会話シーンの複数視点からの自動撮影・編集

西崎　隆志† 　尾形　 涼† 　中村　裕一‡ 　亀田　能成† 　大田　友一†

† 筑波大学大学院システム情報工学研究科　〒 305–8573 茨城県つくば市天王台 1–1–1
‡ 京都大学学術情報メディアセンター　〒 606-8501 京都市左京区吉田本町
E-mail: †{tanishi,kameda,ohta}@image.esys.tsukuba.ac.jp, ‡yuichi@media.kyoto-u.ac.jp

あらまし　本論文では，映像コンテンツの取得を目的とした，会話シーンの複数視点からの自動撮影・編集を行うシステムについて紹介する．本システムの特徴は，撮影・編集に映画や TV の技法を取り入れたことにある．具体的には，構図（フレーミング）を維持した自動撮影と，様々な編集規則を導入できる自動編集を行う．撮影部分は，話者の位置を観測し，話者を捉えた典型的なショットを自動的に撮影するようカメラを制御する．得られた映像は，我々が提案してきた制約充足・最適化に基づく映像編集モデルを用いて編集される．そのために，撮影と同時に会話シーンの状況をインデックス情報として取得する．そして，このような自動撮影・編集を実際の会話シーンに対して適用し，質の良い映像コンテンツを取得できることを確認した．
キーワード　会話シーン，自動撮影，自動編集，映像コンテンツ

Automated Capturing and Editing on Multiple Views for Conversation Scenes

Takashi NISHIZAKI†, Ryo OGATA†, Yuichi NAKAMURA‡, Yoshinari KAMEDA†, and Yuichi OHTA†

† Systems and Information Engineering, University of Tsukuba, 305–8573 Japan
‡ Academic Center for Computing and Media Studies, Kyoto University, 606–8501 Japan
E-mail: †{tanishi,kameda,ohta}@image.esys.tsukuba.ac.jp, ‡yuichi@media.kyoto-u.ac.jp

Abstract　This paper introduces a novel system for automatically capturing conversation scenes on multiple views and for automatically editing obtained videos. The points of this system are camerawork and editing based on movies or TV culture: automated camera control with keeping good composition; automated editing with a variety of editing rules. The video capturing portion of the system detects positions of people, and control cameras for keeping an appropriate picture composition. The editing portion of the system applies our automated computational editing model that can introduce a variety kinds of editing rules. Our system detects speeches and other behaviors of people in conversation scenes during video capturing, which can be indices necessary as the input of the editing process. With those bases, we applied our system to conversation scenes, and check its efficacy for acquiring comprehensible and usefull videos.
Key words　conversation scene, video capturing, video editing, video contents

1. はじめに

映像によって会議の議事録や会話のメモを残したいという要望は大きいが，会議などの日常的な出来事に大きなコストはかけにくいため，専門家を雇って撮影・編集を行うことは一般的に難しい．また，固定カメラによる自動撮影を導入したとしても，得られた映像記録は見づらく，コミュニケーションの様子や場の雰囲気も分かりにくい．そのため，より質の良い映像を撮影・提示できる自動撮影・編集システムが必要となる．また，会話形式を用いて情報提示を行なうことの有効性も実証されてきており [1]，このような意味でも，会話を詳細に撮影・蓄積・提示する技術が必要とされている．

このような問題に対して，関連研究としては，ミーティングの記録システム [2]，机上作業におけるプレゼンテーションの撮影・

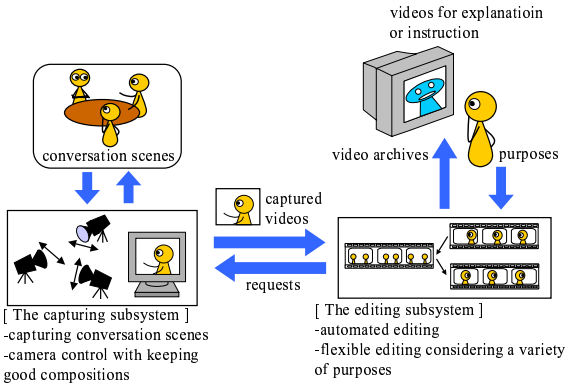


図 1　映像取得・提示の枠組み

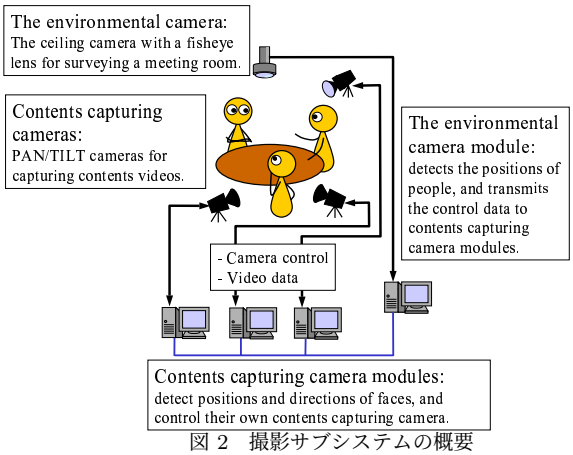


図 2　撮影サブシステムの概要

編集システム [3] [4]，講義の自動撮影システム [5] などがある．以上の研究に対して，本研究の特徴は質の良い映像を取得することを目指した点である．ここで言う “質の良い映像” とは，会話シーンの基本情報（例えば，誰が何を話したか）を伝えるだけではなく，シーンの細やかな雰囲気を伝えることができ，しかも，見ていて飽きない映像のことを指す．そのために我々は，映画や TV 番組で用いられている映像撮影・編集技法を取り入れ，撮影から編集までを一貫して自動的に行うシステムを構築してきた．

2.　会話シーンのための自動映像取得

2.1　自動撮影と自動編集の枠組み

本論文で提案する自動撮影・編集の枠組みを図 1 に示す．この枠組みでは，撮影部分（撮影サブシステム）で会話シーンを撮影し，同時にその状況を表すイベント生起情報を取得する．そして，得られた映像とイベント生起情報を入力データとして，編集部分（編集サブシステム）で映像を編集する．最終的に，イベント生起情報をインデックス情報とし，編集を施された映像とともに会話シーンの映像コンテンツとして蓄積する．本研究では，この枠組みに対して大きな自由度を与えているため，様々な設定を試すことができる．

2.2　自動撮影の設定

撮影サブシステムの概要を図 2 に示す．そのポイントは以下の二点である．

（1）最終的に視聴者にみせる映像を撮影するための “コンテンツ撮影カメラモジュール” と，人物位置を検出するための “環境カメラモジュール” とに役割を分ける．

（2）複数の視点から撮影するように設置したコンテンツ撮影カメラモジュール群を用いて，複数の構図で撮影する．

会話者の行動を制限すること無く，良い構図を維持しながら人物を追跡撮影し続けるためには，（1）の設定が必要となる．移動する人物を追跡するためには広範囲を追跡撮影しなくてはならない．しかし，構図やカメラワーク等の映像の質を適切に保つためには，カメラを頻繁かつ高速(注1)に制御することは好ましくない．また，様々な構図のショットを撮影するために（2）の設定が必要となる．

本システムでは，複数のコンテンツ撮影カメラによって得られた映像を編集サブシステムへの入力とする．この編集サブシステムで撮影された映像データだけではなく，そのシーンで起ったイベントの生起情報も必要となる．イベント生起情報の自動取得については，4. 節で述べる．

2.3　自動編集の設定

会話シーンの映像を撮影・編集する目的には，発話者や聴衆の詳細な様子を伝える，場の雰囲気や会話の盛り上がりを伝えるなど，様々な編集規則があるが，それらが排他的な要素を持っている場合が多い．この問題に対し，本研究の編集サブシステムでは，筆者らが提案した制約充足と最適化に基づいた映像編集モデル [6] を用いた．この編集モデルは，映像編集の規則やパターンを評価関数と制約によって表現し，可能なショットの組合わせの中で最も評価の良いものを求める．これにより，相反する可能性のある幅広い編集規則が利用可能となる．これまでにもイベント駆動型の編集モデルは提案されてきたが [10] [11]，イベントの生起時刻の微妙な違いなどが編集結果に大きな影響を及ぼしてしまう，アルゴリズム中に相反する要求を組み込むのが困難であるといった問題があった．本編集モデルは，その問題点を解決するための手法である．紙面の都合上，本論文では，この編集モデルの詳しい説明を省略するので，詳しくは文献 [6] を参照されたい．

3.　撮影サブシステムにおける処理

以下で撮影サブシステムについて詳しく述べる．

3.1　会話シーンにおける典型的な構図

本研究では複数人の会話シーンを対象としており，映画や TV 番組でよく用いられるショット構成を模倣する [7]．本研究では，それらのショットの基本を一人対一人，もしくは一人対二人とする．そして，それより人数が多い場合には，一人対二人の応用形として考える．より複雑な設定が必要となる状況への対処は今後の課題としたい．このような考え方に基づき，本研究では以下のような構図のショットを扱う．

クロースアップ/バストショット(注2)： 図 3 (a)(b) のように，人物の顔を画面に大きく捉えたショット．顔が正面を向いている

(注1)：カメラを頻繁かつ高速で制御すると，撮影される映像は不快でみづらいものになってしまう．カメラの不必要な PAN/TILT はできる限り避けなければならない．

(a) Close-up/bust shot (full frontal) (b) Close-up/bust shot (oblique or in profile) (c) Over the shoulder shot (d) Long shot capturing all speakers (e) Close-up shot for a table (f) Frontal shot of two speakers (g) Lateral shot of two speakers

図 3　典型的な構図

場合には図 3 (a) のように顔を画面の中心に捉え，顔が横を向いている場合には図 3 (b) のように顔の前に空間を空けるように捉える．

肩越しショット： 図 3 (c) のように，向かい合った相手の肩越しに人物の顔を撮影するショット．本撮影サブシステムでは，相手の肩を捉えながら，顔の前に空間を空けるように顔を捉える．

ロングショット(注3)： 図 3 (d) のように，カメラを引いて人物の全身より広い範囲を捉えたショット．本撮影サブシステムでは，向き合った二人，若しくは三人の人物を捉えたロングショットを扱う．

テーブルのクロースアップショット： 図 3 (e) のように，テーブル上の物体，特に話題の対象になっているものを捉えるショット．

二人以上の人物の前面ショット： 図 3 (f) のように，横に並んで座っている二人の人物を前方から捉えたショット．二人の人物の撮し方はクロースアップ／バストショットの撮し方に準ずる．

二人以上の人物の側面ショット： 図 3 (g) のように，横に並んで座っている二人の人物を横から捉えたショット．二人が向いている方向に絶えず空間を空けるように捉える．

3.2　コンテンツ撮影カメラモジュール

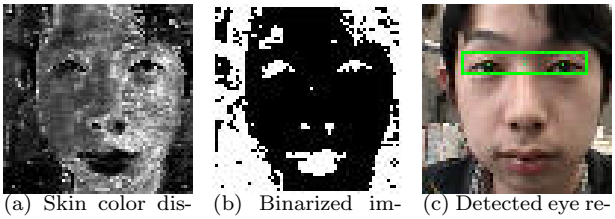
各コンテンツ撮影カメラモジュールは，次節で述べる環境カメラモジュールから大まかな人物の位置情報を取得しており，それを基にカメラの初期 pan/tilt 値を決定する．

顔を追跡撮影するためには顔検出が必要である．それに加えて，クロースアップ／バストショットにおいては，適切な構図を保つために顔の向きの計測も必要となる．本システムでは，Rowley らによる顔検出手法 [8] とテンプレートマッチングを組み合わせることにより，人物の顔の検出・追跡撮影を行っている．

また，小田らによる手法 [9] を基にして目領域を検出し，両目の重心と顔の重心を比較することによって，顔の向きを判別する．まず，図 4(a) に示すような，各画素が肌色にどれほど近いかを表す “肌色距離画像” を求める．次に，図 4(b) に示すような肌色距離画像を二値化したものから，目領域候補群を抽出する．そして，そこから評価関数を適用することによって二つの領域を選びだし，これを目の検出位置とする．この評価関数は，小田らによって提案されているものを本システムに合わせて若干修正して用いた．また，テンプレートマッチングによっても目の位置を追跡する．これら二つの方法で得られた目の位置の重み付け線形和を求め，カルマンフィルタによって平滑化したものを最終的な目の位置とする．得られた目領域を図 4(c) に示す．以上の処理で得られた両目の重心と顔の重心を比較し，顔の向きを判別する．このような手法により，顔を画面上の何処に捉えるべきかが計算され，そのために必要な pan/tilt 角が得られる．

(注2)：画面中での人物のみかけの大きさによって呼び名が変わる．画面中に人物の顔が大きく収まっている場合にはクロースアップショット，人物の胸の高さから上が収まっている場合にはバストショットと呼ぶ．

(注3)：厳密には，立っている人物のみかけの大きさを基準にした呼び名であるが，ここでは座っている人を基準として考えることにする．



(a) Skin color distance image (b) Binarized image (c) Detected eye regions

図 4　目領域検出の概要

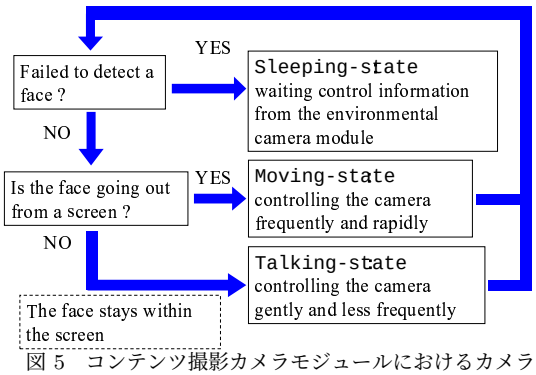


図 5　コンテンツ撮影カメラモジュールにおけるカメラ制御の流れ

本撮影サブシステムでは，顔の追跡状態を三つの状態に分類し，その状態に応じた制御を行う．これらの状態間を推移する大まかな流れを図 5 に，それら状態の説明を以下に示す．

会話状態 (Talking-state)： 顔があらかじめ定められた範囲内で動いている，若しくは停留している状態．カメラを低速・低頻度で制御する．

移動状態 (Moving-state)： 顔が画面端に存在し，顔の動きベクトルが画面の外側を向いている状態．カメラを高速・高頻度で制御する．

休眠状態 (Sleeping-state)： 顔検出に失敗し，追跡撮影が不可能となった状態．再び顔が画面内に現れるか，環境カメラモジュールからの指示があるまでは，何もせず待機する．

3.3　環境カメラモジュール

コンテンツ撮影カメラモジュールの撮影を補助するため，魚

眼レンズを取り付けた環境カメラを会議室内の天井に設置する．フレーム間差分とテンプレートマッチングの二つの手法により人物のおおよその位置を取得し，得られた二種類の位置の中点を最終的な人物の位置とする．テンプレートマッチングのテンプレートは，フレーム間差分によって得られた領域の面積が閾値以上である場合に更新される．

環境カメラモジュールはあらかじめ各コンテンツ撮影カメラモジュールがどのショットを撮影すべきかといった情報を持っている．そして，検出された人物の位置と各コンテンツ撮影カメラの位置・画角から，各コンテンツ撮影カメラが担当するショットを撮影できるか否かを判別し，可能である場合には，初期 pan/tilt 角を各々決定する．また，各コンテンツ撮影カメラモジュールは，撮影中，識別番号と撮影状態の情報を環境カメラモジュールに定期的に送信する．これにより，環境カメラモジュールがコンテンツ撮影カメラモジュールに初期的な指示を送ること，及びコンテンツ撮影カメラモジュールが休眠状態になった場合に，追跡撮影を再開するように指示することが可能になる．

4. イベント生起情報の自動取得

4.1 イベントの種類

会話シーンに対する映像編集では，シーン中で起きている出来事（イベント）に応じてショット切り替えを行う必要がある．本研究では，後述のものを自動で取得して利用する．なお，本編集モデルでは，映像の各ショットの各時刻における良さや全体的・部分的なショットの組み合わせを評価関数や制約を用いたスコアの加減算により評価し，最もスコアの高いものを編集結果として求める [6]．以下では実験で用いた評価関数（スコアの加算）とイベント生起との関係を併記した．

発話の有無: 編集モデルにおける映像セグメント単位（現在 1 秒に設定）で発話があるか否か．話者を撮す各種ショットのスコアを加算する．

物を指し示す指示詞の発話の有無: 主に,「これ」,「この」,「このように」等．現在は机上の物体を指し示すと仮定し，机上のクロースアップショットのスコアを加算する．

話題を切り替える接続詞の発話の有無: 主に,「ところで」,「話は変わって」等．シーンのロングショットのスコアを加算する．

頷き動作の有無: その映像セグメントで人物が頷いているか否か．相手の発話（主に問い）に対する肯定を意味すると仮定し，頷いている話者を対面者の肩越しに撮すショットのスコアを加算する．

会話シーンに対して妥当な編集結果が得られることは尾形らによって既に報告されているが [6]，本研究では，そのためのイベント生起情報を自動で取得しており，その点において，本研究が初めての試みとなっている．

4.2 イベント生起情報の取得

以下，前節で述べた自動取得されているイベントの取得方法について述べる．

発話情報は，市販の音声認識ソフトウェア (IBM 社 ViaVoice) を用いて検出する．

頷き動作は，映像中での顔の位置の変化に評価関数を適用することで検出する．この評価関数は実際の頷き動作のデータを基に設計した．

図 6 コンテンツ撮影カメラモジュールによる撮影結果

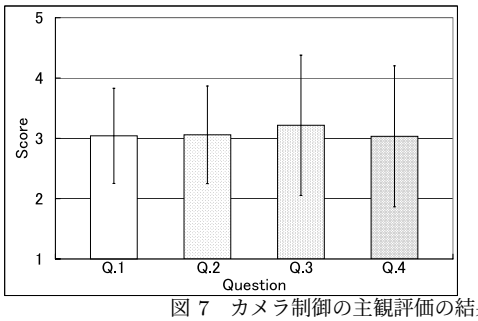


図 7 カメラ制御の主観評価の結果

5. 撮影結果とカメラワークの評価

コンテンツ撮影カメラモジュールにおけるカメラ制御の妥当性を評価するため，複数の会話シーンを撮影し，得られた映像に対して主観評価実験を行った．

5.1 撮 影 結 果

実験例として，二人／三人の人物が席について会話する複数のシーンの映像を各々約 70 秒間撮影した．映像の内容は，最初に二人／三人の人物が机に歩み寄り，机の前の椅子に一対一／一対二で向き合った状態で会話を開始するものである．

コンテンツ撮影カメラモジュールが撮影した会話シーンの映像の例を図 6 に示す．これまでの実験では，人物の顔を低速の pan/tilt 制御（最低速度は約 2.4 deg/sec）と低頻度の位置更新（最低頻度は約 5 sec に 1 回）で追跡撮影し続けることができ，コンテンツとして概ね満足のいく映像を取得できた．

しかし，人物の向き判別の精度が不十分であり，適切な構図を維持できない場合があったため，これを改善する必要がある．

5.2 カメラ制御の主観評価

本撮影サブシステムのカメラ制御に対する主観評価実験を行った．撮影した複数の映像を 30 人の被験者にみてもらい，以下に示す 5 段階評価のアンケートに答えてもらった．

- **Q.1「カメラを動かす速さは適切でしたか？」**
  “遅い” が１，“適切” が３，“速い” が５．
- **Q.2「カメラを動かす頻度は適切でしたか？」**
  “少ない” が１，“適切” が３，“多い” が５．
- **Q.3「良い構図で映されていましたか？」**
  “悪い” が１，“どちらでもない” が３，“良い” が５．

表 1　イベント生起情報の取得精度 1：頷き動作，キーワード検出 [%]

	サンプル数	検出数	誤検出数	Precision	Recall
頷き動作	80	62	6	91	78
指示詞	54	50	5	91	93
接続詞	52	48	0	100	92

表 2　イベント生起情報の取得精度 2：発話の有無 [%]

発話区間数	平均 Precision	平均 Recall
50	88	81

- **Q.4「カメラの動きは自然でしたか？」**
  “不自然” が 1，“どちらでもない” が 3，“自然” が 5．

評価実験の結果の平均を取ったものを図 7 に示す．どの質問に対する回答もおおよそ 3 になっている．Q.1 と Q.2 の質問では 3 が適度であるため，カメラワークの速さと頻度に関しては高評価が得られたといえる．これは，本撮影サブシステムのカメラ制御において，顔の追跡状態が会話状態であるとみなされた場合にカメラを極力低速・低頻度で制御する設定としたことによると考えられる．しかし，Q.3 と Q.4 の質問では 5 が最も良い評価であるので，構図維持とカメラ制御の自然さに改善の余地があるといえる．前節で述べたように，構図維持の評価の低さは顔の向き判別の精度が不十分であるためと考えられ，これを改善する必要がある．また，カメラ制御の自然さの評価が低いのは，カメラを一定速度で制御しているためと考えられる．この加速度を制御することにより，これを改善できると期待できる．

6. イベント生起情報の取得結果と評価

イベント生起情報の取得精度が編集結果に与える影響を調査するため，各イベント生起情報の取得精度を求め，その取得精度に基づいて作ったいくつかの編集映像を比較した．

6.1 イベント生起情報の取得精度

一人での発話，一人対一人の対話，一人対二人の対話の様子を撮影し，得られた複数の映像から各種イベント生起情報の取得精度を求めた．取得精度を表 1，表 2 に示す．表 1 の各イベントは短時間で生起が終了するものであるのに対し，表 2 の発話の有無は発話区間中絶えず生起し続けるものである．

表 1 をみると，各項目の値はいずれも高くなっているが，頷き動作の Recall 値がそれ以外に比べ低いことが分かる．これは，頷き動作の個人差が大きく，現在の頷き動作検出手法では十分に対応できていないためと考えられる．表 2 に示した発話の有無も Recall 値が低い結果となっている．音声認識エンジンは発話情報を一定時間記憶領域に溜め込む仕様となっており，このために認識の遅れ時間や認識漏れが生じてしまうためと考えられる．

次節に述べる実験で，これらの精度が会話シーンの編集に関して十分であるか否かを確認する．

6.2 映像編集結果の評価

本実験で用いた編集規則を表 3 にあげる．本実験の目的はイベント生起情報の取得精度が編集結果に与える影響を調査することであるため，編集規則には一般的と思われるものを用いた．評価関数・制約の働きや編集規則の詳細については, [6] を参照されたい．

図 8　評価実験：5 種類の編集例

以上の編集規則を用いて，イベント検出精度の異なる複数の編集結果を比較する主観評価実験を行った．被験者に見てもらった編集結果の例を図 8 に示す．以下，各編集設定について述べる．

- タイプ 1 は全てのイベント生起情報に誤検出・検出失敗が無いものである．
- タイプ 2 は全てのイベント生起情報が本システムの適合率・再現率で取得された状態である．
- タイプ 3 は発話情報を完全に取得でき，頷き動作が全く取得できなかった場合である．
- タイプ 4 は発話情報は全く取得できず，頷き動作は完全に取得できた場合である．
- タイプ 5 は全てのイベント生起情報が取得できなかったことを想定した例である．

本実験では，4 種の会話シーンに対して以上の 5 種類の編集結果を作成し，これを 30 名の被験者にみてもらい，以下に示す 5 段階評価のアンケートに答えてもらった．

- **Q.1「発話者の様子は分かりましたか？」**
  “分からない” が 1，“どちらでもない” が 3，“分かる” が 5．
- **Q.2「聞き手の様子は分かりましたが？」**
  “分かる” が 1，“どちらでもない” が 3，“分かる” が 5．
- **Q.3「シーン全体の位置関係は分かりましたか？」**
  “分かる” が 1，“どちらでもない” が 3，“分かる” が 5．
- **Q.4「会話の雰囲気は伝わりましたか？」**
  “分かる” が 1，“どちらでもない” が 3，“分かる” が 5．
- **Q.5「ショット切り替えのタイミングは自然でしたか？」**
  “不自然” が 1，“どちらでもない” が 3，“自然” が 5．
- **Q.6「ショット切り替えの頻度に対して，どのように感じましたか？」“退屈” が 1，“適切” が 3，“目まぐるしい” が 5．**
- **自由記述**

6.3 精度評価実験の結果

評価実験の結果の平均を図 9 に示す．いずれの質問への回答をみても，イベント取得に失敗の無いタイプ 1 と，イベントが本撮影システムの適合率・再現率で取得されたタイプ 2 との間に目立った差が無いことが分かる．このことから，ある程度のイベント取得の失敗が編集において許容されること，本撮影サブシステムのイベント検出精度が概ね満足できるものであること等が分かった．

表 3　設定した編集規則

前評価関数 $e_1(t)$：	人物が話していたら “クロースアップ／バストショット” のスコアを加点.
前評価関数 $e_2(t)$：	人物が話していたらその人物を撮す “二人以上の人物の前面ショット” のスコアを加点．三人の会話シーンにのみ適用.
前評価関数 $e_3(t)$：	物を指し示す指示語が発話されたら “テーブルのクロースアップショット” のスコアを加点.
前評価関数 $e_4(t)$：	人物が頷き動作をしていたらその人物の “肩越しショット” のスコアを加点.
前評価関数 $e_5(t)$：	話題を切り替える接続詞が発話されたら “ロングショット” のスコアを加点.
前評価関数 $e_6(t)$：	二人以上の人物が同時に発話していたら “ロングショット” のスコアを加点.
前評価関数 $e_7(t)$：	人物の話が一定時間 (本実験では 8 秒に設定) 以上続いたら聞き手の “クロースアップショット” のスコアを加点．二人の会話シーンにのみ適用.
前評価関数 $e_8(t)$：	人物の話が一定時間 (本実験では 8 秒に設定) 以上続いたら聞き手を撮す “二人以上の人物の側面ショット” のスコアを加点．三人の会話シーンにのみ適用.
後評価関数 $e_9(t)$：	ショットの組合せのスコアは各ショットのスコアを単純に加算したものとする.
制約 c_1：	一つのショットは最低でも 3 セグメント以上は連続しなくてはならない.
制約 c_2：	スコアが閾値以下のショットを解候補となるショットの組合せに含めてはならない.
制約 c_3：	状況説明のために “ロングショット” を映像の最初の 8 秒以内に必ず含まなくてはならない.

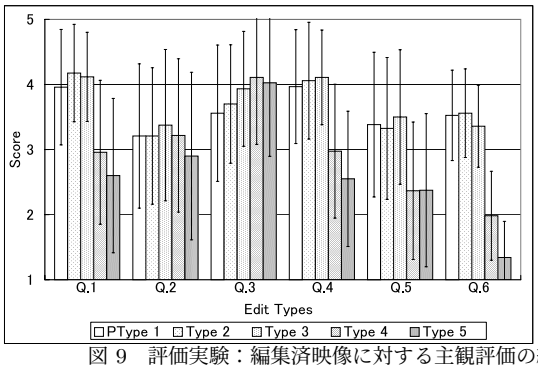


図 9　評価実験：編集済映像に対する主観評価の結果

質問に対する回答において特徴的であったものを以下に述べる．まず，Q.2 において全体的に大きな差がみられなかったことから，シーンのロングショットさえ提示されれば，聞き手の様子がある程度分かるためと考えられる．また，Q.3 において顕著な差が見られなかったことから，ロングショットが短時間で提示されてもシーンの空間的構成が分かると感じる被験者と，長時間提示し続けられないと分かりにくいと感じる被験者に二分されることが分かった．最後に，Q.6 において，本撮影システムの精度でイベント生起情報が取得された場合の評価が最適評価である 3 となっているため，ショット切り替えの頻度が妥当であるとみなせる．

7. ま と め

本論文では，映像コンテンツの取得を目的とした，会話シーンの複数視点からの自動撮影・編集を行うシステムについて紹介した．話者の位置を観測し，話者を捉えた典型的なショットを自動的に撮影するようカメラを制御する撮影サブシステムを構築し，制約充足・最適化に基づく編集モデル [6] に必要となるイベント生起情報の取得の自動化も行った．そして，実際の会話シーンを撮影した例を示し，これに対する主観評価実験の結果より，本撮影手法の有用性を示した．また，自動的に取得したイベント生起情報を基にした編集結果と正解情報を基にした編集結果を比較する主観評価実験を行うことにより，イベント生起情報の取得精度が概ね満足できることを示した．

本研究には多くの課題が残っている．まず，顔の向き判別の精度や頷き動作の精度など，撮影システムの精度を改善する必要がある．さらに，自動検出すべきイベントの種類，各評価関数・各制約のパラメータの設定方法を検討する必要がある．

文　献

[1] T. Nishida, N. Fujihara, S. Azechi, K. Sumi, and H. Yano, “Public opinion channel for communities in the information age,” New Generation Computing, Vol.17, No.4, pp.417–427, 1999.

[2] Y. Rui, A. Gupta, and J.J. Cadiz, “Viewing Meetings Captured by an Omni-Directional Camera,” Proc. of ACM’s Special Interest Group on Computer-Human Interaction 2001, pp.450–457, 2001.

[3] 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, No.11, pp.1606–1617, Nov, 2003.

[4] 尾関基行, 中村裕一, 大田友一, “注目喚起行動に基づいた机上作業映像の編集,” 信学論 D–II, Vol.J88, No.5, May, 2005.

[5] M. Murakami, S. Nishiguchi, Y. Kameda, and M. Minoh, “Effect on Lecturer and Students by Multimedia Lecture Archive System,” 4th International Conference on Information Technology Based Higher Education and Training, pp.377–380, 2003.

[6] 尾形涼, 中村裕一, 大田友一, “制約充足と最適化による映像編集モデル,” 信学論 D-II, Vol.J87, No.12, pp.2221–2230, Dec, 2004.

[7] S. D. Katz（著）, 津谷 祐司（訳）, “映画監督術 SHOT BY SHOT,” フィルムアート社, 1996.

[8] H. Rowley, S. Baluja, and T. Kanade, “Neural Network-Based Face Detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 20, number 1, pp.23–38, Jan, 1998.

[9] 小田竜也，森健策，末永康仁，“顔画像からの目領域推定に関する一手法,” 信学技報，PRMU98-197，pp.39–46，1999.

[10] Y. Atarashi, Y. Kameda, M. Mukunoki, K. Kakusho, M. Minoh, and K. Ikeda, “Controlling a Camera with Minimized Camera Motion Changes under the Constraint of a Planned Camera-work,” Workshop on Pattern Recognition and Understanding for Visual Information Media, in Cooperation with ACCV, 2002, pp.9–14, 2002.

[11] M.Onishi, T.Kagebayashi, and K.Fukunaga, “Production of Video Images by Computer Controlled Cameras and Its Application to TV Conference System,” Proc. of IEEE Conference on Computer Vision and Pattern Recognition, Vol.2, pp.131–137, 2001.