

多重ワーク時における会議中継カメラの自動切替手法

津村 弘輔[†] 住谷 哲夫[†] 高田 格[‡] 重野 寛[‡] 岡田 謙一[‡]

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。一方、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。我々はこのことをふまえて多重ワーク支援という研究に注目した。そこで、本研究では資料作成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。

2. 提案

2.1 作業環境設定

本研究では、Head Mounted Display(以下,HMD とする)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。また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。



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

2.2 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。つまり聞き手の反応に合わせ、反

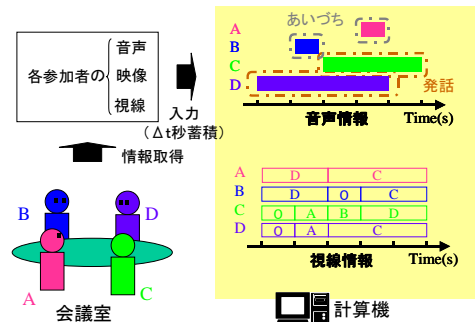


図 2 発話情報、ノンバーバル情報の取得

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、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た。

2.3 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¹⁾。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 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。

2.4 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。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 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 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。

2.5 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。

[†] 慶應義塾大学大学院理工学研究科
Graduate School of Science and Technology, Keio University
[‡] 慶應義塾大学理工学部
Faculty of Science and Technology, Keio University

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	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

3. 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。

3.1 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。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)

抑制すべき発話箇所は 39カ所であった。

3.1.1 比較手法

本手法の有用性を検証するために、以下を用意した。

- 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 Δt 秒発話情報を蓄積し、その情報を基に切替える。

3.1.2 実験結果

表 2 に、既存手法と比較した一致率、切替数の結果を示す。これより、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 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

3.2 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。

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。

3.2.1 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。

3.2.2 実験結果

表 3 に問題の正解率の結果、表 4 にタイプ精度の結果を示す。サブワークである問題の正解率に関して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、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。

表 3 問題正解率

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 %

表 4 タイプ精度

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 %	

4. 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。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。

参考文献

- 1) 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.