グループウェア 11－2
（1995．4．20）

ＴＶ映像表現の分析と遠隔ＴＶ会議システムへの応用

井上智雄　小林武文　岡田謙一　松下温
慶応義塾大学理工学部

遠隔TV会議に対するニーズは増大しているが、対面状況のような会議が可能になっているとは言い難い。我々は、ショットの転換や様々なカメラワークといった映像に加えられた操作は、映像メディアから与えられる一つの独立した記号であり、映像を扱う際にはこれらを考慮すべきであるという立場から、TV会議システムにおける映像表現に着目した。そして、人が一番身近に体験している映像はTV映像であることから、TVの討論番組における映像表現を分析した。TVの討論番組におけるショットを効果的に分類し、一つのショットの持続時間および分類した場面間の遷移確率から、映像のパターンを見いだした。さらに、遠隔TV会議システムに応用するために、分析結果を元にした映像制御アルゴリズムを作成し、実際にTV会議システムを構築した。

Analysis of Pictures on TV Programs and Its Application to a Videoconferencing System

Tomoo Inoue, Takefumi Kobayashi, Ken-ichi Okada and Yutaka Matsushita
Keio University
3-14-1, Hiyoshi, Kohoku-ku, Yokohama 223 Japan
e-mail: inoue@myo.inst.keio.ac.jp

The need for videoconferencing has been increasing these days, but it does not seem that the videoconferencing systems currently in use are satisfactory. Their pictures are often tedious and boring. To solve this problem, we have been paying attention to presentation of pictures on TV, especially on debate programs. Thus several TV debate programs have been selected and analyzed. First, an appropriate classification of the shots has been done. Then duration of each shots and transition probability among classes have been examined. From these, rules for controlling pictures have been obtained. After that, we have made a picture controlling algorithm based on the rules, and have applied it to a videoconferencing system.

1　はじめに

会議はオフィスの活動において非常に重要な位置を占めており、ここ当分なくなることは無いといえる。そして、地理的に遠く離れた人同士の会議では金銭的、時間的コストが大変高くつくために遠隔TV会議に対するニーズが増大している。しかし現在のところこれによって対面状況のような会議が可能になっているとは言い難い。遠隔地の人同士のコミュニケーションにおいては映像情報が重要な役割を持つと言われており、確かに従来の遠隔TV会議システムでも映像は使用されているが、その映像が十分に活用されていない点に現在のTV会議システムの持つ問題点の一つがあると考えられる。

そこで、映像メディアとして先輩である映画やTVに学び、それらの映像に含まれる被写体それ自体以外の映像情報、すなわち映像の演出をTV会議システムに取り入れることを考えた。具体的には、ＴＶ会議への応用を考えて、テレビの討論番組について調査した[1]。まず、画面上に映っている人物の役割（話者、話の相手、第三者）を軸として場面を分類し、その分類に基づいて一つの場面の持続時間や、場面の時間遷移などを調べた。

さらにその結果を元にした映像演出アルゴリズムに従って、コンピュータが半自動的にビデオカメラを制御するようなＴＶ会議システムを試作した。

以下、本稿ではこれまでのTV会議に使われてきた映像情報の問題点を指摘し、その一つの解決策を提示し、この考えに基づいて行ったTV番組の分析について述べ、分析結果のＴＶ会議への応用について述べる。

2　TV会議映像の問題点

これまで研究されてきた多人数のTV会議システムの多くは、一つの画面に参加人数分のウィンドウが常に表示されていたり、あるいは人数分の画面があったりした。また多くの場合、他地点の映像は画面近くに置かれた固定カメラによるものであった。

まず前者について、コンピュータ作業では複数のウィンドウを同時に開いているほうが一般的であるが、これはユーザがそれぞれのウィンドウを自分でコントロールできるからであって、TV会議の場合はコンピュータ上のウィンドウであっても、それらの動きをユーザは制御できないため、その状況は映画やTVと同じものになる。そして映画では複数の画面が同時に存在するのは珍しい。なぜなら、観客がそれらを制御できないので注意を分割することになる。映画の画面は、次々と変わるので、観客は見ていない画面で起こっている大切な動きを見逃しているのではないかと心配することになる[2]。従って、TV会議システムにおいても多くのウィンドウを同時に使用しない方が良いと考えられる。

次に後者について、固定カメラによる単調な映像を飽きずに見続けられる人は少ないだろう。これは、映像固有の特性を考慮していないのである。映像には、対象のサイズを自由に変化させられ、また様々な視点の構図を不連続に提示しうるという特徴がある。さらに実時間の時間経過に関わらない表現も可能であり、映像を認識するということは我々の実寸大の生活世界を認識することとは異質なものだといえる[3]。したがって映像を考慮する場合にはショットの転換や様々なカメラワークといった映像に加えられた操作を考慮するべきであるが、これまで、会議に参加している人物が画面にどのように映されるべきかについては考えられていないようである。

これらの事柄を踏まえてTV番組を考えると、TV番組ではたいてい映っている画面は一つであり、その中で様々な映像操作がされている。例えばTV中継では、カメラは「見るべきもの」にフォーカスする。スポーツ中継では、カメラは「ボール」「選手」を追い、観客席で騒動が起きれば、テレビカメラはそちらを向く。これはそばにいる観客の視線と同じである。つまり、演出、視点の設定を行っているのである。討論番組でも、これは当てはまる。そこで、TV会議のための適切な映像表現を得るために、内容が会議に近いものとして討論番組についての調査分析を行った。

3　TV討論番組の分析

調査対象としたのは以下の番組で、全部で3600場面余り、時間にして10時間余りである。

・テレビ朝日「朝まで生テレビ」
・フジテレビ「報道2001」

・NHK「日曜討論」
・テレビ朝日「サンデープロジェクト」田原総一朗コーナー

まず、画面上に映される人物の役割とそのショットの種類から次のような分類法を考えた。これに従い、人物の映っているすべての場面を分類する事ができた。

1．話者単独場面：話者が一人だけ映っているもの
2．話者と周囲を含む場面：話者と話者の周囲の人物が映っているもの
3．話者と相手を含む場面：話者と話者が話しかけている人物が映っているもの
4．相手単独場面：話者が話しかけている人物が一人だけ映っているもの
5．相手と周囲を含む場面：話者が話しかけている人物とその周囲の人物が映っているもの
6．第3者単独場面：話者でも相手でもない人物が一人だけ映っているもの
7．第3者複数場面：話者でも相手でもない人物が複数映っているもの
8．全体場面：参加者全員が映っているもの

そして、一つの場面の持続時間を秒単位で計り、その分布を調べた（図1）。また場面が切り替わってゆくときに、分類中のどの場面からどの場面へと切り替わってゆくかという遷移状態を遷移確率の面から調べた。ここで、場面の遷移について一つ考慮すべきことがある。場面が遷移する時の主なきっかけとしては、話者の交代があるが、その他に、視聴者の関心を持続させる目的で行われていると考えられる場面の遷移がある。これら二つは性質の異なるものであるので、区別して考えた（表1，表2）。それに従って、ショットの持続時間については話者交代によって次のショットに遷移することになったショット、すなわち話者交代時直前のショットを除いたものについても集計した。

まず図1から全体として、持続時間の短いものほど頻度が多く、長いものになるにつれてその頻度

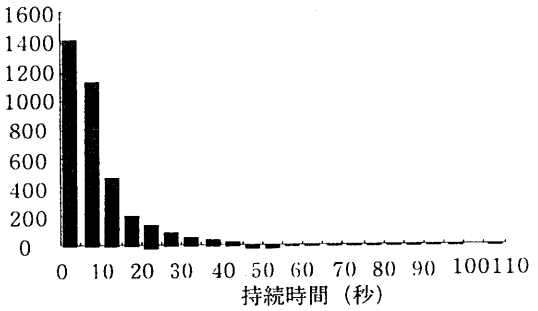

図1　ショットの持続時間度数分布

は少なくなっていくことがわかった。これを分類別に調べると、「話者」の入っている場面は持続時間が比較的長いものもありばらつきが大きいことがわかった。それに比べて「話者」の映らない場面は平均して5秒程度と短く、またばらつきが小さかった。これらは視聴者の関心を持続させる目的が大きい場面で、視覚的な「間」となるような場面であると考えられる。なお、話者交代による影響を除いたものについても全体の傾向に違いは見られなかった。ただし、分類別では、「話者単独場面」に関して持続時間の短いショットが減少した。これは、ある人物が話者であるときに、別の人物が割り込んで話を始めるといった状況で、話者の交代に合わせて

	話者単独	話者を含む	話者と相手	相手単独	相手を含む	第三者単独	第三者複数	全体
話者単独	55.60%	14.08%	10.47%	1.99%	2.17%	2.89%	3.43%	9.39%
話者を含む	57.73%	22.68%	8.25%	3.09%	1.03%	4.12%	1.03%	2.06%
話者と相手	75.90%	9.64%	4.82%	3.61%	2.41%	0.00%	2.41%	1.20%
相手単独	52.00%	28.00%	8.00%	8.00%	0.00%	4.00%	0.00%	0.00%
相手を含む	83.33%	8.33%	0.00%	8.33%	0.00%	0.00%	0.00%	0.00%
第三者単独	45.24%	26.19%	9.52%	4.76%	2.38%	2.38%	2.38%	7.14%
第三者複数	44.83%	31.03%	6.90%	6.90%	0.00%	0.00%	6.90%	3.45%
全体	82.22%	13.33%	0.00%	0.00%	0.00%	0.00%	0.00%	4.44%

表1　話者交代時の場面遷移確率（行から列に遷移）

	話者単独	話者を含む	話者と相手	相手単独	相手を含む	第三者単独	第三者複数	全体
話者単独	6.80%	18.74%	12.82%	13.11%	4.56%	24.76%	14.37%	4.85%
話者を含む	54.66%	6.30%	3.27%	9.07%	2.77%	13.10%	9.07%	1.76%
話者と相手	65.02%	6.90%	3.45%	12.81%	0.99%	4.93%	3.94%	1.97%
相手単独	68.62%	7.45%	12.23%	3.72%	2.66%	4.26%	0.00%	1.06%
相手を含む	45.95%	32.43%	12.16%	4.05%	1.35%	1.35%	2.70%	0.00%
第三者単独	63.34%	10.22%	3.24%	2.24%	0.25%	14.21%	3.74%	2.74%
第三者複数	62.07%	18.10%	2.59%	2.16%	0.00%	6.03%	6.90%	2.16%
全体	66.29%	11.24%	8.99%	7.87%	3.37%	1.12%	1.12%	0.00%

表2　話者交代時以外の場面遷移確率（行から列に遷移）

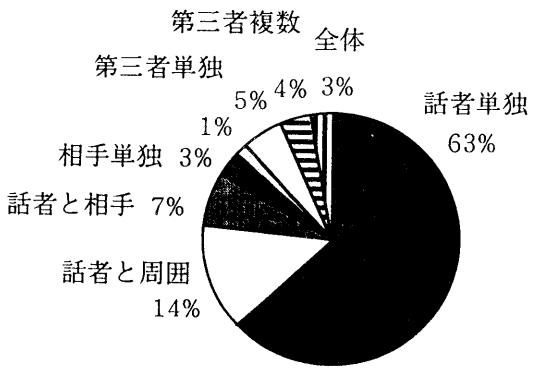

図2　全資料に占める各分類の時間の割合

ショットが切り替わることの影響が取り除かれたためであると考えられる。

次に、場面の遷移確率から遷移の中心は「話者」の映る場面であることがわかった。また、話者交代時には8割以上が話者の映る場面に遷移するのに比べて、それ以外の時には話者の映っている場面からは比較的他の場面に遷移することが多く、その後話者の映る場面に戻るパターンがよく見られた。さらに、全資料時間に占める分類別時間を調べたところ（図2）、「話者単独場面」が6割以上あり、「話者と周囲を含む場面」「話者と相手を含む場面」も合わせると8割以上であった。

これらのことから、一般に討論番組の映像は話者を中心として展開し、時折視聴者の関心を持続させるための比較的短い話者以外の場面に映り、再び話者に返るという流れを持っていることがわかった。

4　TV会議への応用 －映像演出アルゴリズム－

TV討論番組の分析結果を元にして映像演出アルゴリズムを作成し、これに従って、コンピュータが半自動的にビデオカメラを制御するようなTV会議システムを試作した。システムは1地点に複数の参加者（本システムでは4名）がいるような互いに離れた2サイト間の会議を想定した。

まず、作成したシステムの構成を図3に示す。ビデオカメラは、パンやズームをコンピュータからの信号で制御できる。参加者は普通のテレビ会議のように、遠隔地の相手と音声と映像のリンクによってCRTモニタを見ながら会話をす

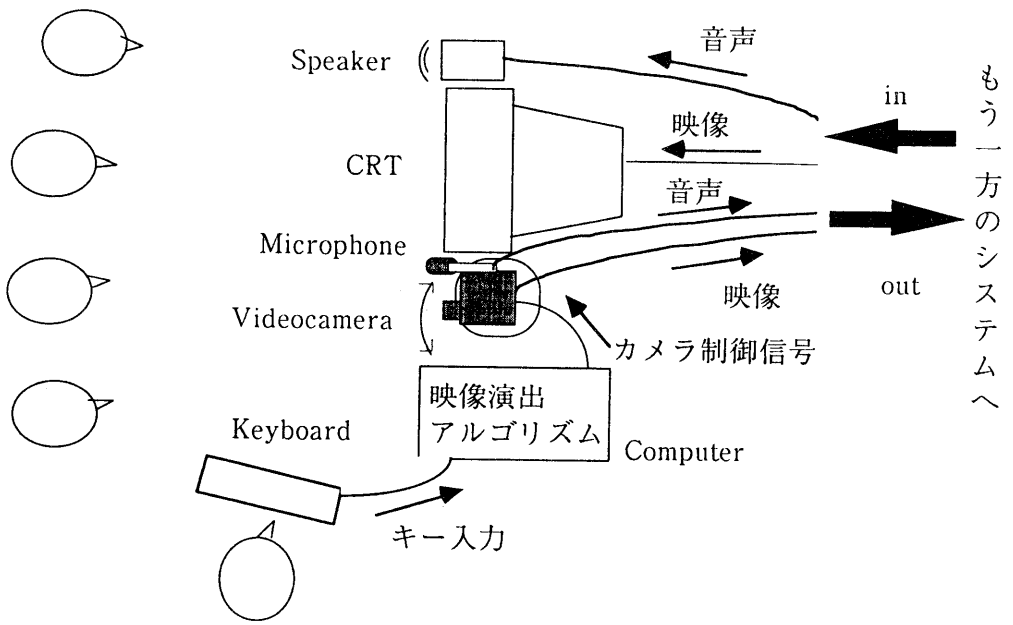

図3　システムの構成

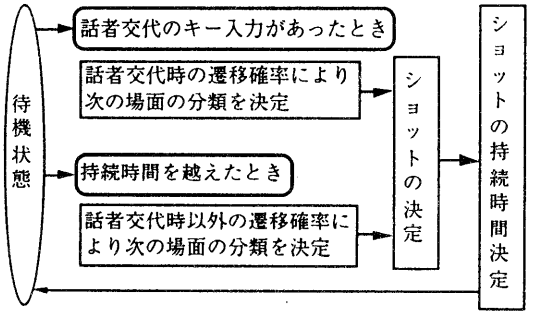

図4　映像演出アルゴリズム

ることができる。本システムは、会議の参加者とは別に一人オペレータを使用する。オペレータは、その時点での話者をコンピュータに指示するという役割だけを担う。これにより話者が誰なのかという点も考慮して総合的なカメラの制御が映像演出アルゴリズムによって行われる。つまり、オペレータは話者を指示するが、カメラは必ずしも話者に向けられるわけではなく、周囲の人物も一緒に映したり、全体をとらえたりといったことが自動的に行われるのである。

映像演出アルゴリズムは、TV番組の分析にから得られた、分類場面間の遷移確率と場面ごとのショットの持続時間の確率分布をもとにして作成した。まず、映像演出アルゴリズムの流れを図4に示す。

4.1　遷移確率

番組分析においては、会話の「相手」を考慮して場面を8種類に分類した。しかし、これをシステムに適応するときには、TV番組とTV会議の違いを考えねばならない。TV番組は一方向的なものであるのに対し、TV会議は双方向的なものである。つまりTV番組においては、「話者」も「相手」も基本的にスタジオという同一の場所にいて、それらの映像は一方的に視聴者に送られるのだが、TV会議では、「相手」は話者のいる側の人物にも、別サイドの側の人物にもなりうる。また、TV番組の分析結果から、場面の遷移においては「話者」が特に重要であることがわかった。これらのことから、まず話者を重視したシステムを試作し、そのための映像演出アルゴリズムも、TV番組分析結果のうち「話者」の映る場面に注目して作成した。

具体的には、TVの全場面のうちの、「話者単独場面」、「話者と周囲を含む場面」、「第3者単独場面」、「第3者複数場面」、「全景場面」だけを利用して、新たに遷移確率行列を構成した（表3，表4）。

4.2　ショットの持続時間

ショットの持続時間については、TV分析で得た、話者交代時直前のショットを除いたものをもとにしている。なぜなら、話者の交代を指示するのはオペレータであり、アルゴリズムに必要なのはそれ以外の場合の映像の変化を作ることだからである。場面ごとのショットの持続時間の度数を5秒単位で求め、その度数に応じた確率でそれぞれの中央値の時間だけショットが持続するものとした（表5）。

4.3　実際のショットの種類

現在本システムでは、1サイトのビデオカメラは一台である。また、使用したカメラの操作はパン（左右方向の回転角度）と

表3　映像演出アルゴリズムの場面遷移確率（話者交代時）

	話者単独	話者と周囲	第三者単独	第三者複数	全体
話者単独	65.12%	16.49%	3.38%	4.02%	10.99%
話者と周囲	65.88%	25.88%	4.71%	1.18%	2.35%
第三者単独	54.29%	31.43%	2.86%	2.86%	8.57%
第三者複数	52.00%	36.00%	0%	8.00%	4.00%
全体	82.22%	13.33%	0%	0%	4.44%

表4　映像演出アルゴリズムの場面遷移確率（話者交代時以外）

	話者単独	話者と周囲	第三者単独	第三者複数	全体
話者単独	9.78%	26.96%	35.61%	20.67%	6.98%
話者と周囲	64.39%	7.42%	15.43%	10.68%	2.08%
第三者単独	67.20%	10.85%	15.08%	3.97%	2.91%
第三者複数	65.16%	19.00%	6.33%	7.24%	2.26%
全体	83.10%	14.08%	1.41%	1.41%	0%

表5　ショットの持続時間確率

場面 / 持続時間(秒)	話者単独	話者と周囲	第三者単独	第三者複数	全体
2.5	17.27%	32.22%	77.33%	63.32%	43.81%
7.5	27.45%	44.39%	20.65%	31.00%	38.10%
12.5	20.27%	13.13%	1.76%	3.93%	13.33%
17.5	12.55%	4.06%	0.25%	0.87%	3.81%
22.5	7.36%	3.10%		0.87%	0.95%
27.5	5.45%	1.43%			
32.5	3.09%	0.48%			
37.5	2.27%	0.72%			
42.5	1.45%	0.24%			
47.5	0.55%				
52.5	0.73%				
57.5	0.73%				
62.5	0.36%				
67.5					
72.5					
77.5	0.09%				
82.5	0.27%				
87.5					
92.5					
97.5		0.24%			
102.5					
107.5	0.09%				

ズームのみである。4人の場合、これで10種類のショットがとられる。実際には、遷移する場面の分類が決定された後、その時点の話者との関係から適当なショットが選ばれる。例えば「話者と周囲」の場合、画面に映るのは二人か三人か、また右の人か左の人かあるいは両方かなどがあり得る。このように、一つの場面分類にいくつかのショットがある場合にはそのうちの一つを等確率で選ぶ。

また、実際のTV番組において映す対象が同一でショットを変える操作としては、ズームの他に複数のカメラを切り替えることによって同じ人物を違うアングルで映す操作が一般的に見られるが、本システムでは次の場面が全く同じ対象を映すことになった場合には、ズームによって異なるショットとなるようにしている。

4.4　話者側の映像

ここまでは、相手側のサイトに話者がいる場合について述べてきたが、話者側の映像については触れていない。つまり、話者がいない側の状況をどのように映すかという問題がある。これはTV番組とTV会議の違いに起因する問題である。TV番組はTVの中に必ず話者がおり、第三者としての視聴者に対して映像を提供しているので、TV番組の分析からは、本質的に話者側の映像についての知識は得ることができない。そこで、これらについては実験的に適切な方法を見つけてゆく必要がある。

5　むすび

これまでにTV討論番組の分析から、典型的な映像表現のパターンを得ることができた。そして、これらを元にして自動的に映像の演出を行うアルゴリズムを作成し、TV会議への適用を試みた。現段階では、話者中心というもっとも本質的と思われる部分だけを取り出しており、またシステム自体も簡単なものであるが、変化のある映像をTV会議に持ち込むことになるので、次はその効果を調べたい。また、話者側の映像についても考えなければならない。なお、本システムではショットの持続時間は度数分布の中央値といういくつかの固定時間から選ばれるが、連続的な確率分布の方が望ましいだろう。さらに今後は、カメラの台数を増やし、一層効果的な映像の演出を可能にしたい。

参考文献

[1] 黒須正明：リアルタイム通信における画像表現法の考察，平成5年度情報技術標準化センター成果報告会, pp.33-41, 1994.

[2] Clanton, C., Young, E. : Film Craft in User Interface Design, SIGSHI'94 Tutorial Notes, 1994.

[3] 中島義明, 井上雅勝：映像の心理学, 大阪大学人間科学部紀要, Vol.19, pp.1-26, 1993.