多重ワークにおける会議中継カメラの自動切替え手法

住谷哲夫[†] 津村弘輔[††] 高田格[††]
重野寛[†††] 岡田謙一[†††]

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え手法を提案した．まず，複数台のカメラが撮影した複数の映像・音声，センサを用いて取得した参加者の視線データをメモリ上に Δt 秒間蓄積する．次に，その Δt 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えの抑制，適切な発話者への切替えを行う．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く比較的内容理解度が高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 Method of Remote Conference Cameras for Multitasking Workers

Tetsuo Sumiya,[†] Kousuke Tsumura,[††] Itaru Takata,[††]
Hiroshi Shigeno[†††] and Ken-ichi Okada[†††]

The purpose of this paper is to support multiple works between deskwork and browsing remote conference. We proposed automatic camera switching method of remote conference for multitasking worker. First, the audio/video streams are captured by multiple video cameras, and participants' gaze data are captured by multiple sensors. Those streams are not directed and transmitted in real time, but are buffered on a memory for Δt seconds. During the buffering, the conversational and nonverbal history of the scene is collected by various sensors simultaneously. Next, participants' reaction is presumed based on the conversational and nonverbal history. Finally, the buffered streams are directed and merged into one stream based on the participants' reaction. A prototype system was implemented and evaluated by experiments. We evaluated switching adequacy and effect in multiple works environment. The experimental results indicated that the proposed method puts less stress and gets content across easily, in particular, the effect was high in multiple work environment.

1. はじめに

近年，情報社会の進展によりオフィスワーカの仕事量は増加しており，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[1]．このような多重ワークの支援を目的とする研究として，ユーザの仕事の時間管理を目的としたもの[2),3)]，複数ワークを頻繁に切り替え人間の視覚的短期記憶の性質を検討しているもの[4)]，メインワーク以外の情報を周辺提示するもの[5)~7)]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は会議の出席や資料作成など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．そこで，本研究では資料作成など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[8)~10)]，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1つの映像を選択する研究[11)~13)]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

† 株式会社エヌ・ティ・ティ・ドコモ
NTT DoCoMo, Inc.
†† 慶應義塾大学大学院理工学研究科
Graduate School of Science and Technology, Keio University
††† 慶應義塾大学
Keio University

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え手法を提案する．本手法では，複数台のカメラが撮影した複数の映像・音声，センサを用いて取得した参加者の視線データをメモリ上に Δt 秒間蓄積する．次に，その Δt 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えの抑制，適切な発話者への切替えを行う．実際に提案手法に基づき映像切替え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え手法について，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え手法

本章では，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え手法[13] 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えでは，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音声データから発話情報を蓄積しておく．最後に，蓄積された発話情報を基にシーンの状況を判断し，1 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等の映像編集が可能となる．これにより従来のリアルタイムの切替えでは不可能であった切替えによる演出，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り替える「ずり上げ切替え」などの録画番組に用いられる切替え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え手法を提案する．

ここで，今回の実験では会議参加者と遠隔参加者とのインタラクションはとらないものとする．理由は今回の実験で対象としている会議は，実際に会議場で行われている会議に参加することができない参加者が，遠隔地からその会議を聴講するという場面を想定しており，なかでも同時刻に見ていることが重要な会議である伝達会議を想定し実験を行った．つまり，伝達会議であれば基本的にインタラクションをとる必要はなく，そのような会議を想定しているため，今回はインタラクションをとらないものとする．

3. 提　　案

3.1 作業環境設定

本研究では，**図 1** 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，複数作業の同時状況把握を容易にし，複数作業の切替えを支援できる．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

前述のとおり，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え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り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とおりである．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．たとえば，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ほど，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，

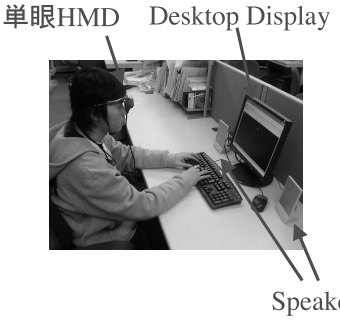

図 1 HMD とデスクトップディスプレイを用いた多重ワーク支援環境

Fig. 1 Multitasking support environment with HMD and Desktop Display.

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え手法を提案する．

また，今回想定している多重ワークでは対話ワークと非対話ワークを同時に行うことを目的としている．具体的には，対話ワークとは会議などの相手とのリアルタイムの対話を必要とする同期的なワークのことをいい，非対話ワークとは文書作成などの個人で中断や継続を制御することが可能な非同期的なワークのことを指す．そして対話ワークは複数人で行うワークになり，非対話ワークは基本的に 1 人で行うタスクとなる．しかしながら今回の提案ではすべてを網羅しているのではなく，メインワークとして非同期ワークであるデスクワーク，サブワークとして同期ワークである遠隔会議参加の支援を目指している．

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，センサを用いて取得した参加者の視線データをメモリ上に Δt 秒間蓄積する．次に，その Δt 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えの抑制，適切な発話者への切替えを行い，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え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 2 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 Δt 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，Δt 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 1 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，だれがだれを見ているかという視線に関するノンバーバル情報の Δt 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．

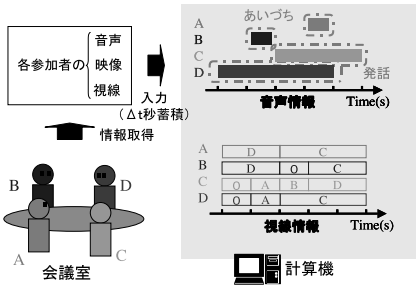

図 2 発話情報，ノンバーバル情報の取得
Fig. 2 Acquisition of utterance information and nonverbal Information.

ここで，当然ノンバーバル情報としてのあいづちには音声をともなうものとともなわないものが存在する．しかし，音声をともなうものもともなわないものも同じあいづちという意味合い，つまり発話者に対する反応という点では同じであると考えた．そこで会議参加者が発話者の発言内容に対して反応を持った場合には，音声（うん，なるほど，といったような発言）を発してもらい，あいづちとすることにした．つまり，今回は実験映像を作成するにあたって，あいづちを行う際は必ず音声を発してもらったため，今回の実験映像に関しては声を発しないあいづちは存在しないことになる．

また，それぞれの認識判断の根拠として，テレビ番組や討論番組の録画映像の分析や，実際の会議風景を撮影し，それぞれの参加者の発言を細かく分析した．その結果として，ほとんどのあいづちが 1 秒以下で行われていることが分かった．そこで今回は観察結果から我々の主観で 1 秒という閾値を決定し，これを用いて発話情報とノンバーバル情報の分類を行った．

3.5 聞き手の反応に基づく映像切替え

映像切替え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えを行う．映像切替え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 3**）．図 3 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

表 1 映像切替えの種類と動作条件
Table 1 Type of image switching and operation conditions.

発話状況	映像切替えの動作条件	映像切替えの種類
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 発話者のポイントが聞き手のポイント以下	発話者の映像に切り替え 直前の映像を持続
発話が重複した場合	他発話者より 2 ポイント以上高い 他発話者との差が 1 ポイント以下	ポイントが高い発話者の映像に切り替え 全体を映すシーンカメラの映像に切り替え

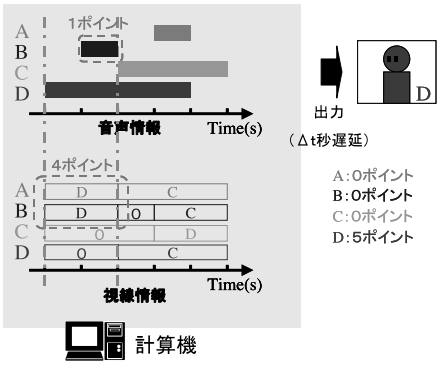

図 3 聞き手の反応に基づく映像切替え（新たな発話が生じた場合）
Fig. 3 Image switching based on reaction of the listener (New utterance occurrence).

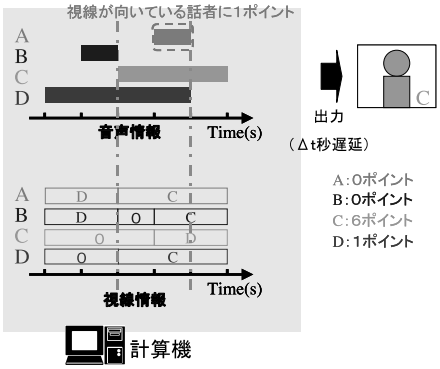

図 4 聞き手の反応に基づく映像切替え（発話が重複した場合）
Fig. 4 Image switching based on reaction of the listener (The utterance overlap).

る．このように新たな発話があったとき，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り替わる．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発話者の映像へは切り替えず，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え数を減らし，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 4**）．発話が重複しているときにあいづちが生じた場合は，あいづちが生じたときにあいづちを打った人物が視線を向けている参加者に 1 ポイント加算される．図 4 の例では，C が 6 ポイント，D が 1 ポイント取得している．このように発話者同士のポイントを比較し，他の発話者より 2 ポイント以上高い場合はその発話者の映像に切り替わり，低いポイントの発話者へは切り替えない．また，発話者同士のポイントを比較しその差が 1 ポイント以下であった場合は，どちらかの発話者に切り替わ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．

4. 実　　装

提案手法に基づいて，会議映像を自動切替え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 5** に示す．縦 5 m×横 10 m 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 映像切替えアルゴリズム

映像切替えアルゴリズムの構成図を**図 6** に示す．各カメラにより撮影された複数の映像は提案アルゴリズムに基づいて選択され，その映像が出力される．

本研究では映像切替え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えを行う．

(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 Δt 秒間蓄積される．この Δt 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 Δt 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，次に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（Switcher）へ入力される．このスイッチャを制御し，(2) で決定したカメラの映像を選択し出力する．

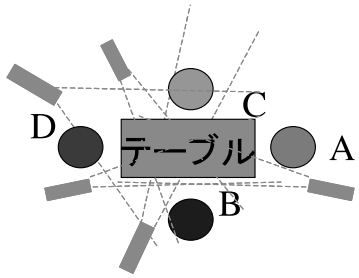

図 5 会議空間のレイアウト
Fig. 5 Layout of the conference room.

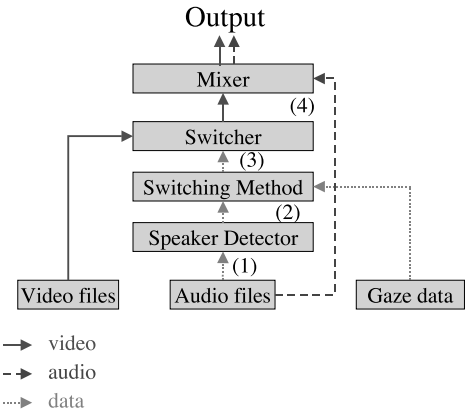

図 6 映像切替えアルゴリズムの構成図
Fig. 6 Architecture of image switching algorithm.

(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，Speaker Detector, Switching Method, Switcher, Mixer の各モジュールを 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 Elecom 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 1.5 GHz，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の切替えが適切であるか，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 切替えの適切度

切替え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，映すべき発話を本手法により映せた箇所の一致率，映像を通して切替えの生じた回数を評価項目とした．抑制すべき発話は以下の 3 項目を満たす発話と定義した[14]．

- だれ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

映すべき発話は以下の 2 項目のいずれかを満たす発話と定義した．

- 重要な内容を含む言葉
- 完全な会話へ発展する言葉

今回のそれぞれの発話の定義は実験者の主観的観点から決定した．まず，抑制すべき発話箇所であるだれかの発話に対する返答とはあいづちが考えられ，これらは単なる返答ということなので，これをだれかの発話に対する返答とした．次に，相手が返答する必要のない言葉に関しては，「そうなんだ.」といったようないわゆるぼやきに近い発言と定義した．最後に，完全な会話へ発展しない言葉に関しては，繰返しの意味を込めた聞き返しという言葉であり，今回はこれらの発言を完全な会話へ発展しない言葉として定義した．

次に，映すべき発話箇所である重要な内容を含む言葉とは，時間や場所，人などの重要なキーワードを含むような言葉であり，今回はこれらを含む言葉を重要な内容を含む言葉と定義した．同じく，完全な会話に発展する言葉とは，質問や話題転換といった発言を含むものを完全な会話に発展する言葉として定義した．

また，抑制すべき箇所の一致率 P を以下のように定義した．

- 一致率 P

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

（N_m：抑制すべき/映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/映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，映すべき箇所は 53 カ所であった．これらの決定方法は上記定義に基づいて，実際に撮影された会議映像・音声を通して，実験者の主観で決定した．

また決定箇所の同定の妥当性に関しても，今回の会議において話されている会議内容は，我々があらかじめ考えた内容・シナリオを基に行っているため，すべての会話内容を最初から把握している．そのため，実際の会議を分析してこの部分はどこに該当するのかを判断した．つまり発話箇所の同定に関しては，あらかじめ分かっている会議内容を基に決定したため妥当であると考える．

表 2 一致率と切替え数の評価結果（既存手法との比較）
Table 2 Evaluation both match number and switching number (Comparison of the existing method with our method).

評価項目	提案手法	発話切替え	状態遷移	タイムシフト
抑制箇所 P	73%	3%	12%	31%
映す箇所 P	74%	98%	78%	92%
切替え数	38 回	126 回	126 回	77 回

表 3 一致率と切替え数の評価結果（ノイズ混入時との比較）
Table 3 Evaluation both match number and switching number (Comparison of the mixed noise with our method).

評価項目	提案手法	ノイズ 5%	ノイズ 10%	ノイズ 20%
抑制箇所 P	73%	71%	64%	54%
映す箇所 P	74%	74%	68%	62%
切替え数	38 回	38 回	44 回	44 回

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．

- 発話自動切替え手法
  発話時に話者の映像に自動的に切り替える．
- 状態遷移自動切替え手法
  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り替える．
- タイムシフトを用いた切替え手法
  本手法と同様に Δt 秒発話情報を蓄積し，その情報を基に切り替える．

また，本研究では映像切替えの妥当性について検証するため，だれがだれ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．目視で取得したとは，まず図 5 のようにそれぞれの会議参加者の映像を，各参加者の正面映像を取得するようにセットされたカメラを用いて会議参加中の様子を映像・音声とともに取得し，次にその取得したそれぞれの映像を実際に見ることで，同じく図 5 の会議空間のレイアウト情報，つまりだれがどこに座っているかという情報からその人が今どの方向を見ているのかという視線情報を取得した．

今回は視線情報の正確性をあげるために目視による取得を行ったが，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 5%，10%，20%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 5%，10%，20%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，切替え数の結果を示す．この結果よ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いる．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分かる．一方，映す箇所に関しては既存手法よりも多少低い結果となっている．これは，他の手法は発言に対してフィルタリングをほとんど行わず多くの箇所を映すため，映すべき箇所もカバーできたためと考えられる．しかし，切替え数に注目すると本手法は発話切替え手法の 30%，状態遷移切替え手法の 30%，タイムシフトを用いた切替え手法の 49%の切替え数で済んでいる．よって，映す箇所を完全にカバーすることはできないが切替え自体の回数を減らすことができ，視聴者に負担が少なく内容理解度も比較的高い映像生成が可能であるといえる．

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，切替え数の結果を示す．実際に 5%，10%，20%のノイズを混入した映像を被験者に見てもらったところ，被験者の主観的観点から，10%までなら被験者は違和感なく会議映像を見ることができたということが分かった．つまり，10%までならノイズを混ぜても結果はあまり変わらないといえる．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考えられる．

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．また，実験後にアンケートをとり定性評価を行った．

被験者の作業環境は図 1 に示すとおりである．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

表 6 アンケートの結果
Table 6 Results of the quesionnaire.

質問項目	提案手法の平均得点	既存手法の平均得点	Wilcoxon 符号付順位検定 P 値
1. 切替えのタイミングは適切だったか	4.4	2.2	***0.00391
2. 見たい映像に切り替わっていたか	4.2	2.2	**0.00781
3. 見やすい映像だったか	3.8	2.3	*0.01563
4. カメラの切替えに違和感を感じなかったか	4.1	2.2	*0.01563
5. ストレスを感じなかったか	3.6	2.1	*0.01563
6. 議論の流れがつかめたか	4.2	3.0	*0.02734
7. 話の内容を理解しやすかったか	3.7	2.4	*0.03125

(n = 20, ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$)

表 4 問題正解率
Table 4 Accuracy rate of question.

作業環境	提案手法	既存手法
多重ワーク	66.9%	53.9%
個別ワーク	76.9%	75.0%

表 5 タイプ精度
Table 5 Accuracy rate of typing.

作業環境	提案手法	既存手法
多重ワーク	96.7%	95.0%
個別ワーク	98.5%	

タイプワークを行い，HMD 画面でサブワークである会議の閲覧を行う．タイプワークは 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，(3) アンケート用紙に記入する，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え手法を用意した．また，多重ワーク環境ではない視聴者に対しての影響も調べるため，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．

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分かる．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，既存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分かる．

5.2.3 アンケート評価

本手法を用いた生成した映像が，視聴者にどのような影響を与えるかを確かめるために映像の主観評価を行った．前述の実験を行った被験者 20 名に対し，多重ワーク環境における提案手法，既存手法で生成された各映像についてアンケートに 5 段階で評価してもらった．アンケートの質問項目を**表 6** に示す．映像の見映えに関する質問（項目 2，3，5），映像の切替えに関する質問（項目 1，4），内容の理解に関する質問（項目 6，7）を用意した．

アンケートの結果を表 6 に示す．各質問は「まったくあてはまらない」，「あまりあてはまらない」，「どちらともいえない」，「ややあてはまる」，「かなりあてはまる」の 5 段階にそれぞれ 1 点から 5 点を与え，映像別に質問に対する平均得点を求めた．表中の各項目の値は評価値の平均得点である．さらに，評点に有意差があるか確認するため Wilcoxon の符号付順位検定 p 値[15),16)] を求めた．

映像の見映えに関する質問の結果から，項目 2 に関しては危険率 1%以下，項目 3，5 に関しては危険率 5%以下で違和感なく見やすい映像であるという評価を得た．また，映像の切替えに関する質問の結果から，項目 1 に関しては危険率 0.1%以下，項目 4 に関しては危険率 5%以下で違和感のない自然な切替えになっているという評価を得た．本手法により聞き手の反応の少ない映像への切替えが抑制され，聞き手の注目を集めている重要な発言を映すことで見やすい映像，違

和感のない映像切替えが可能だったためと考えられる．内容の理解に関する質問の結果から，項目 6，7 に関しては危険率 5%以下で内容が理解しやすい映像になっているという評価を得た．本手法により，内容理解に無駄な発話が抑制され，重要な発話が映されたためと考えられる．一方で，項目 5，7 に関しては他の項目に比べて若干低い結果となった．これは多重ワーク環境で実験を行ったため負荷を感じていたこと，メインワークに集中し話しが聞き取りにくかったことが考えられる．しかし，そのような環境でも項目 6 に関しての結果が既存手法よりも高い結果となったことから，本手法によって生成された映像は多重ワーク環境でも見やすく，議論の流れをつかみやすい映像であるといえる．

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え手法を提案した．まず，複数台のカメラが撮影した複数の映像・音声，センサを用いて取得した参加者の視線データをメモリ上に Δt 秒間蓄積する．次に，その Δt 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えの抑制，適切な発話者への切替えを行い，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く比較的内容理解度が高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．

また今回の提案では会議としてインタラクションをとる必要性の少ない伝達会議を想定している．しかし，当然インタラクションをとる必要がある会議も十分想定される．そこで Δt 秒間遅延した映像からリアルタイムの映像へ倍速再生などを用いることによって映像を滑らかにつなぎ，スムーズにインタラクションをとれるようにする手法などを今後研究していく必要がある．

さらに今回は，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分かった．しかし視線情報は目視によって取得した理想的なデータを用いたため実用的ではないと考えられる．そこで，今後はより実用的なシステムを作るために，そのようなあいづちに対しても対処する必要があり，画像処理の観点から顔の動きを取得しうなずきを認識させることで，聞き手の反応を推定するパラメータを増やしていくなど，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 Meyer, D.: Multitasking makes you stupid, studies say (2003). http://www.dfw.com
2) Rondenstein, R.: Owntime: a system for timespace management, *Proc. CHI99* (1999).
3) O'Conaill, Frohlich and David: Timespace in the workplace, *Proc. 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 (2003).
5) Maclntyre, B.: Support for multitasking and background awareness using interactive peripheral displays, *Proc. ACM* (2001).
6) Dahley, A.: Water lamp and pinwheels: ambient projection of digital information into architectural space, *Proc. CHI98* (1998).
7) Somervell, J. and McGrickard, D.S.: Secondary task display attributes-optimizing visualizations for cognitive task suitability and interference avoidance, *Proc. Symposium on Data Visualisation 2002* (2002).
8) 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).
9) Liu, Q., Kimber, D., Foote, J., Wilcox, L. and Boreczky, J.: Flyspec:a multi-user video camera system with hybrid human and automatic control, *Proc. ACM Multimedia* (2002).
10) Lee, D.-S., Erol, B. and Graham, J.: Portable meeting recorder, *Proc. ACM Multimedia* (2002).
11) 竹前嘉修，大塚和弘，武川直樹：対面の複数人対話を撮影対象とした対話参加者の視線に基づく映像切替え方法とその効果，情報処理学会論文誌，Vol.46, No.7 (2005).
12) 井上亮文，平石絢子，柴 貞行，市村 哲，重野 寛，岡田謙一，松下 温：シナリオ情報によるオーケストラ演奏のカメラワーク生成手法，情報処理学会論文誌，Vol.46, No.1 (2005).
13) 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).
14) Ward, N.: Using prosodic clues to decide when to produce back-channel utterances, *Proc. International Conference on Spoken Language Processing* (1996).

15) 涌井良幸，涌井貞美：図解でわかる統計解析用語辞典，日本実業出版社 (2003).
16) 緒方裕光，柳井晴夫：統計学—基礎と応用，現代数学社 (1999).

(平成 18 年 3 月 27 日受付)
(平成 18 年 10 月 3 日採録)

住谷 哲夫（学生会員）

2004 年慶應義塾大学理工学部情報工学科卒業．2006 年同大学院前期博士課程修了．現在，NTT ドコモ株式会社に勤務．在学中，自動撮影技術，ウェアラブルコンピューティングの研究に従事．

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．現在，同大学院前期博士課程在学中．自動撮影技術，ウェアラブルコンピューティングの研究に従事．

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．自動撮影技術，ウェアラブルコンピューティングの研究に従事．

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助教授．工学博士．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，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 年，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．