑制，適切な発話者への切替を行い，視聴者への負担が少なく内容理解度の高い映像を生成する．評価実験を通じて，本手法を用いて視聴者への負担が少ない映像生成が可能であること，特に負荷がかかった状態で話の内容を理解しやすい映像生成が可能であることを確認した．

今回の論文においては，視線情報に目視によって取得した理想的なデータを用いたが，評価実験により本手法は視線情報のエラーに対しても耐性があることがわかった．よって，今後は視線情報の取得に自動検知システムを用い，より実用的なシステムの構築を目指す．

謝辞 本研究の一部は，21 世紀 COE プログラム研究拠点形成費補助金のもとに行われた．ここに記して謝意を表す．

## 参考文献

- 1) Star-telegram.com and David Meyer. Multitasking makes you stupid, studies say. <http://www.dfw.com>, 2003.
- 2) Roy Rondenstein. Owntime: a system for timespace management. Proceedings of CHI99, 1999.
- 3) O 'Conaill, Frohlich, and David. Timespace in the workplace. Proceedings of CHI95, 1995.
- 4) 横澤. 変化検出課題における視覚的短期記憶の性質. 心理学研究, 73, 6, 464-471, 2003.
- 5) Blair MacIntyre. Support for multitasking and background awareness using interactive peripheral displays . Proc. Of ACM, 2001.
- 6) Andrew Dahley. Water lamp and pinwheels: ambient projection of digital information into architectural space. Proceedings of CHI98, 1998.
- 7) Jacob Somervell and D.Scott McGrickard. Secondary task display attributes-optimizing visualizations for cognitive task suitability and interference avoidance. Proceedings of the symposium on Data Visualisation 2002, 2002.
- 8) V. M.Gonzales and G. Mark. Constant, Constant, Multi-tasking Crazines. In Proceedings of CHI '04, pp 113?120, 2004.
- 9) Mary Czerwinski, Eric Horvitz, and Susan Wilhite.: A Diary Study of Task Switching and Interruptions, Proceedings of CHI 2004, pp.175-182 (2004)
- 10) C.S.Pinhanes and A.F.Bobick. Approximate world models: Incorporating qualitative and linguistic information into vision systems. Proc. AAAI '96, pp. 1116?1123, Aug 1996.
- 11) 尾関基行, 伊藤雅嗣, 中村裕一, 大田友一. 複合コミュニケーション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.
- 12) 井口泰典, 土居元紀, 真鍋佳嗣, 千原國宏. 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.
- 13) 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. 情処研報, CVIM-125-5, Jan 2001.
- 14) 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 (掲載予定), Vol.47, No.3, March 2006.
- 15) Nigel Ward. Using prosodic clues to decide when to produce back-channel utterances. In Proc International Conference on Spoken Language Processing., 1996.