

## 文殊の知恵システムにおける視線情報伝達に関する一検討

坂本 信樹, クスタルト ウイドヨ, 青木 輝勝, 安田 浩

東京大学先端科学技術研究センター

〒153-8904 東京都目黒区駒場 4-6-1

E-mail: sakamoto@mpeg.rcast.u-tokyo.ac.jp

あらまし 本稿では、実物大映像・視線一致等を実現した高臨場感型ビデオ会議システム「文殊の知恵システム」において、参加者の頭部の動きにも対応して自由な視点位置からの視線一致を保持するため、どのような映像を提示すべきかを確認する実験を行った。その結果、観察者の視点に関わらず、参加者が常にカメラを見つめるような状況が最も望ましいことが分かり、この結果を利用してカメラの移動とディスプレイへの映像表示の方法について検討した。また、ディスプレイ上の共有空間におけるウィンドウの大きさを変化させ、対話相手の注目しているウィンドウをどの程度まで理解できるかを実験で確認し、共有空間設計への適用について考察した。

キーワード ビデオ会議システム, 視線一致, 共有空間

## A Study on the Gaze Communication for MONJUnoCHIE System

Nobuki SAKAMOTO, Kustarto WIDOYO, Terumasa AOKI, and Hiroshi YASUDA

Research Center for Advanced Science and Technology, The University of Tokyo

4-6-1 Komaba Meguro-ku, Tokyo 153-8904, JAPAN.

E-mail: sakamoto@mpeg.rcast.u-tokyo.ac.jp

**Abstract** This report describes the Gaze Communication for 'MONJUnoCHIE' videoconference system, which realizes teleconferences with eye-contact, life-size image and proximity to displayed persons. At first we examined what image should be displayed to support eye-contact from any point of view, and we found out that attendees should keep on looking at the camera behind the display independently of the observer's point of view. Based on the result we discussed some image displaying methods. Secondly we examined the accuracy of the recognition of windows in the displayed shared space between 2 remote sites. And we discussed applying the result to MONJUnoCHIE system.

**key words** videoconference, eye-contact, shared space

## 1. はじめに

近年、高速大容量通信ネットワーク技術及び画像処理技術など基盤技術の発達によって各種の新サービス開発が盛んになってきており、中でもテレワーキング（遠隔地間コラボレーション、テレビ会議など）アプリケーションに大きな期待が寄せられている。

筆者らはこれまで文献[1]にて実物大表示・視線一致等により高臨場感を実現したビデオ会議システム「文殊の知恵システム」を提案している。その中でも視線一致は、対話者が複数存在する場合に相手を指定するだけでなく、特に創造会議等において相手の意図を読み取る上で重要な情報である。また協同作業を行っている対話相手の注目点を理解することは、相手が自分との協同作業を正しく行っているかを確認するために有用である。よって視線情報の伝達と適切な理解は、ビデオ会議システムにおいて現実の会議での対面状況と同様に実現されるのが望ましい。

そこで本稿では視線情報の伝達と適切な理解の支援を目的とした2つの実験を行い、そこで得た知見を提案システムの設計へ適用することを考える。具体的には、現実の対面状況と同様に1サイトに一人の参加者がいる場合を考え、参加者の頭部の動きに対応し自由な視点位置で相手との視線一致を感じられる手法に関して考察を行う。次に、システム利用者のディスプレイ上での注目点が相手サイトにおいてどの程度の精度で理解されているかを確認する実験を行い、ディスプレイ上に表示する協同作業空間の大きさ・配置について検討し、今後のシステム設計への適用を考える。

本稿ではまず第2章で文殊の知恵システムの概要と特徴について述べる。第3章では、現システムで問題となる頭部位置変化による視線の不一致について述べ、参加者の頭部位置が移動した場合に視線一致を保持するための映像提示手法を確認する実験を行い、具体的な実装について検討する。第4章ではシステム利用者のディスプレイ上での注目点が、相手サイトにおいてどの程度の精度で理解されているかを確認する実験を行い、協同作業空間の設計の検討を行う。最後に第5章で本稿のまとめと今後の検討課題について述べる。

## 2. 文殊の知恵システムの概要

### 2.1. 従来のマルチメディア会議システムの問題点

従来のビデオ会議システムは、PCを用いたもの[2]と大型スクリーンを用いたもの[3]に大別できる。PCを用

いたものでは通常、各参加者の顔画像や資料データがウィンドウ表示される。大型スクリーンを用いたシステムでは資料提示用に計算機を導入し相手映像と会議資料を別のディスプレイに写しているものが多い。これらの方法では、対話相手がどの資料を見ているかという視線情報が直感的に得られない。相手がどこを見ているかということは、相手が何を考えているかを推測する大きな手助けとなるため、特に相手との駆け引きが活発に行われる意志決定や創造型の会議において、上述の問題は会議進行上の大きな妨げとなりうる。

### 2.2. 文殊の知恵システムでの視線一致の実現

文殊の知恵システムで使用する40インチ(縦60cm横80cm)のディスプレイは、特定の入射角からの映像のみを結像し、反対側からの光は直線的に透過する性質をもっている。これを図1の様に配置することで、ディスプレイ上の対話相手の顔映像を見つめることが実質的にはその裏にあるビデオカメラを見つめることとなり視線が一致する。

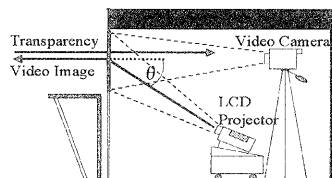


図1: 文殊の知恵システム端末の構成

### 2.3. 文殊の知恵システムでの共有資料提示手法

文殊の知恵システムでは想定される資料として、お互いが紙に印刷して所有する資料と、一方だけが所有する電子的データ(文書ファイルなど)が挙げられるが、前者はカメラによって自サイトの机上映像が相手サイトに送られるようにする(機のシームレス化)ことで相手が相手机上の紙資料に着目していることが自サイトから見て分かるようになる。しかし解像度の問題から紙の存在は分かるが書いてある内容までは分からないため、対話者間で共有は出来ていない。また電子的データについては、一方が所有する電子的データを相手に提示する場合、各資料をディスプレイ上にウィンドウ表示する。相手映像と資料ウィンドウを同じスクリーン上に映し、各サイトから見た資料提示位置を、ガラス板を挟んで向かい合ったような配置に調整することとする。これにより相手がどの資料ウィンドウに注目しているかがその視線方向から直感的に理解できるようになり対話者間で資料の共有が可能となると考えられる。

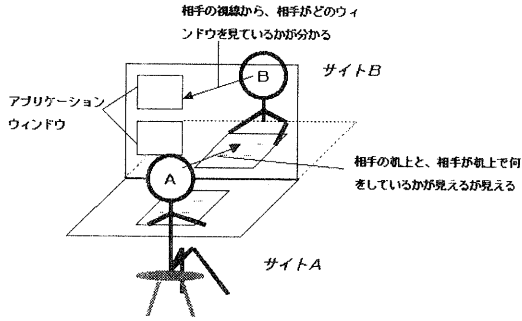


図2：資料提示の概略図

### 3. 自由な視点からの視線一致を保持する手法

#### 3.1. 文殊の知恵システムにおける視線一致の現状

2.2 で説明した視線一致実現方式ではカメラ位置が固定のため、参加者がカメラの正面に座っている状態ではうまく視線一致を支援できるが、参加者の頭部が上下左右に移動すると斜め方向から相手の顔映像を見つめることになり、視線延長線上にカメラはないため視線が一致しない。その反面、視線一致を感じる許容限がある程度存在すること[4]や、カメラと頭部の距離が開いていることから、頭部位置の微小な変化では視線一致に不自然さは感じられにくいことも予想される。

一方の参加者の頭部位置が大きく変化した状況を上から見た図を、図3に示す。参加者Aがディスプレイに向かって右方向に頭部を移動させた場合、ディスプレイ上のBの映像を見つめると、カメラAにはAがカメラの右側を見ている映像が撮影される。一方、Bの前のディスプレイにはAが左に移動した映像が表示される。これをBが見つめると、カメラBにはBがカメラの右側を見ている映像が撮影されることになる。このためディスプレイを見つめる角度が許容限を超えると、両参加者は相手が自分よりも右側を見ていると感じるようになると予想される。

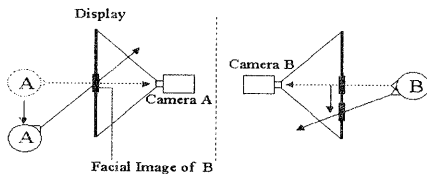


図3：頭部の動きと視線方向の変化

#### 3.2. 視線一致率に関する主観評価実験

3.1 で述べた現状を踏まえ、自分の頭部位置が大きく変化した際、対話相手の視線方向がどのような場合に最も視線が一致していると感じられるか（以後、視線一致率と呼ぶ）を確認するための実験を行った。実験の模式図

を図4に示す。実験手順は以下の通りである。

(a)カメラ位置を固定し、一方の端末正面に座る人物が、ディスプレイ上で横方向10cmおきの計7個所を見つめ、映像を相手サイトに映像を送る。

(b)被験者は7つの各映像について視線の一致具合を5段階評価「一致している」「やや一致している」「どちらとも言えない」「あまり一致していない」「一致していない」で答える。

(c)被験者はこの判定を20cmづつ横にずれた計7個所の頭部位置から行う。

実験条件は以下の通りである。

- ・解像度は40インチディスプレイにVGAの映像
- ・被験者数20人

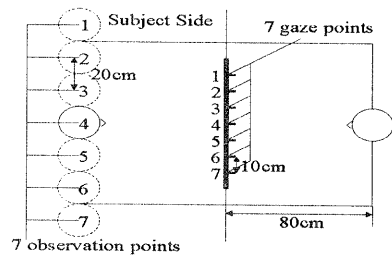


図4：実験の模式図

表1：提示映像の視線方向と観察者の視点位置による視線一致率(%)

		提示される顔映像の視線方向							
		1	2	3	4	5	6	7	
被験者の頭部位置	1	一致している	0	0	15	60	0	0	0
		一致していない	60	30	15	5	45	70	85
	2	一致している	0	0	5	70	0	0	0
		一致していない	65	45	15	0	40	70	85
	3	一致している	0	0	20	85	5	0	0
		一致していない	75	35	15	5	25	65	80
	4	一致している	0	0	5	85	5	0	0
		一致していない	80	65	15	0	15	55	75
	5	一致している	0	0	0	75	10	0	0
		一致していない	85	60	20	0	30	55	75
	6	一致している	0	0	0	65	10	0	0
		一致していない	70	50	20	0	15	35	70
	7	一致している	0	0	0	65	5	0	0
		一致していない	75	55	30	5	20	40	60

実験結果を表1に示す。表1の「一致している」「一致していない」の和が100%でないのは、5段階評価の「やや一致している」「どちらとも言えない」「あまり一致していない」の3つを計上していないためである。

表1より被験者の頭部位置に関わらず、提示映像が中心（ディスプレイ上4の位置）を見つめている場合に最も視線が一致しているとの結果が読み取れる。また、被験者の観察点が中央から離れるにつれ、提示映像が中心（ディスプレイ上4の位置）を見ている場合でも「一致している」という評価が下がっている。例えば、観察点が7の場合の視線一致率は55%であり、観察点が4

の位置の場合（85%）と比べて低い値となっている。しかし、提示映像が他の方向を向いている映像の場合に視線一致率がほぼ0%であり、他の提示映像と比較して、提示映像が中心を見ている場合が最善であるということが言える。以上の結果から、視線一致を保持するためには、相手の頭部位置に関わらず自分がディスプレイ裏のカメラを見つめることが必要であることが分かる。

### 3.3.自由な視点からの視線一致を保持する手法

3.2の議論より、視線一致を保持するためには、参加者がカメラを常に見つめ続ける状況を作る必要があることが分かった。それを実現するために、現在、相手の映像を基に三次元モデルを構築して視点位置に対応した適切な映像を随時合成表示する手法が考えられている[5]が、現状ではリアルタイム性や臨場感において問題があり実用的ではない。そこで本システムでは映像そのものには処理を施さない手法を考えることとする。対話相手の映像に処理を施さずに用いる手法としては、以下の3つの手法が考えられる。

(1)カメラAの位置をAの頭部の動きに追従させて視線の延長線上に移動させる手法（図5）

(2)Aの眼と位置固定カメラAを結ぶ直線と、ディスプレイ面の交点に、Bの顔映像が映るようにディスプレイ上での映像表示位置を移動させる手法（図6）

(3)(2)の手法にカメラのパン操作による頭部追跡を加えた手法（図7）

(1)の手法では、Aの顔はカメラAの撮影範囲の中央に位置し続けるため、頭部の移動に関わらずBの前のディスプレイ映像で、Aは画面の中央に位置し続ける。そのためBはカメラBを真っ直ぐ見つめることになり視線が一致しつづける。しかし本手法ではカメラを可動にするため機器が比較的大規模になり、空間的な制約やコストの面での問題がある。

(2)の手法では、頭部の移動に伴いディスプレイ上の映像を付き従うように移動させることで、カメラの位置を固定のままで、参加者は常にカメラを見つめることが可能となる。映像の表示位置を決めるための頭部位置検出にはカメラの映像を用い、画像の表示位置調整はソフトウェア的に容易に解決できると考えられる。

(3)の手法では、(2)の手法に加えて光軸と視線が一致するため、正面映像で常に対話できるようになる。

以上の議論から、本システムにおいては(3)の手法が適していると考えられる。

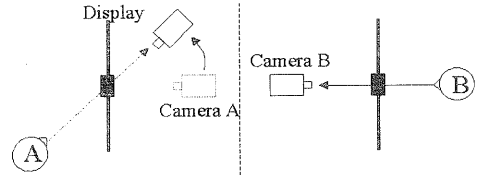


図5：手法(1)における映像表示位置の変化

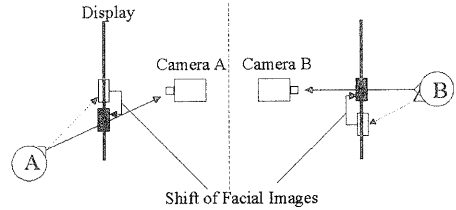


図6：手法(2)における映像表示位置の変化

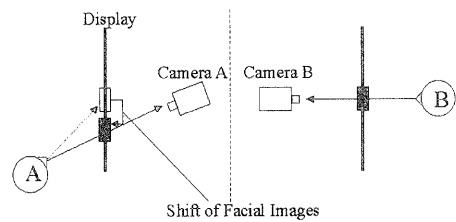


図7：手法(3)における映像表示位置の変化

## 4. ディスプレイ上における対話相手注目点の理解精度に関する検討

### 4.1. 対話相手の注目点を理解する効用

文殊の知恵システムでは、遠隔地間で対話相手に対するプレゼンテーションや協同作業を行い、そこでの作業がより付加価値の高い成果を生み出せる環境を作り出すことを目的としている。対話相手が画面上のどこを見ているかがわかるという効用（Gaze Awareness）は、対話相手の思考プロセスの推測や対話相手へのアドバイスをする上で有益な情報源になることが既に文献[6]で述べられている。

一般的な電子的な資料を用いた会議では、資料ファイルを各人のパソコンに送り各々が自分の前にあるパソコンのディスプレイを見る形式や、大きなスクリーンにプロジェクターを用いて皆が同じ映像を見る形式などがあるが、前者は注目点が他者からは全く分からず、後者は参加者の視点位置と大画面との対応を推測せねばならず、いずれも参加者が何に注目しているか、という推測が直感的には困難である。これは紙資料を用いる場合でも同様である。文殊の知恵システムでは、2.3で述べた資料提示方法により相手の注目しているウィンドウを視線方向

から理解することが出来、上述の問題点が解決されると期待される。すなわち、実際の対面会議において、対話相手がどの資料に注意を向けているかを視線方向から判断する効果を、遠隔地間でも得られると期待される。以下ではそれを確認するための実験を行い、得られた結果をシステム設計へ適用することを考える。

#### 4.2. 注目点の理解精度に関する実験

注目点の理解精度に関し、共有空間に表示するウィンドウはどの程度の大きさで、また何個のウィンドウの同時表示まで、視線方向から識別できるかを確認するための実験を行った。被験者の前のディスプレイには、一人の人物の実物大上半身映像に矩形領域分割され各領域に一つのウィンドウを割り当てた映像をオーバーラップさせて合成した刺激映像を提示する。被験者は映像中の人物がどのウィンドウ（領域）を見ているかをウィンドウに割り当てられた番号で解答用紙に記入して答える。最初に練習用課題としてランダムに視線を散らした映像と各領域を順番に見つめる映像に対して解答した後で、再びランダムに視線を散らした映像を提示して解答を得た。

提示刺激映像の作成では、被写体はディスプレイ正面に座り、ディスプレイの中心点に視点が来るように頭の位置を決めた。そして被写体の前のディスプレイ上に刺激映像と同じく複数のウィンドウで分割された映像（中央に点をうってある）を提示し、領域の中心点を3秒間見つめる。これを各領域についてランダムに行い、ディスプレイ背後にあるビデオカメラで録画した。被写体の位置は通常システム利用時を想定してディスプレイ中央から60cm離れた点とした。自然な対話の状態を模擬するため、頭部の動きには特に制限は設けていない。カメラパラメータは、顔の大きさがディスプレイ上に30cmで表示されるよう調節した。刺激映像の例を図8に示す。

実験条件としては、ディスプレイ領域分割は、9分割（一つの領域が縦20cm横27cm）、16分割（一つの領域が縦15cm横20cm）、25分割（一つの領域が縦10cm横16cm）の3パターンである。各パターンとも顔を表示する領域にはウィンドウを表示していない。また被験者数は12人である

#### 4.3. 実験結果と考察

各ウィンドウの正解率を、ウィンドウ配置とともに図8に示す。各領域の上の数字がウィンドウ番号を示し、下の数値が正解率である。

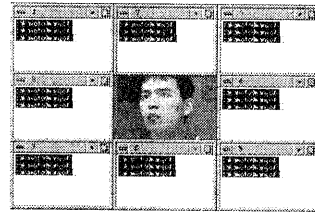


図8：刺激映像の例（ウィンドウ1に注目）

1 100%	2 100%	3 100%	1 91.7%	2 100%	3 100%	4 83.3%
8 100%	9 100%	10 100%	12 100%	13 100%	14 100%	5 91.7%
7 100%	6 100%	5 100%	11 83.3%	9 91.7%	8 100%	6 91.7%

(a)ディスプレイ9分割 (b)ディスプレイ16分割

1 66.7%	2 100%	3 100%	4 100%	5 91.7%
16 75.0%	17 66.7%	18 58.3%	19 58.3%	6 83.3%
14 58.3%	12 75.0%	11 100%	10 100%	7 100%
13 66.7%	15 58.3%	12 75.0%	10 100%	8 66.7%
13 66.7%	12 75.0%	11 100%	10 100%	9 66.7%

(c)ディスプレイ25分割

図9：ウィンドウ配置と各ウィンドウの正解率

図9(a)に示す様に、9分割では全解答者が正しく解答した。即ち9分割でのウィンドウの大きさに関しては、どのウィンドウを見ているかという視線情報が正しく伝達されたと言える。

ウィンドウの16分割でも、図9(b)に示す様に各ウィンドウともほぼ正しく視線情報が伝達されている。不正解の解答を見ると、ウィンドウ11に関しては、ウィンドウ12を見ているという解答であり、ウィンドウ4に関してはウィンドウ5を見ているという解答であった。いずれも不正解の解答は、縦方向に隣合ったウィンドウを解答している。

一方、ウィンドウ25分割では、図9(c)に示す様に、正解率の高いウィンドウと低いウィンドウに大きな差が出た。全解答数44のうち、縦に隣接するウィンドウを解答したものが28(63.6%)、横に隣接するウィンドウを解答したものが11(25.0%)、斜めに隣接するウィンドウを解答したものが5(11.4%)であった。誤答の傾向を矢印で示したものを図10に示す。矢印の始点が正解のウィンドウで、誤答したウィンドウに向かって矢印が伸びている。矢印の太さは誤答数の比例している。

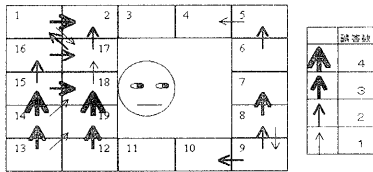


図 10：誤答の傾向

図 10 中の矢印からも、横方向よりも縦方向の誤答が多いという傾向が分かる。特に、ディスプレイの中央の高さにあるウィンドウと、一段下のウィンドウについての誤答が多かった。この結果から縦方向に隣接させることは避けた方がよいと思われる。また横方向に誤答したものは全て端の列のウィンドウを一つ内側のものに答えている。2、3、4 や、10、11、12 を誤答したものは無くこれらは正しく理解されたと考えられる。

実験後に被験者の意見を聞くと、「目の動きよりむしろ首の動きが判断の材料となった」と答えた被験者が多かった。横方向よりも縦方向に誤答が多かったのは、横方向に並ぶウィンドウよりも縦方向の方が密に並んでいることや、刺激映像の頭部の動きが横方向の変化に対して縦方向の変化では頭部の動きがあまり誘起されなかったことに起因すると考えられる。また縦方向の誤答が、特に中央に偏った理由としては、人間の視覚刺激を認知できる視野が、眼の上方向に 35 度、下方向に 50 度であること [7] から、上方にあるウィンドウを見るときは頭部を上方向に向ける動きが誘起され、それを基に判断をつけられるが、中央及び下方向にあるウィンドウを見るときは頭部の動きが誘起されにくいから、判断が難しいということが考えられる。

また、密集している左側については正解率が低いことから判断が難しかったことが分かり、25 分割のウィンドウの大きさでは、端の列の隣にウィンドウを置いても判断が困難であることが分かる。これは本来人間が向かい合った状況での視線方向の推測が、視線の 3 次元ベクトルで判断されるのに対して、本実験では視線が 2 次元映像に投影されているため奥行き情報が消失し、視線の 2 次元ベクトルが複数のウィンドウにまたがったためと考えられる。しかし誤認パターンとして、14 を 12 と間違えたものや、2 と 16 を間違えたものが無いことから、斜め方向の隣接に関しては、視線方向に直交する向きでは可能であることが分かる。25 分割のウィンドウの大きさで認知可能と思われるウィンドウ配置を図 11 に示す。

以上の実験結果をまとめると、9 分割、16 分割の大きさのウィンドウでは、対話相手の視線から注目している

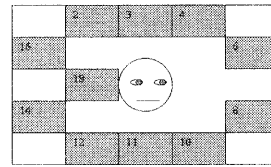


図 11：25 分割で認知可能なウィンドウ配置

ウィンドウが判断できるが、25 分割の大きさでは、縦方向の隣接と、端の列では横方向の隣接するウィンドウについて、判断が難しくなることが分かる。この実験結果より、注目点の理解精度を考慮してウィンドウ表示手法を設計する場合、9 個、16 個までのウィンドウならば比較的自由に配置できるが、25 個を超えるウィンドウを表示する場合には、

- ・縦方向に隣接させない
  - ・ディスプレイ端で横方向に隣接させない
  - ・視線方向と同じ向きに斜め方向に隣接させない
- などの点に注意を払う必要がある。

## 5. まとめ

本稿では、筆者らの提案する「文殊の知恵システム」において、参加者の頭部の動きにも対応して自由な視点位置からの視線一致を実現するため、カメラの移動とディスプレイへの映像表示の方法について検討した。また、対話相手の注目しているウィンドウをどの程度まで理解できるかを実験し、共有空間設計への適用について考察した。今後は本稿での検討事項を実装し、実際に利用した場合においてどのような効果があるかを評価してゆく予定である。

## 参考文献

- [1]安田,青木,クスタルト,坂本,鈴木,佐分,“文殊の知恵システム：意志決定可能な視線一致型テレビ会議システム”, 情報研報 98-AVM-23,1998.
- [2]日経コミュニケーション, “特集, 離陸するパソコン会議システム”, 日経 BP 社, No.223, pp. 56-73, 1996.
- [3]岡田, 松下, “臨場感のある多地点テレビ会議システム: MAJIC”, 情報処理学会論文誌, Vol. 36, No. 3, pp. 775-783, 1995.
- [4]佐藤,三浦,永田,“映像電話における撮像管の位置に関する検討”,電気四学会連合大会, 発表番号 1998 (1967).
- [5]村井,加藤,井口,“テレビ会議システムにおける対話者情報表示法の検討”, 計測自動制御学会, 第 13 回ヒューマンインタフェースシンポジウム論文集, pp.673-678,1997.
- [6]Ishii, H, “Translucent Multiuser Interface for Realtime Collaboration”, the IEICE Transactions on Fundamentals of Electronics, Communications and Computer Science, Vol.75, No.2, 1992.
- [7]Hatada, T., Sakata, H. & Kusaka, H., “Psychophysical Analysis of the ‘Sensation of Reality’ Induced by Visual Wide-Field Display”, SMPTE Journal, Society of Motion Picture & Television Engineers, Vol.89, pp.560-569, 1980.