

多人数参加型テレビ会議システムにおける発言者拡大映像の作成

富野 剛† 井上 亮文†
市村 哲† 松下 温††

従来のテレビ会議システムは、表示サイズや表示解像度が限られているため「誰が発言者かわかりにくい」または「発言者の表情をとらえにくい」などの問題点があった。著者は、会議室内の参加者の中から発言者を自動的に検出し、検出した発言者を拡大表示して遠隔地に伝送することが可能なテレビ会議システムを構築することを狙いとして研究を実施している。本論文では、このような1拠点に10人程度が参加できる1拠点多人数参加型テレビ会議システムを構成するための、マイクロフォンアレイと映像処理を用いた映像表示法について述べる。汎用PCと市販音響機器のみから構成されていることが特徴であり、特殊なハードウェアを使用することなく、通常の会議室環境において10人程度の会議参加者から発言者の方向を検出して拡大表示することが可能である。マイクロフォンアレイからの音声情報をデジタル処理して発言者の位置を特定する手法と、複数の人物を含む映像をデジタル処理して発言者の人物の上半身映像を拡大表示する手法によって実現した。また、紙をめくるノイズが発言と間違われぬよう工夫を行った。

A Speaker Zooming Method for Room-to-room TV Conference

TAKESHI TOMINO,[†] AKIFUMI INOUE,[†] SATOSHI ICHIMURA[†]
and YUTAKA MATSUSHITA^{††}

Due to the limitation of display size or resolution of traditional TV conference systems, we can not see detailed expression on speaker's face shown on a TV screen, or can not even see who is speaking from the remote site. We have been developing a TV conference system where a speaker is automatically identified and his/her face is zoomed in on the TV screen, so that remote participants can better read speaker's facial expression. In this paper, a microphone array and video processing for TV conference systems are described. Our microphone array system is unique because it consists only of one consumer PC, an extra PC sound card, two general cameras, and four omni directional microphones; no other special hardware or special operating system is required. It has a capability that a speaker can be detected out of more than 10 participants in a meeting room. The system expands speaker's upper-body image based-on the speaker's position detected with the microphone array. Then we propose a method to prevent the system from mistaking paper flip-flap noise for voice from a speaker.

1. はじめに

近年、ブロードバンドを用いた常時接続環境が整備され、従来はテキスト中心だったネットワークコミュニケーションの形態が、音声や映像などマルチメディア情報を利用した形態へと変化してきた。この中でも、企業を中心として、インターネットを利用したテレビ会議システムを導入する例が増えている。

遠隔地への会議出席は、交通・宿泊費などが多大に

かかるため、それらを極力圧縮したいという企業は多い。さらに、危機管理の一環としてテレビ会議を導入する企業も増えている^{1),2)}。

現在、テレビ会議システムの形態は、テレビ会議システム専用のカメラやディスプレイ、通信装置を用いることで実現される「1拠点多人数参加型」と、パソコンとソフトウェア、ヘッドセットなどを利用するだけの「単独参加型」に大別できる。1拠点多人数参加型は1拠点で多くの人が利用でき、画質・音質ともに優れているが、かかる費用が高いという問題がある³⁾。一方、単独参加型では、1拠点で利用できるのは1人または多くても2,3人に限られており、画質・音質も低いが、費用が安いという利点がある。

しかしながら、「1拠点多人数参加型」のテレビ会

† 東京工科大学

Tokyo University of Technology

†† 住宅情報化推進協議会

Forum for Agreeable Living with Intelligence, Communication & Electronics

議システムを多くリリースし、テレビ会議業界を牽引している TANDBERG 社の売り上げが近年伸びていることや⁴⁾、国内外問わず、多くの導入実績が見られることから⁵⁾、会議室どうしを接続して、報告会、レビュー会、意思決定会議、経営会議などを行いたいという要求は依然として多いことが分かる。また、たとえば、1 拠点 10 人まで参加できるテレビ会議システム⁶⁾ が販売されているように、企業の拠点間会議では、1 拠点に 10 人程度が会議に参加できるシステムを用いたいというニーズがあることが推測できる。

しかし、1 拠点に多人数が参加した場合は、画面の表示サイズが限られており、「相手の表情をとらえにくく発言者が判別しにくい」「誰が参加しているのかわかりにくい」などの問題点があった。そこで、著者らは、発言者が容易に特定でき、その表情が分かりやすいテレビ会議システムを構築することを目的として、複数の会議参加者の中から発言者を自動的に検出する機能と、その検出した発言者を拡大表示する機能について研究を行っている。

従来研究としては、発言者拡大映像を得るために、発言者を認識しカメラワークを用いてテレビ会議システム映像の演出を行うものや、発言者を唇の動きから自動認識するものなどがあつた^{7)~10)}。しかし、会議参加者全員にマイクロフォンを用意したり、顔をつねに正面から撮影したりしなければならないなどの問題があつた。

なお、本論文では、1 拠点 10 人程度が参加できるテレビ会議システムを「多人数参加型テレビ会議システム」または「1 拠点多人数参加型テレビ会議」と呼ぶこととする。

発言している人を検出する方法としては、当初、映像情報のみを用いる方法について検討した。具体的には、予備実験として人の唇の動きから画像認識を用いて会議中の話者を特定する方法について、市販の画像認識ライブラリ¹¹⁾ を利用した実験システムを構築して性能を調査した。しかしながら、会議中話者がテレビカメラ設置方向を向いていることはむしろ少なく、横の参加者と話したり、机の上の資料を見ながら顔を伏せた状態で話したりすることが多く、必要な認識性能が得られないことが分かった。この経験から、本研究では、人の向いている方向に認識性能が影響されにくい音声情報を映像処理と組み合わせることにより、検出精度を高めることにした。

このような多人数参加型テレビ会議システムを構成するため、本論文では、拡大対象である発言者を特定するために、映像処理による会議参加者位置推定と、

マイクロフォンアレイによる発言者方向推定を組み合わせさせた手法を提案する。

マイクロフォンアレイは今回新規に開発したものであり、PC 1 台と安価な市販音響機器（サウンドカード、マイクロフォン、マイクミキサ）のみから安価に構成できることが特徴である。

従来のマイクロフォンアレイでは、3 つ以上の音声入力を同期することができる専用ハードウェアを使用して 0~360° の話者方向を推定している^{12),13)}。しかし、本研究では、OS の処理遅延の影響で 2 つの音声入力の同期しかできない汎用 PC サウンドカードを用いてこれを可能とした。また、2 つのマイクロフォンを結んだ直線上に音源が近づくと角度推定誤差が大きくなるという従来の遅延時間推定法の問題に対処した。加えて、紙めくりノイズが人の声と間違つて認識される問題に対処する方法を確立しシステムに実装した。

さらに、フレーム間差分、HSV 表色系による肌色認識を組み合わせると発言者方向検出精度を向上させることにより、10 人程度の会議参加者からなるテレビ会議環境において、特殊なハードウェアを使用せず、また顔の向きにも依存しない発言者特定を可能にした。

2. 要求条件と提案手法

2.1 要求条件

特殊なハードウェアを使用しないテレビ会議システムの実装にあたり、要求条件を以下のように定めた。

- (1) PC と市販撮影・音響機器のみで構成する。
- (2) 少ないマイクロフォンで 360° 範囲を認識する。
- (3) 2 台の民生用 DV カメラを用い会議室全体映像と話者拡大映像を作成する。
- (4) 紙をめくる音を発言者として誤認識しないようにする。

テレビ会議システムへの応用を想定した場合、生産まで見こしてシステムを設計する必要がある。このことから著者らは、多数のマイクロフォンを用いたり、特殊なハードウェアを導入したりすることなくシステムを構築できることが望ましいと考えた。さらに、専用ハードウェアを用いずにシステムを構成できれば、故障時に容易かつ迅速に取替えが可能、新たに金型や生産ラインなどを構築することなく量産が可能、資源の調達幅が広く可能、などの利点がある。そこで、要求条件 (1), (2) を定めた。

また、DV カメラ 1 台のみでは、着席位置によっては横顔しかとらえられない参加者が存在してしまうことが考えられるため、要求条件 (3) を定めた。さらに要求条件 (3) は、発言者以外の参加者の様子や参加

状況を伝えるための要求条件でもある。

さらに、紙めくりノイズが人の声と間違っ認識される問題に対処するため、要求条件(4)を定めた。

本研究では、以上の条件を満たし、発言者を自動で認識し、表情が分かるように拡大した映像を遠隔地に送信することで従来のテレビ会議システムの問題点を解決することを目指した。

2.2 提案手法

発言者認識の精度を上げるための手法として、マイクロフォンアレイと映像処理を組み合わせた。後述の実験結果が示すように、マイクロフォンアレイだけでは必要とする認識精度を達成できないためである。

2.2.1 マイクロフォンアレイ

360°方向を識別できるマイクロフォンアレイの構成について述べる。著者らは、少ないマイクロフォンで多角度の音源方向の特定が可能な遅延時間推定法(各マイクロフォンに到達する音の時間差から音の到来方向を求める方法)を用いることとした。

一般的なPCサウンドカード(またはPC内蔵サウンドデバイス)には、モノラルマイクロフォン入力端子1つと、ステレオライン入力端子1つしか備っていない。しかも、モノラルマイクロフォン入力端子からの入力と、ステレオライン入力端子からの入力は、サウンドカード上で合成されてしまうために、ソフトウェアで、モノラルマイクロフォン入力端子からの入力と、ステレオライン入力端子からの入力を区別して処理することは不可能である。このことは、PCサウンドカード1つでは、3本のマイクロフォンからの入力音声を処理して音源方向を認識することができないことを意味している。

そこで著者らは、当初3つのマイクロフォンからの音声入力を1台のPCで同時に行うために、USBマイクロフォン端子変換アダプタ(マイクロフォンをPCのUSB端子に接続するための変換アダプタ)を用いた。このアダプタをPCに2つ増設することにより、内蔵サウンドデバイスとあわせて3つのマイクロフォン入力をPCに取り入れるようにした。しかしながら、3つのマイクロフォンを1カ所に置いて同一音源からの音を同時入力してみたところ、PC上のソフトウェアには、異なったタイミングで入力された音響データとして認識されることが分かった。実測の結果、3つの入力音響信号間には、平均780 μ sの時間のズレが生じていた。44,100Hzで入力音声をサンプリング(1サンプルの時間長は22.68 μ s)している本システムにおいて、780 μ sの誤差は30サンプル以上の誤差であり、マイク間距離を40cmに設定した場合には数度か

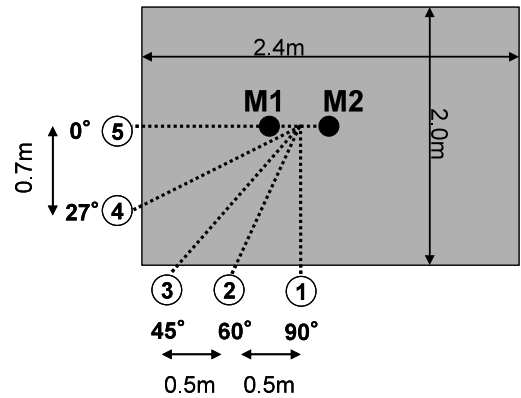


図1 話者位置推定実験(90°範囲)

Fig. 1 Experiment on speaker recognition (90-degree).

ら数十度のランダムな話者位置推定の誤差として現れてしまうことを意味している。

この原因としては、OSが3つのサウンドデバイスからの入力をタイムスライス処理するために時間差が発生してしまうということや、各サウンドデバイスの種別または個体差によりPCに取り込む速さが異なっていることが考えられる。USBマイクロフォン端子アダプタの代わりにPCIバス用サウンドカードを追加した場合にでも、依然としてランダムな時間的ズレが存在し要求条件の精度を確保できないことが分かった。

次に、2本のマイクロフォンM1およびM2を、互いに40cm離して設置したマイクロフォンアレイを用いて認識率を測定した。この条件において入力音声を44,100Hzでサンプリングした場合には、180°の範囲を102方向に分別できることとなる。ただし、2つのマイクロフォンを結んだ直線上に音源が近づくと遅延時間のわずかな差で音源方向推定結果が大きく異なるという現象が起きる。

そこで、2マイクロフォンを用いた話者方向推定精度を調べる実験を行った。空調やPC数台が稼働する会議室の中で、200cm×240cmのテーブルの上に2つのマイクロフォンを40cm離して設置し、図1に示される地点①~⑤(90°, 60°, 45°, 27°, 0°)に会議参加者を配置した。そして、前述の要求条件に基づき、 $\pm 15^\circ$ までの誤差を許容するようにして(目標とする性能は、360°方向に着席した10名を分別できること、すなわち許容角度誤差 $= \pm 18^\circ$ であるが、会議参加者が必ずしも均等幅に着席しないことを考慮し、360°方向に着席した12名を分別できる性能を有するかどうかである許容角度誤差 $= \pm 15^\circ$ によって判断することとした)、各地点の話者位置認識率を測定した。

実験結果を表1に示す。実験結果に基づけば、地点

表 1 話者位置認識率
Table 1 Speaker recognition rate.

話者位置	角度 (degree)	認識率 (%)
①	90	97.5
②	60	87.0
③	45	75.2
④	27	33.3
⑤	0	21.0

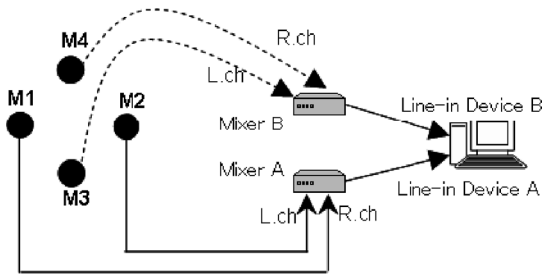


図 2 マイクロフォンの構成と概観
Fig. 2 Microphone array.

①～地点③の話者分別性能については、地点④の話者と地点⑤と比較し、2倍以上の精度で検出可能であり、地点④、⑤話者を $\pm 15^\circ$ の精度で分別することはほとんど不可能であるといえる。当初予想したとおり、2つのマイクロフォンを結んだ直線上に音源が近づくと、極端に認識精度が下がってしまう結果となった。

以上の検証結果に基づき、本マイクロフォンアレイを図2のように構成した。2本のマイクロフォンを使った話者位置推定実験から得た知見を利用し、M1とM2のマイク対(マイク間距離40cm)と、M3とM4のマイク対(マイク間距離40cm)とで、認識する方向を、要求条件を満たす角度範囲で分担するように工夫している。

具体的には、M1とM2のマイク対、M3とM4のマイク対のどちらか一方が、各マイク対を結んだ直線に対して $40 \sim 135^\circ$ の話者方向推定結果を出力した場合にはその値を採用し、もし、両方のマイク対が $0 \sim$

45° または $135 \sim 180^\circ$ の話者方向推定結果を同時に出力した場合には各マイク対の平均角度を採用するようになっている。

次に紙めくりノイズが、人の声と間違っ認識されることへの対処について述べる。

本システムにおいては、人の声以外のノイズを軽減する目的で、 $300 \sim 2,500$ Hz 以外の周波数成分の除去を行っていたが、紙をめくる音がある一定値以上の音量になると発音者の発音と誤認識されることが分かった。紙をめくる音の周波数成分が $300 \sim 2,500$ Hz に多く存在するため、雑音として除去できなかったことが原因であると考えられる。紙をめくるときに発生する音の周波数成分は人の声の周波数成分と重なる部分が多いと多く、紙めくりノイズを効果的に除去する技術は現段階において確立されていない。本システムのマイクロフォンアレイの用途は、音源の方向を特定することであって、雑音を除去した会話音声を作成することではない。このような、紙めくりノイズを発音者と誤認識しないようにするという観点から実施された研究例や報告は見受けられなかった。

そこで著者らは、紙資料を会議参加者に配布し、この紙資料を積極的に使うように指示して会議を行わせる実験を行った。そして、人の音声と紙をめくる音とをサンプル数 2,048 で高速フーリエ変換を行い、両者の周波数スペクトルを比較した。その結果、両者の周波数スペクトル間で、最もパワーの大きい周波数に違いがあることが分かった。本実験では、以下の①～③の紙めくりノイズを発生させ、それぞれの場合において最もパワーの大きい周波数を求めた。データサンプル数はそれぞれ 120 である。

- ① ノート：ノートのページをめくる音
- ② ホッチキス：ホッチキス留めの紙の束をめくる音
- ③ 1枚紙：1枚の紙をくしゃくしゃにする音

実験の結果、上記3種類の紙めくりノイズの場合は、いずれも、最もパワーの大きい周波数が $1,300$ Hz 以上であることがほとんどであった。なお、この実験に先立ち、年代の異なる被験者 15 人を対象とした実験によって、男性の声の場合は約 $400 \sim 700$ Hz、女性の声の場合は約 $600 \sim 1,000$ Hz の範囲に最もパワーが最大となる周波数成分が存在することを調べておいた。このことから、入力信号の最もパワーの大きい周波数について $1,300$ Hz を境として分別することにより、紙めくりノイズと人の声とを判別できる可能性が高いと判断した。

そこで、音声信号の最もパワーの大きい周波数が $1,300$ Hz 以上であった場合には、その音声信号をノイ

ズと見なして話者方向推定の処理を行わないようにシステムを改良し、この効果を確認する実験を行った。具体的には、この紙ノイズ除去処理を加えない場合と加えた場合とで、紙ノイズを話者として誤認識した回数を比較した。表2の実験結果が示すように、紙ノイズ除去処理後は誤認識回数が約1/3に減少する結果を得た。

2.2.2 映像処理

映像処理では、フレーム間差分認識と肌色認識とを行っている。フレーム間差分認識とは、連続する2枚の画像から、画素の差分をとることで動物体を検出する方法である。差分値を単純な計算で検出でき、かつ、前景と背景を区別する必要がないという長所がある(図3)。次に肌色認識では最初RGB画素値から検出を試みた。デジタルカメラで撮影した顔画像のサンプルを用いて各画素のRGB値を256段階で調べた結果、特に傾向は見られず特徴量を検出するのは難しいことが分かった。

そこで、RGB画素値をHSV表色系に変換し、再度

表2 ピーク周波数成分の分布
Table 2 Peak frequency.

雑音種類	雑音除去なし (誤認識回数)	雑音除去あり (誤認識回数)
ノート	40	15
ホッチキス	37	23
1枚紙	51	5
合計	128	43

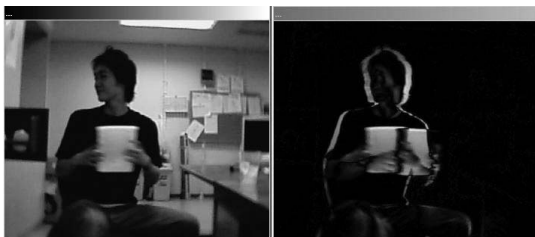


図3 フレーム間差分認識
Fig.3 Interframe detection method.

表3 人物の肌のHSV値
Table 3 HSV data of human skin.

肌 サンプル	A	B	C	D	E	F	G	H	I
H (色相)	19	11	17	8	10	4	22	21	12
S (色彩)	86	95	115	89	112	87	140	68	30
V (明度)	180	208	141	173	159	184	136	206	246

各要素の画素値を求めた(表3)。HSV表色系とは、色合いを表す色相:H、色の鮮やかさを表す彩度:S、色の明るさを表す明度:Vで表される表色系である¹⁴⁾。

実験した結果から、 $4 < H < 35$ の範囲が肌色であることが多いことが分かった。なお、色そのものは色彩であるHで表され、明度であるVとは独立しているため、肌色を検出する際に明るさの変化に依存しないという肌色認識には好都合な特徴がある。

3. システム構成

本システムは、図4のような会議室環境を想定しており、会議参加者から、発言者を拡大して表示できるように設計されている。

映像入力機器として民生用のDVカメラを使用している。DVカメラ2台からの入力映像と、マイクロホンアレイから得られた話者位置推定データを統合して話者の検出と拡大をしている。その際に、紙めくりノイズの誤認識を防ぐための処理を加えている。

1地点の会議用映像として、発言者3人まで拡大したものを映像上部に、会議室全体を映したものを映像下部に出力している(図5)。そしてこの映像を他の地点に送信し、同時に、他の地点の映像を受け取ることで会議システムを構成している。システムの処理の流れは図6のようになっている。

1サイト分の全体システムは図7のようになっている。図5のような会議映像は、図7の「映像・音声処理PC」によって作成され、そのRGB出力がダウンコンバータによってDV映像信号に変換され、もう1台の「遠隔会議PC」に入力される。遠隔会議PCは、テレビ会議ソフトが稼動しているテレビ会議用クライアントPCである。遠隔会議PCにとっては、映像・音声処理PCから出力された会議映像は、DVカメラからの映像として認識されるため、このテレビ会議ソフトの入力とすることができる。すなわち、

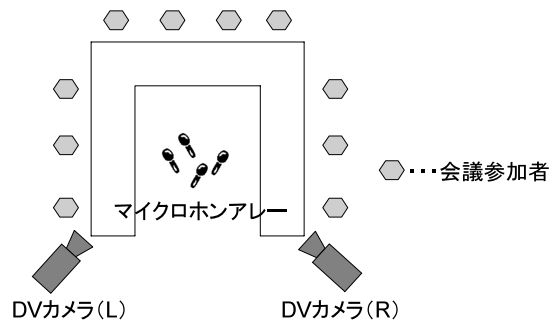


図4 会議室想定環境
Fig.4 Meeting room setting.



図 5 実行画面
Fig. 5 Execution screen.

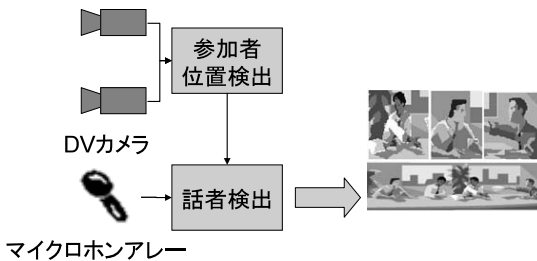


図 6 処理の流れ
Fig. 6 Flow of processing.

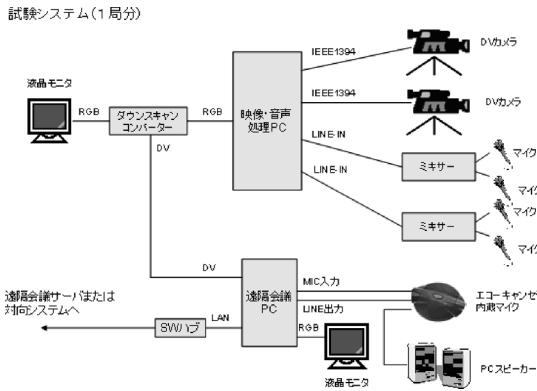


図 7 全体システム図 (1 サイト分)
Fig. 7 System architecture (one site).

映像・音声処理 PC は遠隔会議 PC にとって仮想カメラデバイスという位置づけとなる。

4. 実験と評価

著者らが試作した多人数参加型会議システムにおいて、多人数の被験者を投入して行う被験者実験を行った。会議室は、空調、PC プロジェクタ、PC のファンな

どの環境雑音が存在する一般的な会議室である。開発したマイクロフォンアレイを会議机 (360 cm × 360 cm) の中央に配置し、10 人の被験者を会議机の周りに着席させて実験を実施した (図 4 のとおり着席させた)。実験 1 では本システムの認識精度について検証した。実験 2 では、「発言者が誰か分かるか」「発言者の表情がとらえられるか」などのアンケート調査を行った。

4.1 実験 1：認識精度

開発したテレビ会議システムを実際に使用して評価した。評価環境は図 4 のように設定し、参加者 10 人で本システムの認識精度を調べた。のべ 2 時間半にわたり、拡大表示に映る会議参加者を記録した。なお、会議では報告や発表など 1 人が継続的に話すことが多く、特定の人物が「主たる発言者」となることが多い。このことから、本実験を行うにあたっては、所定時間内の発言時間が最も長い人を「主たる発言者」と定義し、主たる発言者と会話している発言者を「従たる発言者」と定義した。

認識率は、ある発言者が話した時間長に対する、左上・中央上・右上の発言者拡大位置のいずれかに当該発言者が表示された時間長の割合として算出した。実験の結果、主たる発言者の認識率は 96% となった。また、主たる発言者と従たる発言者の両方が同時に拡大表示された認識率は 94% であった。

なお、映像処理を行わずに、マイクロフォンアレイの認識結果のみを用いた場合には、主たる発言者の認識率は 86% であった。映像処理を行わなかった場合は、人が存在しない領域や人と人との間を拡大表示することがあり、このことが認識率の低下の原因となった。

4.2 実験 2：話者拡大機能の有無による評価実験

計 13 人の学生を被験者とし実際に会議システムを使用させ、後に、このうち 11 人に対してアンケートを行った。

被験者 10 人が会議室 A に、他の 3 人が別の会議室 B に入り、会議室 B の地点から見た、会議室 A の映像を評価した。そして、話者映像の拡大あり、拡大なしの 2 つの状況について、以下の A ~ F の項目についてアンケート調査を行った。そして各項目について点数 (1. 非常に悪い ~ 5. 非常に良い) で評価してもらい、各項目の平均点数を調べた。

- A. 誰が発言しているの分かりやすかったか
- B. 発言者の切替えはスムーズか (映像の拡大ありのみ)
- C. 発言者の表情は読み取れたか
- D. 会議の臨場感はあるか
- E. 実際に同じ会議室にいるように見えたか

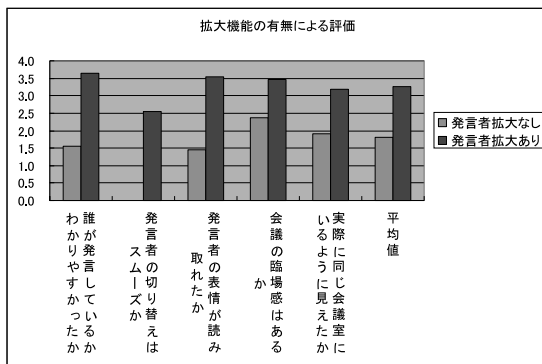


図 8 拡大映像有無による評価の違いのグラフ

Fig. 8 Evaluation with or without video processing.

F. 発言者の拡大は必要だと思うか

被験者の関心が高く、積極な議論をしてもらえらること、個人に焦点が当たる状況を作り出すことを狙いとして、「就職活動について」「自己PRについて」などの内容を議題として設定した。

実験の結果、図 8 から分かるように、すべての項目において、発言者拡大ありの方が高い評価を得た。この結果をウィルコクソンの符号付き順位和検定（有意水準 5%）を用いて検定した結果、すべての項目について有意差があることが分かった。また、話者拡大機能を「必要と思う」か「必要としない」かの 2 択によって調査を行ったところ、被験者全員が「必要だと思う」を選択し、同様に、ウィルコクソンの符号付き順位和検定（有意水準 5%）を用いて検定した結果、有意差があることが分かった。

5. おわりに

マイクロフォンアレイを用いて話者の特定を行い、さらに、2 つの DV カメラから得られた映像からその話者を拡大表示する 1 拠点多人数参加型テレビ会議システムを提案した。

映像処理による会議参加者位置推定と、マイクロフォンアレイによる発言者方向推定を組み合わせた結果、特殊なハードウェアを使用することなく話者映像の拡大表示が可能となった。話者拡大機能の有効性を評価する被験者実験の結果、本システムが提供した話者拡大機能は有効であるとの結果を得ることができた。

参 考 文 献

- 1) CNA レポート—電話会議システム業界専門マーケットリサーチ & コンサルティング。
<http://cnar.jp/> (2006/3/5 アクセス)
- 2) 宮崎県教育研修センター：テレビ会議システム

を利用した連携システムの構築に関する研究—知的障害養護学校における取組を通して (2003)。
<http://mkkc.miyazaki-c.ed.jp/> (2006/3/5 アクセス)

- 3) 橋本啓介：テレビ会議市場についての一考察，CNA レポートジャパン (2002)。
<http://cnar.jp/TV-Kaigi.pdf> (2006/3/5 アクセス)
- 4) CNA Report Japan—電話会議・テレビ会議・web 会議専門ニュースレター，Vol.6，No.12 (2004)。
- 5) TANDBERG—導入事例。
<http://www.tandbergpartner.jp/> (2006/3/5 アクセス)
- 6) ギンガネット。
<http://www.ginganet.co.jp/business.html> (2006/3/5 アクセス)
- 7) 大西正輝，影林岳彦，福永邦雄：視聴情報統合による会議映像の自動撮影，電子情報通信学会論文誌 D-II，Vol.J85-D-II，No.3 (2002)。
- 8) 福井康仁，川村光弘，尾崎知幸，菅原一孔，小西亮介：読唇単語認識システムのハードウェア実現に関する研究，第 46 回自動制御連合講演会 (2003)。
- 9) 井上智雄，岡田謙一，松下 温：テレビ番組のカメラワークの知識に基づいた TV 会議システム，情報処理学会論文誌，Vol.37，No.11，pp.2095–2104 (1996)。
- 10) 井上亮文，吉田竜二，平石絢子，重野 寛，岡田謙一，松下 温：映画の映像理論に基づく対面会議シーンの自動撮影手法，情報処理学会論文誌，Vol.45，No.1，pp.212–221 (2004)。
- 11) NVision：顔画像認識ライブラリ。
<http://www.n-vision.jp/> (2005/11/30 アクセス)
- 12) 西浦敬信，中村 哲，鹿野清宏：マイクロフォンアレイを用いた HMM に基づく音源識別の評価，電子情報通信学会音声研究会，Vol.100，No.522，pp.31–36 (2000)。
- 13) 西浦敬信，山田武志，中村 哲，鹿野清宏：マイクロフォンアレイを用いた CSP 法に基づく複数音源位置推定，電子情報通信学会誌，Vol.J83-D-II，No.8，pp.1713–1721 (2000)。
- 14) 橋本 聡，藤本研司，中村 納，南 敏：顔領域抽出に有効な修正 HSV 表色系の提案，テレビ誌，Vo.149，No.6，pp.787–797 (1995)。

(平成 17 年 11 月 30 日受付)

(平成 18 年 5 月 9 日採録)



富野 剛

1981年生。2005年東京工科大学工学部情報通信工学科卒業。現在、同大学大学院工学研究科博士前期課程在学中。グループウェア、ヒューマンインタフェースの研究に従事。

DICOMO2005 優秀プレゼンテーション賞受賞。



井上 亮文 (正会員)

1999年慶應義塾大学理工学部計測工学科卒業。2001年同大学院理工学研究科修士課程修了。2005年同大学院理工学研究科後期博士課程修了。博士(工学)。現在、東京工科大学コンピュータサイエンス学部助手。

マルチメディアオーサリング、実世界指向インタフェース、ネットワークセキュリティの研究に従事。2005年下期末踏ソフトウェア事業採択(開発代表者)。



市村 哲 (正会員)

1989年慶應義塾大学理工学部計測工学科卒業。1994年同大学大学院理工学研究科博士後期課程修了。博士(工学)。同年富士ゼロックス(株)入社。1997~1999年富士ゼロックスパロアルト研究所(FXPAL)駐在。2002年より東京工科大学助教授。

グループウェア、ネットワークサービス、生体情報活用等の研究に従事。『IT TEXT 基礎 Web 技術』、『IT TEXT 応用 Web 技術』(オーム社)。DICOMO 2003 & DICOMO2005 優秀論文賞受賞。ACM 会員。電子通信情報学会会員。



松下 温 (フェロー)

1963年慶應義塾大学工学部電気工学科卒業。1968年イリノイ大学院コンピュータサイエンス専攻修了。1989~2002年慶應義塾大学理工学部教授。2002~2006年東京工科大学コンピュータサイエンス学部教授。

2003~2006年同大学コンピュータサイエンス学部長。2006年より住宅情報化推進協議会会長。マルチメディア通信、コンピュータネットワーク、グループウェア等の研究に従事。情報処理学会理事、同学会副会長、マルチメディア通信と分散処理研究会委員長、グループウェア研究会委員長、電子情報通信学会、情報ネットワーク研究会委員長、MIS 研究会委員長、バーチャルリアリティ学会サイバースペースと仮想都市研究会委員長、情報処理学会 ITS 研究会委員長等を歴任。郵政省、通産省、建設省、農水省、都市基盤整備公団、行政情報システム研究所等の委員長、座長、委員を多数歴任。特に国土交通省、住宅情報化標準策定委員会委員長、経済産業省総合エネルギー調査会ルータ装置基準委員会委員長、最高裁判所専門委員を務める。『やさしい LAN の知識』(オーム社)、『201x 年の世界』(共立出版)等著書多数。1993年情報処理学会ベストオーサ賞、1995年および2000年情報処理学会論文賞、2000年情報処理学会40周年記念論文賞受賞、2000年VR学会サイバースペース研究賞、2001年5月情報処理学会功績賞。情報処理学会フェロー、電気通信学会フェロー、人工知能学会、ファジイ学会、IEEE、ACM 各会員。