LK-010

話者の注目喚起行動による机上作業映像の自動編集
— ユーザインタフェースの側面からの評価 —

An Editing based on Behaviors-for-Attention for Desktop Manipulation Videos
— Evaluation of the User Interface —

尾関 基行 † Motoyuki Ozeki
中村 裕一 ‡ Yuichi Nakamura
大田 友一 † Yuichi Ohta

1. まえがき

本稿では，机上作業プレゼンテーションの自動撮影システムにおいて，話者（出演者）が自らの行動によって映像を編集つつ作業するという枠組みについて，ユーザインタフェースの側面から検討した結果を述べる．

近年の映像メディアを取り巻く環境の進歩は目覚ましく，遠隔講義や遠隔会議，映像を用いた資料やマニュアルなど，一般の人々が身近に映像を制作し利用することに関心が高まっている [1]–[3]．このような背景から，我々は，料理や科学実験などの机上作業プレゼンテーションを対象とした自動撮影システムを構築してきた．

本システムでは，複数台の首振りカメラを自動制御し，それぞれの撮影対象に応じたカメラワークを用いて追跡撮影する [4]．撮影と同時に，指示動作など話者が視聴者の注目を集めるために行う行動（以下，注目喚起行動と呼ぶ）に基づいて，撮影されたショット群の中から完成映像として記録するものを一つ選択する [5]．この注目喚起行動に基づいた自動編集手法について，これまでに，テレビ番組映像の編集にも注目喚起行動が深く関連していること，また，話者の視点から編集された映像が視聴者にとっても満足できる映像となることを明らかにした．しかし本手法の特徴は，話者が注目喚起行動によって自ら映像を編集するという点であり，ユーザインタフェースの側面からの検討が不可欠である．

我々はこの問題について，被験者に本システムを使って映像編集しながら作業してもらい，その使用感をアンケート評価によって調べた．関連研究として，講義撮影システムによる被撮影者への影響についての報告 [6]-[8] があるが，被撮影者自身が編集に協力するという枠組みについての評価は行われていない．本研究では，撮影されているという受け手としての影響だけではなく，自らが編集しているという送り手としての影響も調べる．

本稿では，2. で撮影システムと自動編集手法の概要について説明し，3. で評価実験の目的と実験で比較する編集手法を挙げた後，4. で実験とその結果について述べる．

2. 撮影システムと自動編集手法の概要

2.1 撮影システムの構成

撮影システムの概要を図 1 に示す．本研究では，料理や工作，科学実験，組立て作業などの机の上で行う作業について，1 人の話者が 1〜3m 程度の範囲を移動しながら説明する場面を撮影対象とする．

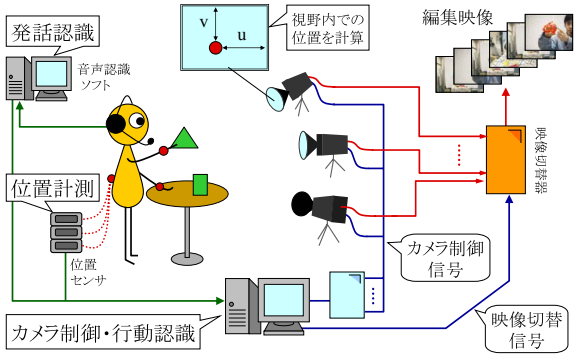

図 1: 撮影システムの概要

自動撮影モジュールでは，話者の両手首と腰の位置を位置センサで計測し，そのデータをもとに複数台の首振りカメラを制御する．各カメラは予め指定した一つの対象を常に追跡し撮影する．撮影したショットは映像切替器に入力し，自動編集モジュールより送られる切替信号によってオンラインで編集される．本稿の実験では 4 台のカメラを用いて，話者と作業領域を含んだミディアムショット[1]と，手元（右手・左手・両手の中間）のクロースアップショットを撮影する．

自動編集モジュールでは，話者の注目喚起行動を認識し，その結果に基づいて映像切替器に切替信号を送る．注目喚起行動の検出には，位置センサから得られる両手首と腰の位置データ，音声認識ソフトから得られる発話データ，自動撮影モジュールで計算されているカメラ視野内における手の位置データを利用する．注目喚起行動の検出方法とそれに基づいた編集ルールについては 2.3 で述べる．

2.2 自動編集に対する考え方

机上作業映像では，「話者と作業領域を含んだミディアムショットで作業の全体的な流れを掴ませつつ，手元や物体などのクロースアップショットで注目すべき箇所を強調する」という編集パターンが基本である．編集の自動化は，いつ・どのクロースアップショットに切り替え，いつミディアムショットに切り戻すかを決定する要素（編集トリガ）を定義し，これを検出もしくは算出することで実現できる．

これについて我々は，編集トリガとして「視聴者の注目を集めるために話者が行う行動（**注目喚起行動**）」に

†筑波大学大学院 システム情報工学研究科
‡京都大学 学術情報メディアセンター

[1] ミディアムショットとは人物の七分身まで含んだショット．本研究で用いるミディアムショットは，ショットの下部に作業机の一部を含んだもの．図 3 中に一例を示している．

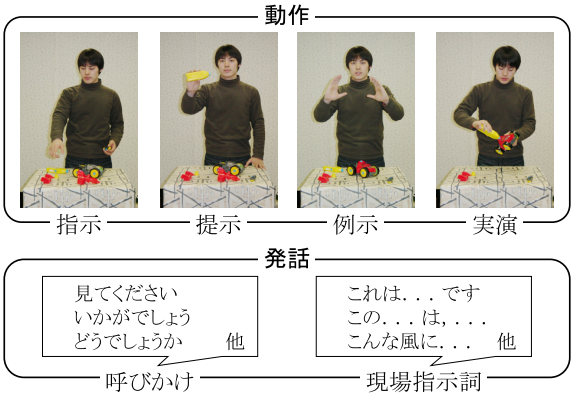

図 2: 注目喚起行動の例

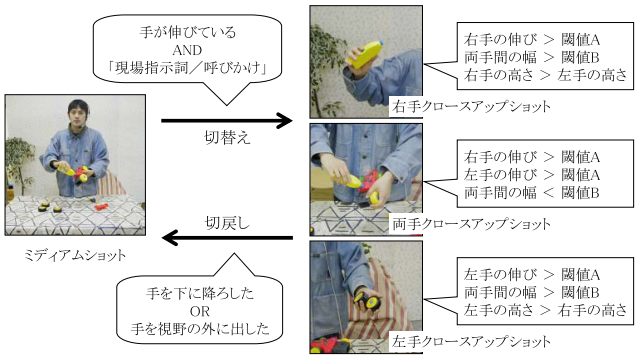

図 3: 注目喚起行動による自動編集ルール（閾値 A と閾値 B は経験的に決定）

着目する．注目喚起行動が起こるとそれが示唆する注目箇所を撮したクロースアップショットに切り替え，それから続く一連の作業が終わるとクロースアップショットからミディアムショットへ切り戻す．

注目喚起行動の例を図 2 に示す．これらの行動は，机上作業プレゼンテーションに頻繁に現れるもの，もしくは作業映像で重要となる物体や操作に対して注目を集めるものである．ただし本研究では，これらの行動を個別に検出して利用するのではなく，まとめて一つの編集トリガとして扱う．

2.3 注目喚起行動による自動編集

注目喚起行動の起こりとそれに続く一連の作業の終了を検出することで，編集を自動化することができる．検出手法は，話者が意図的に利用することから，以下の条件を満たすものが望ましい．

- 制約が少なく，話者への負担が少ないほうが良い．つまり，自然な行動を検出することができ，検出のために特別な行動をとる必要がないことが望ましい．
- 検出に失敗した時に話者が再試行できるほうが良い．つまり，話者が直感的に理解できる単純な検出手法であることが望ましい．

これに対して本研究では，話者の手の位置と発話を組み合わせることで，以下に述べるようにして注目喚起行動を検出する．システムは，これらの検出結果をもとに図 3 のルールに従って編集を行う．

注目喚起行動の検出: 手が体の前方に伸びている時に，注目喚起行動の “呼びかけ” もしくは “現場指示詞” の単語が発声されたら注目喚起行動が起こったとする．どのクロースアップショットに切り替えるかは，右手と左手の位置関係によって決定する．

一連の作業終了の検出: カメラの視野から手が出たこと，もしくは話者が手を下に降ろしたことで判断する．作業の終了時以外にもカメラの視野から手が出ることがあるが，それが作業の終了であるか途中であるかを識別することは今後の課題とする．

表 1: 論点 (β) のための評価項目

(a) 思い通りに切り替わりましたか？
(b) 作業は妨げられませんでしたか？
(c) 自然な行動（動き・発話）だと思いますか？
(d) 長時間（30 分以上）使い続けられますか？
(e) 使い慣れれば便利だと思いますか？

表 2: 論点 (α) のための評価項目

1. 話者による編集は無理がある
2. 話者による編集は可能ではある
3. 話者による編集は十分可能である
4. 他人に編集されるよりも良い

3. ユーザインタフェースの評価

3.1 評価実験の目的

本稿の評価実験では，次の二つの論点について検討する．

論点 (α): 話者が注目喚起行動によって自ら映像を編集しつつ作業するという枠組みが可能であるか

論点 (β): 注目喚起行動による自動編集が他の典型的な自動編集手法に比べて優れているか

これらについて調べるため，6 人の被験者に次節で述べる四つの編集手法を体験してもらう．論点 (β) に関しては，各編集手法を体験する毎に表 1 のアンケートに回答してもらい，同時に各編集手法の成功率を調べる．また，すべての撮影を終えた後で，使いたいと思う順番に編集手法に順位を付けてもらう．論点 (α) に関しては，すべての撮影を終えた後で表 2 のアンケートに回答してもらう．

3.2 比較する編集手法

本実験で比較する編集手法を以下に挙げる．自動編集手法 (A)〜(C) の比較により論点 (α) を，自動編集手法 (A) と手動編集手法 (D) の比較により論点 (β) を検討する．

(A) 注目喚起行動による自動編集
(B) カメラ指定キーワードによる自動編集
(C) 足スイッチと手の位置による自動編集
(D) 他人による手動編集

自動編集手法 (B)・(C) は以下の三つの基準により選んだ．編集手法 (A) が動作と発話の両方を用いた編集であるのに対し，編集手法 (B) は発話のみを用いたもの[2)]，編集手法 (C) は動作のみを用いたものといえる．

- 話者自身がショットを選択するもの
- 両手で作業しながら利用できるもの
- 切り替えるタイミングとどのショットを選ぶかの両方が指定できるもの

以下に (B)〜(D) の実現方法を挙げる．

(B) カメラ指定キーワードによる自動編集: 音声認識の結果から「右手・左手・両手」というキーワードを検出すると，それぞれ該当するクロースアップショットに切り替える．ミディアムショットに切り戻すには「全体」と発声する．

(C) 足スイッチと手の位置による自動編集: 足下に置かれたスイッチ（足スイッチ）を踏むと，その時の手の位置関係から，編集手法 (A) と同様にしてクロースアップショットに切り替える．ミディアムショットに切り戻すにはもう一度足スイッチを踏む．本実験では 2 箇所に足スイッチを置いた．

(D) 他人による手動編集: 筆者の一人がスタジオの様子をモニタで見ながら映像切替器のボタンを押すことでオンライン編集する．なお，筆者は机上作業プレゼンテーション撮影の研究に 4 年以上携わっており，切替えのタイミングなどは心得ている．

4. 実験

4.1 実験手順

本実験では，6 人の被験者に車の模型を組み立てる作業（2 分程度）を実演してもらった．プレゼンテーションの手順（シナリオ）は模造紙に大きく図示してスタジオの前方に掲示した．シナリオにはショットを切り替えるべき箇所が 8 箇所指定してあり，被験者はそれに従って編集を行う．編集結果は，前方に設置されたディスプレイで常に確認することができる．

実験の手順を以下に示す．

1) 「(D) 他人による手動編集」を通してオンライン編集を体験してもらう
2) 編集手法 (A)〜(C) について以下を繰り返す（編集手法の順番は被験者毎に変更）
    1. 編集手法について説明し，練習を兼ねて一度リハーサルを行う
    2. 本番を 2 回行う（編集成功率はこの 2 回で計算）
    3. 表 1 の評価項目について 5 件法で採点してもらう
3) 編集手法 (A)〜(D) について，使いたいと思う順番に順位を付けてもらう（総合順位は 1 位 4 点…4 位 1 点で計算）
4) 以上の撮影を踏まえて，注目喚起行動による自動編集について表 2 の中から該当するものを一つ選んでもらう

[2)] 編集手法 (B) では完成映像の音声にカメラ指定キーワードが含まれてしまうが，今回の実験の目的はユーザインタフェースの評価であるため，この問題については考えない．

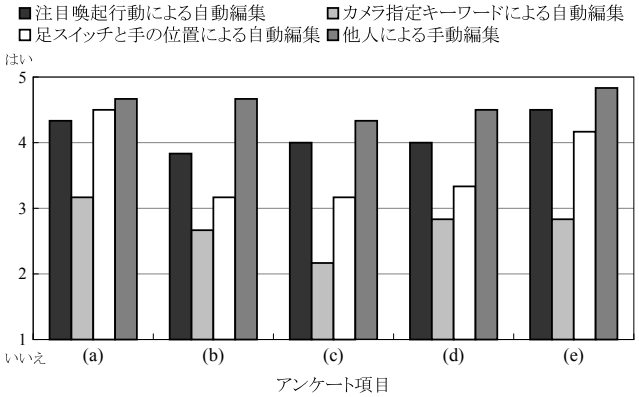

図 4: 5 件法による表 1 のアンケート評価の結果（グラフ縦軸の評価値は各項目について平均値を計算したもの）

表 3: 編集の成功率（単位 %）

	編集手法			
	(A)	(B)	(C)	(D)
切替え成功率	94.8	78.1	86.4	100.0
切替え・切戻し成功率	92.7	75.0	85.4	93.8

4.2 結果と考察

論点 (β) についての検討

まず，論点 (β)「他の典型的な編集手法に比べて提案手法が優れているか」について検討する．

アンケート評価の結果を図 4 に，編集の成功率を表 3 に示す．また，6 人の被験者による机上作業プレゼンテーションの一例を図 5 に示す．以下，自動編集手法 (A)〜(C) についてそれぞれ考察する．

(A) 注目喚起行動による自動編集: アンケート評価では全体的に高い評価が得られ，使いたい順位でも「(D) 他人による手動編集」と並んで 1 位となった．クロースアップショットへの切替えの失敗は，主に音声認識ミスに因るものであった．

(B) カメラ指定キーワードによる自動編集: 全体的に低い評価となり，使いたい順位でも最下位となった．プレゼンテーションとは関係のないキーワードを発声することに大きな違和感を感じた被験者が多く，音声認識ミスによる編集の失敗も比較的多かったことが原因と考えられる．

図 5: 6 人の被験者による机上作業プレゼンテーションの例

(C) 足スイッチと手の位置による自動編集: アンケート評価では,「(a) 思い通りに切り替わりましたか？」と「(e) 使い慣れれば便利だと思いますか？」において高い評価が得られた．しかし，移動する度に足スイッチの場所を確認しなくてはいけないことが煩わしいという意見もあった．

以上より，その他の典型的な自動編集手法に比べて提案手法の評価が全体的に高く，自然な行動として現れる動作と発話を組み合わせた手法の有効性が確認できた．

論点 (α) についての検討

次に，論点 (α)「話者自身が編集するという枠組みが可能であるか」について検討する．

まず図 4 の結果では，前述した通り,「(D) 他人による手動編集」には劣るものの提案手法も十分に高い評価が得られていることがわかる．更に，被験者が使いたいと思う順位では手動編集と並んで 1 位となった．編集の成功率も 90 %を越え，十分に実用的であると考えられる．

また，表 2 のアンケートでは,「2. 可能ではある」と答えた人が 2 人,「3. 十分可能である」と答えた人が 3 人,「4. 他人による編集より良い」と答えた人が 1 人であり,「1. 無理である」と答えた人はいなかった.「4. 他人による編集より良い」と答えた被験者に理由を尋ねると,「自分の切替えのタイミングが他人とは違うので，自分で選べたほうが良い」ということであった．

以上より，話者自身がプレゼンテーションを行いながら映像を編集することは，可能〜十分可能であることが確認できた．

問題点

問題点として，被験者ほぼ全員から「行動してから切り替わるまでに時間がかかる」ことが指摘された．現在のシステムでは，音声認識処理の影響により，行動してからショットが切り替わるまでに最長で 2 秒ほどかかる．これについては今後改善していく必要がある．

また,「編集のために決まった行動をとることに拘束感がある」という意見が多かった．一方で「システムに慣れたらもっと使いやすいと思う」という意見も多かったため，今後もユーザ評価実験を継続し，システムに慣れることで話者の感じる拘束感がどう変化するかを調べていきたい．

5. むすび

本稿では，机上作業プレゼンテーションの自動撮影システムにおいて，話者が注目喚起行動によって自ら映像を編集するという枠組みについて，そのユーザインタフェースの側面から検討した．本システムを使用した被験者のアンケートをまとめた結果，(α) 話者自身が編集するという枠組みが可能であること，(β) 他の典型的な編集手法に比べて提案手法が優れていることを確認した．

今後は，(1) 概要だけ示したシナリオを参考に自由にプレゼンテーションしてもらった場合と，(2) 聞き手とのインタラクション（質疑応答）を行う場合について，それぞれ今回と同様の評価実験を行う予定である．また，今回の実験では被験者全員が本システムを初めて使ったが，同じ被験者で実験を繰り返すことで，システムに慣れると評価が良くなるかどうかを調べていきたい．

参考文献

[1] 宮崎英明，亀田能成，美濃導彦，“複数のカメラを用いた複数ユーザに対する講義の実時間映像化法,” 信学論（D-II），Vol.J82-D-II，No.10，pp.1598-1605，1999.

[2] 先山卓朗，大野直樹，椋木雅之，池田克夫，“遠隔講義における講義状況に応じた送信映像選択,” 信学論（D-II），Vol.J84-D-II，No.2，pp.248-257，2001.

[3] 大西正輝，村上昌史，福永邦雄，“状況理解と映像評価に基づく講義の知的自動撮影,” 信学論（D-II），Vol.J85-D-II，No.4，pp.594-603，2002.

[4] 尾関基行，中村裕一，大田友一，“机上作業シーンの自動撮影のためのカメラワーク,” 信学論（D-II），Vol.J86-D-II，No.11，pp.1606-1617，2003.

[5] 尾関基行，中村裕一，大田友一，“注目喚起行動を用いた机上作業映像のための自動編集手法,” 画像の認識・理解シンポジウム，2004.（採録決定）

[6] 村上正行，田口真奈，溝上慎一，“日米間遠隔一斉講義における講師・受講生の評価変容の分析,” 日本教育工学会論文誌，Vol.25，No.3，pp.199-206，2001.

[7] 村上正行，西口敏司，亀田能成，美濃導彦，“講義自動撮影システムの導入に伴う講師・受講生への影響,” 信学技報 MVE，pp.29-32，2003.

[8] 望月俊男，他，“教室の授業と連携した e-Learning とその評価分析,” 教育システム情報学会誌，Vol.20，No.2，pp.132-142，2003.