MM-Space: 動的投影を用いた頭部運動の物理的補強表現に基づく会話場再構成

大塚和弘† 熊野史朗† 三上弾†
松田昌史† 大和淳司†

複数人物による対面会話場面を実世界に再構成するシステムを提案する．本研究では，時間・空間を隔てた会話の状況を，あたかもその場にいるかのように観測できるシステムの実現を目指し，会話者の顔画像をスクリーンに投影しつつ，その人物の頭部運動をスクリーンの物理的な運動として補強して提示するという表現モダリティを提案する．本システムでは，会話者の配置に合わせて，別地点に複数のプロジェクタ，及び，透過型スクリーンが配置され，各スクリーンに等身大の人物像が投影される．スクリーンにはアクチュエータが接続され，計測された会話者の頭部姿勢と同期して，スクリーンの姿勢が動的に制御される．人物の頭部運動には，視線方向の遷移に伴う首振りや頷きが含まれ，このスクリーンの物理的な運動と画像運動との相乗効果により，会話者の動作がより明確に観測者により知覚されることが期待される．さらに枠なし透明スクリーンへの背景なし人物像の投影により，遠隔人物が観測者の部屋の背景に重畳表示され，高い存在感の醸成も期待できる．本稿では，このスクリーンの動きの効果として，会話者の視線方向がより分かりやすく，その結果，話し手の話し掛ける相手がより正確に理解されるという仮説を立て，被験者実験によりこの仮説の妥当性を示唆した．

MM-Space: Re-creating Multiparty Conversation Space based on Physically Augmented Head Motion by Dynamic Projection

Kazuhiro Otsuka,† Shiro Kumano,† Dan Mikami,†
Masafumi Matsuda† and Junji Yamato†

A novel system is presented for reconstructing multiparty face-to-face conversation scenes in the real world; its key component is based on highly effective dynamic displays that augment human head motion. This system aims to display and playback recorded conversations as if the remote people were taking in front of the viewers. This system consists of multiple projectors and transparent screens attached to actuators. Each screen displays the life-size face of a different meeting participant, and are spatially arranged to recreate the actual scene. The main feature of this system is dynamic projection; screen pose is dynamically controlled in synchronization with actual head motions of the participants to emulate their head motions, including head turning, which is typically accompanied with shifts in visual attention. Our hypothesis is that physically augmented screen movement with image motion can boost viewers' understanding of the evolution of others' visual attention. Experiments suggest that viewers can more clearly discern the visual focus of attention of meeting participants, and more accurately identify the addressees.

1. はじめに

人と人とが対面で行う会話は，もっとも基本的なコミュニケーションの形態であり，日々，情報の伝達・共有，他者の感情や意図の理解，グループでの意志決定が会話を通じて行われている．昨今，空間を隔てた会話を実現するため，映像通信の活用が期待されているが，いまだ対面の会話ほど自然なコミュニケーションを実現するに至っていない．著者らは，より良い遠隔コミュニケーション環境を実現するには，まず，人の会話のメカニズムについて理解を深める必要があるという思想のもと，会話シーン分析の研究を進めている[15]．その一例として，頭部動作や発話区間といった非言語行動の計測，人物間のインタラクション（e.g.「誰が誰に応答しているか」）や会話の構造（e.g. モノローグ，ダイアローグ），対人感情（e.g.「誰が誰に共感しているか」）の自動推定などがあげられる．現在，このような会話シーン分析の成果を実際のコミュニケーション・システムへと活用すべく，本稿では複

† 日本電信電話 (株)，NTT コミュニケーション科学基礎研究所
NTT Communication Science Labs., NTT Corporation

数人物の対面会話の場を，別の地点で再構成するという初期課題への取り組みを紹介する．

対面会話においては，視線や頭部ジェスチャ，顔表情，韻律など様々な非言語情報が交わされている[2]．我々は，遠隔コミュニケーション環境では，これらの非言語情報が十分に伝達されないことが，その不自然さの一因であると考え，これを何らかの手段で補い提示する必要があると考えている．様々な非言語情報の中でも，本研究では，頭部の姿勢や運動として表出される情報に着目している．その一つが「誰が誰を見ているか」という視線方向（視覚的注意の焦点とも呼ばれる）である．視線には，他者のモニタリング，自身の態度・意図の表出，会話の流れの調整といった機能がある[9]．人は，興味の対象を視野の中心に捉えることから，対象との位置関係に応じた頭部姿勢（顔向き）として視線方向が表出される．さらに，頷き，首振り，傾げといった頭部ジェスチャも重要な非言語情報である．話し手の頭部ジェスチャは，話し掛け・問いかけや発話の強調のサインであり，また，聞き手の頭部ジェスチャは，傾聴や承認，同意・不同意，理解の程度を表すサインとして機能する[12]．

本研究では，このような非言語情報を含めた会話の場を実空間上に再構成し，時間・空間を隔てた会話の状況を，あたかもその場にいるかのように観測可能なシステムの実現を目指す．特に視線方向や頭部ジェスチャの表出を担う頭部運動に着目し，人物像を投影するスクリーンの物理的な運動として頭部運動を補強的に表現する方法を提案する．本システムは，会話者の人物配置に合わせて，別地点に複数のプロジェクタ及び透明スクリーンを配置し，等身大の人物像を投影する．スクリーンにはアクチュエータが接続され，会話者の頭部運動と同期してスクリーンの姿勢が制御される．このスクリーンの物理的な運動と投影画像上の運動との相乗効果により，会話者の頭部動作がより明瞭に知覚でき，その結果，視線方向や話し手の話し掛ける相手（受け手と呼ぶ），及び，頭部ジェスチャが，より明瞭に観測者によって理解されることを期待する．また，透明スクリーンへの背景除去された人物像の投影により，遠隔人物の顔が実際の部屋の背景に重畳されて表示され，高い存在感の実現も期待される．

特に本稿では，頭部運動の補強表現の効果として，会話者の視線方向の理解が促進され，その結果，受け手人物がより正確に同定できるという仮説を立て，被験者実験によりその妥当性を示唆する．なお，会話場の再現には，視線方向以外の非言語情報の表現も重要であると考えられるが，本稿では，遠隔システムの典型的課題である「視線不一致」の現象と関連して，視線方向の正確な伝達をとくに先決な課題として取り上げる．また，本稿のシステムは，オフライン動作を対象とするが，双方向の遠隔会話を実現するシステムの構成要素として位置づけられる．

頭部運動の物理的補強表現は,「バイオロジカルモーション」[7],「心的帰属」[5] と呼ばれる人間の知覚の性質に関連する．これら理論によると，人間には，光点や図形の動きを人や生物の動きに「見立て」，その原因を対象の心的状態や社会的文脈に帰属させて解釈する傾向がある．この理論に基づき，正方形スクリーンという単純形状であっても，人間の頭部運動に近い動きで動かせば，観測者はスクリーン平面を顔面と見立て，その動きから会話者の動作やインタラクションをある程度，読み取ることができると期待される．

また，人間の視覚は，特に周辺視野に表れた動きに対する感度が高いことが示唆されている[10]．これにより観測者の周辺視野に位置する人物の動作も，そのスクリーンの動きからより明瞭に察知できることが予想される．このように本研究は，現実に近い運動の提示がコミュニケーション行動の理解に有用であるとの仮説のもと，スクリーン姿勢の変化として，人物の頭部運動を補強表現する方式を提案するものである．

本稿は以下のように構成される．第 2 節では関連研究に対する本研究の位置づけを述べる．第 3 節では提案システムの構成を記し，第 4 節では実験結果を示す．最後に第 5 節において結びと今後の展望を述べる．

2. 関 連 研 究

従来のテレビ会議システムやテレプレゼンスシステムにおいて，円滑な会話を妨げる原因の一つとして，「誰が誰を見ているか」,「自分に視線が向けられているか」が分かりにくいという視線不一致の問題がある．この問題の緩和には，地点間において人物の空間配置に整合性を持たせることが有効とされる[19][6]．例えば，各人物に個別のディスプレイを割り当て，それを人物の配置に合わせて設置するシステム[19] や，大型ディスプレイで囲まれた空間内で，ある地点の人物の画像を別地点の同一相対位置に表示するシステム[6] が提案されている．このような空間的整合性によって，顔向きの手がかりが適切に伝達され，視線方向の理解が向上すると考えられている．しかし，会話者の画像が空間に分散されて表示されることにより，全人物を一つの視野内に納めることができず，ときに人物間の相互作用が分かりにくいという側面もある．本稿の提案システムでは，この空間的整合性の考え方を取り入れつ

つ，スクリーンの物理運動による補強表現により，周辺視野に位置する会話者の動作の知覚を増強する狙いをもつ．また，縁なし透明スクリーンへの背景なし人物像の投影により，従来システムを上回る存在感・同室感を得ることを狙っている．

また，人物像を CG アバタ/キャラクタとして 3 次元形状も含めて再構成・表示することで，視線不一致の問題解決や，より高い存在感の実現を目指すアプローチがある[8]．しかしながら，現状の技術では，人の微細な表情まで十分に再現可能な水準には達しておらず，3 次元ディスプレイや HMD などの表示デバイスの利用も自然なユーザ行動の制約要因となっている．本研究では，これらアプローチとは異なり，2 次元画像から他者の表情を読み解く人の知覚能力の高さを活用し，それを補助するためスクリーンの物理運動を用いるという新しい表現法を探る立場をとる．

さらに近年ではテレプレゼンスロボットと呼ばれる，ユーザの遠隔操作により会話への参加を可能とするロボットが注目を集めている[4]．高い存在感を与える点が利点とされるが，一方，機械としての存在感が突出し，ユーザの個人性が表現されにくいという問題がある．また，究極の個人化として，アンドロイドの利用が検討されているが[18]，いわゆる「不気味の谷」[13] の克服が長期的課題とされる．また，ロボットには，人の非言語行動を表出する手段としての可能性もある．例えば，頷きの表現や，ロボットアームによるジェスチャ表現[1]，ユーザとの対人距離が可変なディスプレイによる近接学（プロセミクス）の再現[14] などが試みらている．本研究では，ロボティクスによる動き表現の可能性に着目し，複数人会話における対人視線方向，及び，話し掛けの方向性の提示のため，スクリーン姿勢の動的制御による頭部運動の補強表現を提案する点において新規性を有すると考える．また，提案システムでは，機械としての存在感を極力消し，また，CG 画像の不気味さ・不自然さを排除し，さらにその人自身の存在感を高めるため人物の 2 次元画像を（最小限の画像処理のみで）透明スクリーンへ投影するという方式をとる．

3. 提案するシステム: MM-Space

図 1 に提案システム（*MM-Space* と呼ぶ☆1）の外観を示す☆2．また，図 2 には，本システムのブロック図，図 3 には空間配置をそれぞれ示す☆3．実際に会話を行う空間（図 1(c)）では，会話人物の顔画像や音声を取得する（図 2 の入力部，図 3(a) の配置）☆4．また，会話を再現する空間（図 1(a), (b)）では，会話者の配置にあわせて，複数のプロジェクタ，スクリーン，アクチュエータ，スピーカが配置され（図 3(b)），各人の顔画像が透過型平面スクリーンへ投射され，そのスクリーンの姿勢が人物の頭部運動と同期して制御される．なお，本稿の提案システムは，頭部運動の物理的補強表示の効果を検証するため，オフラインの再生に特化した初期システムであり，今後，実時間・双方向システムへの拡張が予定されている．

図 4 に，スクリーン，プロジェクタ，及び，アクチュエータの外観を示す．各スクリーンは，透過アクリル板に光線拡散物質が配合されたものである．また，その背後には，背面投射用の LCD プロジェクタが設置される．各スクリーンは，その下部よりアクチュエーターにより支持される．このアクチュエーターは，鉛直軸周りの回転運動（パンと呼ぶ），及び，水平軸周りの回転運動（チルトと呼ぶ）の運動を生成する 2 つのユニットから構成される．以後，これを PTU（パンチルトユニット）と呼ぶ．この PTU によりスクリーンの姿勢が動的に制御される．

図 2 の処理部では，顔姿勢追跡，背景除去，制御信号生成，画像写像生成が実行される．顔姿勢追跡部では，撮影された画像から各人物の顔の位置と姿勢が推定される．なお，頭部の姿勢計測のためモーションキャプチャなどのセンサを利用することもできる．背景除去部では，入力画像から背景領域を除去し，人物領域のみの画像が生成される．制御信号生成部では，計測された人物の頭部姿勢に基づいて，スクリーンの姿勢を制御するための制御信号を生成する．画像写像生成部では，スクリーンの姿勢と連動して，歪みのない投影像を得るため，入力画像（または背景除去後の画像）に写像を施す．以下，各部の概要を説明する．

顔の位置と姿勢の計測　図 5(a) のような撮影画像から人物の画像上での顔の位置，及び，顔姿勢（頭部姿勢）の計測が行われる．その方法として，STCTracker（疎テンプレートコンデンセーション追跡法）[11] と呼ばれる顔姿勢追跡法を用いる．STCTracker では，顔

☆1 MM-Space の *MM* は，Multimodal, Multiparty, Meeting, Motion, Minimalism 等を象徴する．

☆2 http://www.brl.ntt.co.jp/people/otsuka/Interaction2012.html にてデモムービーが視聴できる．

☆3 この様な人物配置を採用した一つの理由としては，遠隔システムへの拡張を念頭にし，遠隔の参加者（本稿でいう観測者）を配置する空間を設けるという意図がある．また，画像上の顔方向のみからは対人視線方向が判別しづらい状況を作り，提案方式の効果をより明瞭に浮き上がらせる意図もある．

☆4 なお，各人物毎に個別のカメラ，マイクを用いる代わりに，文献[16] のような全方位センサも利用可能である．

図 1　提案システム（MM-Space）の外観．(a) 背景除去あり，(b) 背景除去なし，(c) 元になった会話シーン
Fig. 1　Overview of proposed system. (a)background removal, (b)without background removal, (c)original meeting scene

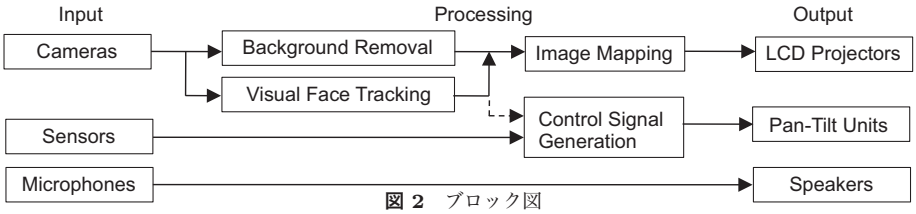

図 2　ブロック図
Fig. 2　Block diagram of system

の 2 次元位置，及び，3 軸周りの回転角，スケーリング係数，照明変動係数からなる 7 次元の状態変数が，GPU 実装されたパーティクルフィルタにより逐次推定される．図 5(b) の例では，顔姿勢の推定結果がメッシュとして重畳表示されている．また，顔姿勢の計測には磁気式のセンサを用いることもできる．

背景除去　入力画像に対して背景除去の処理が施される．図 5(c) に結果例を示す．本システムでは，混合正規分布に基づく背景モデル化[17]，及び，動的輪郭法による前景抽出[3] を組み合わせた方法が使用される．

アクチュエータの制御信号の生成　PTU の機械的特性を加味しつつ，できる限り実際に近い頭部運動を再現することを狙い，計測された頭部姿勢の時系列データから，PTU に入力される制御信号が生成される．なお，頭部姿勢は画像として既に表現されている点，及び，スクリーンの姿勢を，観測者からみてスクリーン像が良好に視認できる範囲に保つという観点より，PTU の制御姿勢は，頭部姿勢角度に一定のスケーリング（例えば 0.4）が施されたものとする．図 6 には，頭部姿勢角（パン方向）の計測値（黒線），及び，生成された制御信号の姿勢角（緑線），及び，その速度成分（青線）の時系列をそれぞれ例示する．計測信号に含まれる微小ノイズや微小運動を除外しつつ，大きな姿勢変化のみを再現していることがわかる．この処理は，1）ダウンサンプリング，振幅のスケーリング，ゼロ値のシフト，時間差分からなる前処理，2）速度のピーク区間の検出と選別，3）最低・最高速度の制約下での波形の整形，からなる．また，チルト運動成分についても，パン運動と同様の手順を踏むが，特に人物の頷きに焦点を当て，連続した運動の後，直立姿勢に回帰するよう制御信号が生成される．こうして得られた制御信号が，映像の再生と同期して逐次 PTU に入力される．

投影画像の生成　スクリーン上に歪みのない画像を投影するため，撮影された画像から人物の顔を中心とする部分画像が切り出され，スクリーンとプロジェクタの相対位置関係，及び，スクリーンの姿勢に応じた写像変換が行われる（図 5(d) は結果の一例）．この変換は，幾つかの座標系の間の写像の系列に分解して，記述することができる．図 7 に，これら座標系の間の関係を図示する．この写像の系列は，スクリーン上の一点，$\mathbf{q}$，を始点とし，PTU 座標系上の点，$\mathbf{p}_{\mathrm{PTU}}$，プロジェクタ座標系上の点，$\mathbf{p}_{\mathrm{Proj}}$，画像座標系上の 1 点，$\mathbf{p}_{\mathrm{Img}}$，ウィンドウ座標系上の 1 点，$\mathbf{w}$，へと至る；$\mathbf{q} \rightarrow \mathbf{p}_{\mathrm{PTU}} \rightarrow \mathbf{p}_{\mathrm{Proj}} \rightarrow \mathbf{p}_{\mathrm{Img}} \rightarrow \mathbf{w}$．スクリーン座標系と PTU 座標系の間の変換は，並進，及び，スクリーンの回転（パンとチルト）として記述される．PTU 座標系とプロジェクタ座標系の変換は，並進と回転，また，プロジェクタ座標系と画像座標系の間には透視投影を仮定する．最後に画像座標系からウィン

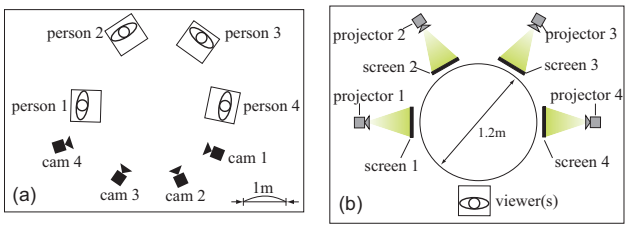

図 3　空間配置, (a) 実際の会話場面, (b) 再現空間
Fig. 3　Spacial configuration of people and devices, (a)meeting room, (b)viewer's room.

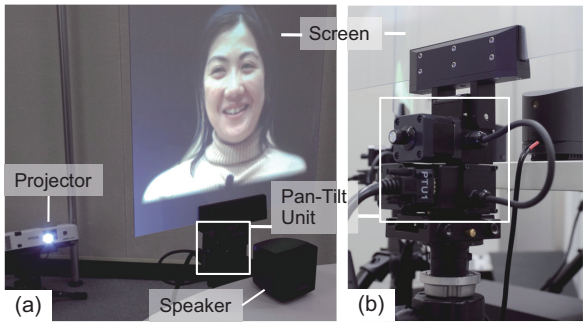

図 4　動的ディスプレイ装置, (a) スクリーン，プロジェクタ，及び，スピーカ, (b) パンチルトユニット
Fig. 4　Dynamic display devices, (a)screen, projector, and speaker, (b)Pan-Tilt Unit from in side view

ドウ座標系への変換は，LCD プロジェクタのレンズシフトを反映した原点シフトが行われる．

この変換において，機器の設置後には定数となるパラメータの値は，事前のキャリブレーションにより推定される．まず，既知の異なる PTU の姿勢について，スクリーンの 4 つの角に対応するウィンドウ座標の値の実測値が取得される．その後，この実測値と変換式により計算される理論値の間の二乗距離の合計の最小化により，未知パラメータが計算される．

4. 実　　験

本稿では，動的スクリーンによる補強表現の効果として，会話者の視線方向，及び，受け手の同定の正確さ，分かりやすさに焦点を絞り，スクリーンの動きあり（動的条件と呼ぶ）と，動きなし（静的条件と呼ぶ）の二つの表示条件を対比した被験者実験を行った．

4.1　使用機器，及び，動作設定

カメラには，Point Grey Research Dragonfly^{TM} を使用し，30 [frame/sec] にて VGA 画像を取得した．PTU には，Directed Perception PTU-D46-17 を用いた．スクリーン（G-Screen Through 型 透過率 97%）のサイズは 45cm×45cm，重量は約 570g であった．プロジェクタには EPSON 1925W (4000[lm] XGA 解像度）を用いた．なお，本実験では，視線方向の表現に焦点を当てるため，パン運動のみが再現され（0.4 倍のスケーリング），チルト姿勢は直立に固定した．また，画像処理による微小な画像欠損等による観測者の注意の散逸を防ぐため，投影画像は背景除去なしの原画像とし，磁気式センサ（Polhemus FASTRAK^{TM}）による頭部姿勢計測値を用いた☆．

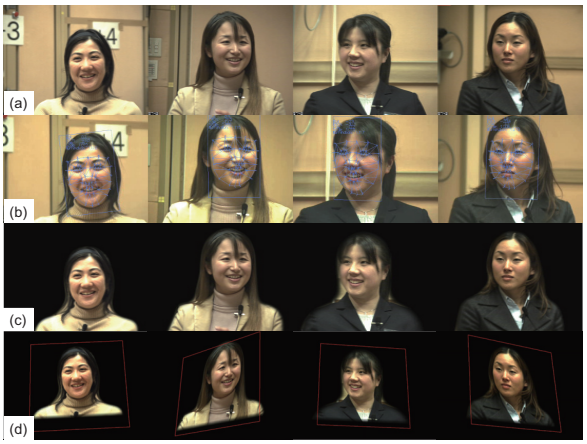

図 5　画像処理の様子, (a) 撮影画像 (b) 顔姿勢追跡, (c) 背景除去, (d) 投影画像（赤枠は説明用）
Fig. 5　Image Processing, (a)Original images, (b)Faces under pose tracking, (c)Background-removed images, (d)Projection images

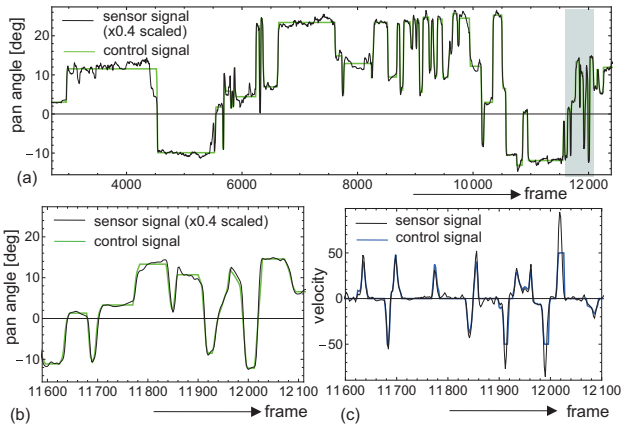

図 6　頭部姿勢時系列（パン角度）の例（×0.4 スケーリング），黒線：計測信号，緑線：PTU 制御信号（角度），青線：PTU 制御信号の速度成分，(a) 会話セッション全体 (∼5.2[min])，(b) 部分拡大（∼17[sec]），(c)(b) の区間の速度成分
Fig. 6　Example time series of pan angle and its velocity, black line is original signal, green line is generated PTU control signal, and blue line (a)whole meeting duration, (b)close-up over grayed interval in (a), (c)velocity of pan angle during same period as in (b)

4.2　実　験　1

動的条件，静的条件における観測者の定性的な振る舞いの違い，及び，観測者の受けた印象を分析した．

データ　4 人の女性（20 代後半）グループ 2 組による 3 つの会話セッションを用いた．会話参加者は，与

☆ ただし，顔領域抽出は顔姿勢追跡による．センサは会話者の頭部にヘアバンドで不自然さなく固定された（図 1，図 5 参照）．

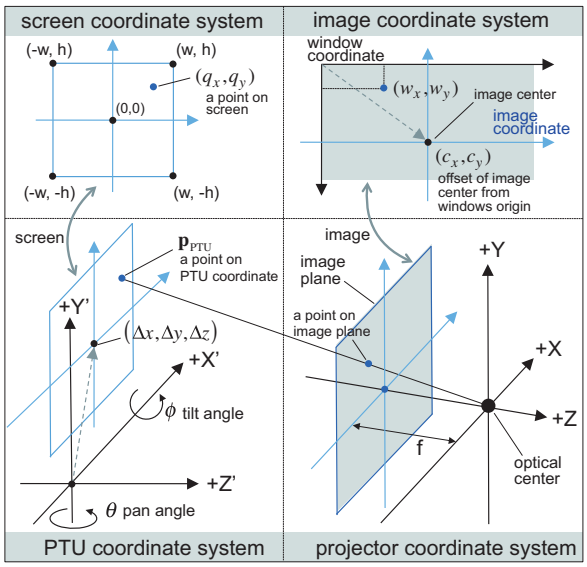

図 7　投影のための座標系
Fig. 7　Projector-Screen coordinate systems

えられた議題について 5 分以内にグループとしての結論を出すよう指示された．グループ 1 による会話セッションでは，C1:「恋愛と結婚は一緒か別か」，C2:「専業主婦に税制優遇措置を与えるべきか否か」，グループ 2 による会話 C3 では「安楽死は合法化すべきか否か」が議題として，実験者により選択された．各会話 C1,C2,C3 の長さは，それぞれ 5.1 分, 5.5 分, 5.2 分であった．

被験者　会話に参加していない 8 名の女性（20 代〜30 代）が本実験に参加した☆．

手続き　図 8☆☆のように，被験者は一人ずつ別個に会話セッションをその開始から終了まで連続して視聴した．各被験者は，動的条件，静的条件，それぞれ 1 つの会話セッションを視聴した．会話，及び，表示条件，その提示順番は，被験者毎に混合された．

質問紙　全ての実験（後述の実験 2 も含む）が終了した後，被験者が質問紙への回答を行った．ここでは会話の流れ，表情，視線，感情，雰囲気について各項目の分かりやすさ，及び，身近さ，のめり込み，楽しさの各項目の印象の強さについて，動的条件と静的条件を比較し，順位をつけた．また，自由記述形式にて実験の感想を記入させた．

定性的分析　視聴中に撮影された被験者の正面ビデオ画像に基づいて，被験者の視線方向（いつ誰を見ていたか）のアノテーションを手動で作成し，視線行動の分析を行った．このアノテーションは，会話，及び，本実験に参加していない 1 名の女性により行われた．図 9 にはスクリーンの物理的運動が，観測者の視線行動に影響を与えたと解釈しうる一例を示す．この場面では，話し手である P2 が，聞き手 P1 の反応を確認しようと P1 の方に視線を向ける場面である．動的条件で視聴した三人中二人が，P2 が視線を P1 に向けた直後に，視線を P2 から P1 へと変化させ，その結果，P1 が P2 に向かって頷く場面を確認している．一方，静的条件で視聴した二人の観測者は，このタイミングで視線を P2 を P1 へと変化させることがなく，P1 の頷きを見逃している．この例では，動的スクリーンによって，表示された人物の視線変化に誘導されて観測者の視線が変化したとの解釈が成り立つ．今後，より多くの事例を含めたより精緻な分析が待たれる．

結果と議論　質問紙への回答を分析した結果，特に「視線」と「楽しさ」の項目について，8 名中 7 名が動的条件をより優位と記入し，片側二項検定によって統計的有意差 ($p < 0.05$) が確認された．また,「会話の流れ」と「身近さ」については，8 名中 6 名が動的条件を優位と回答した．また，自由記述欄からも，8 名中 5 名より,「動くスクリーンによって，視線の方向が分かりやすかった」という回答を得た．これら結果より，動的なスクリーンにより視線方向が分かりやすいという仮説の妥当性が示唆された．また，4 名より，スクリーンの動くノイズ音を指摘された．

4.3　実　験　2

頭部運動の物理的補強表現の効果を定量的に検証することを狙いとして，話者の同定，及び，その受け手の同定の正確さと明瞭さの観点から評価を行った．

データ　実験 1 とは異なり，文脈情報を除くため，被験者は，会話全体から切り出された数秒程度の短い映像（クリップと呼ぶ）を視聴した．会話セッション C1, C2, 及び，C3 の各々について，それぞれ 18 個のクリップを手動で抽出した．各クリップには，一人の話者が他者（一人または複数）に対して話し掛ける/問いかける場面を含む．なお，その相手として複数の可能性がある場面のみを対象とした．例えば，図 3(a) において，人物 2 が人物 1 に話し掛ける場合など，人物の顔向きから一意にその相手が特定できる場面は除外した．また，視聴するクリップの時間順序は，文脈手がかりを排除するためランダムとした．クリップの長さは最小 1.9 秒，最大 14.3 秒，平均 5.4 秒であった．

被験者　実験 1 と同じとした．

質問紙　話し手と受け手の同定の正確さと明瞭さを評価するため，表 1 の設問を用いた．Q1-1 と Q2-1 で

☆ 会話者，及び，会話内容への親和性が高いと思われる同性，及び，近い年代を被験者とした．

☆☆ 本図の撮影に用いたカメラの画角上，図 1(a)(b)，図 3(b) の配置と若干異なって見えるが，実際は同一である．

図 8　被験者実験の様子
Fig. 8　A subject participating in experiments

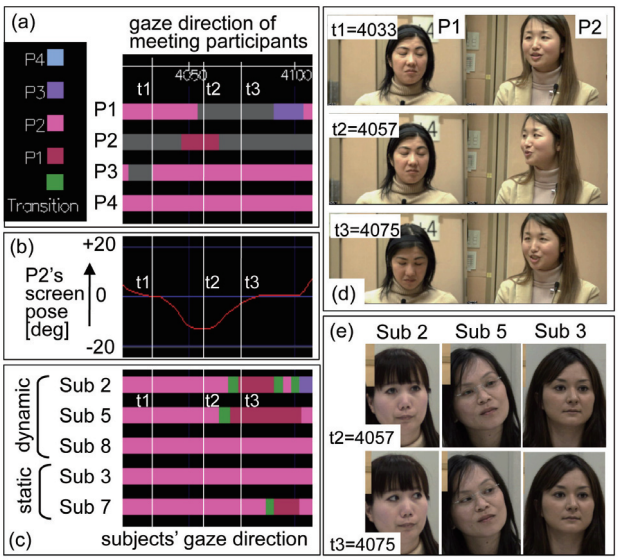

図 9　観測者の振る舞い，(a) 会話者の視線方向の時系列，(b) 人物 P1 のスクリーンのパン角度，(c) 観測者の視線方向の時系列，(d) 会話者 P1 と P2 の様子，(e) 観測者の様子
Fig. 9　Behavior analysis of viewers. (a)time series of gaze directions of meeting participants, (b)time series of P1's pan angle for PTU control, (c)time series of viewers' gaze directions, (d)snapshots of meeting participants P2 and P1, (e)snapshot of viewers

は人物の番号を一つ回答させた．Q1-2 と Q2-2 では，7 ポイントのスケール（-3 (非常に分かりにくい) から+3 (非常に分かりやすい)）より一つを選択させた．

正解データ　設問 Q1-1 については，発話内容の書き起こしより話者が特定され，設問 Q2-1 については，話し手の視線方向より受け手が特定された．なお，受け手については，クリップの時間内で話者が複数の人物に視線を向けていた場合，それら人物を正解とした．

手続き　図 8 のように，8 名の被験者は各々別個に実験に参加した．各被験者は，二つの条件（動的条件，静的条件）について，各一つの会話セッションから抽出された 18 個のクリップを視聴した．各クリップの再生の時間順序はランダムとした．各クリップの視聴後，20 秒間で質問紙への記入を行った．なお，視聴条件と会話セッションの組み合わせは，実験参加者ごとに混合された．

表 1　実験 2 の設問
Table 1　Questionnaire for Experiment 2

	Questions
Q1-1	話し手は誰だったと思いますか？
Q1-2	話し手は誰だったかよく分かりましたか?
Q2-1	話し手が話し掛けていた相手は誰だったと思いますか?
Q2-2	その相手は誰だったかよく分かりましたか?

結果と議論　Q1-2 と Q2-2 については各条件，各クリップ毎に平均の正答率（正解した被験者数／そのクリップを視聴した被験者数）が計算された．なお，受け手の同定（Q2-2）については，質問紙に回答された受け手が，正解の中に含まれていた場合には正解と判定した．明瞭性の設問（Q1-2 と Q2-2）については，各条件，各クリップ毎に平均の評定値が計算された．図 10 には，動的条件，静的条件のそれぞれについて平均正答率，及び，平均の明瞭性評定値のそれぞれ全体平均を記す．図 10(a) より，話し手の同定については両条件とも高い正答率が得られたことが分かる．図 10(b) より，話し手同定の明瞭性についても，両条件で高い評価が得られた（評定値の＋ 2 は,「わかりやすい」に対応）．正答率，明瞭性ともに，両条件間に統計的有意差は無かった（Mann-Whitney の U 検定の 5%水準☆）．なお，話し手の同定は，次の受け手の設問の前提となるものであり，両条件間で差があることはシステムの設計上，想定していない．また，各人物の音声がスクリーン前方の個別スピーカから再生されたため，音の手がかりも有効に作用したと考えられる．

次に，図 10(c) には，設問 Q2-1 の受け手の同定の正答率を示す．動的条件では 81.5%，静的条件では 67.6%と，動的条件がより高い正答率を示し，Mann-Whitney の U 検定の結果，統計的有意差が認められた ($p = 0.019 < 0.05$)．図 10(d) には，設問 Q2-2 の受け手の同定の明瞭性に関する結果を示す．動的条件の評定値がわずかに静的条件を上回るが，T 検定の結果，統計的有意差は示唆されなかった ($p = 0.12 > 0.05$)．以上の結果より，スクリーンの物理的な運動による補強表現が，受け手の同定の正確さに寄与しうることが示唆されたといえる．

5.　結びと今後の展望

本稿では，複数人物の会話場面を表示するための新しいシステムを提案した．本システムは，人物像を投影するスクリーンの物理的な運動として，人物の頭部

☆ 本稿では，データが正規性の条件を満たす場合には T 検定，そうでない場合には Mann-Whitney の U 検定を用いた．

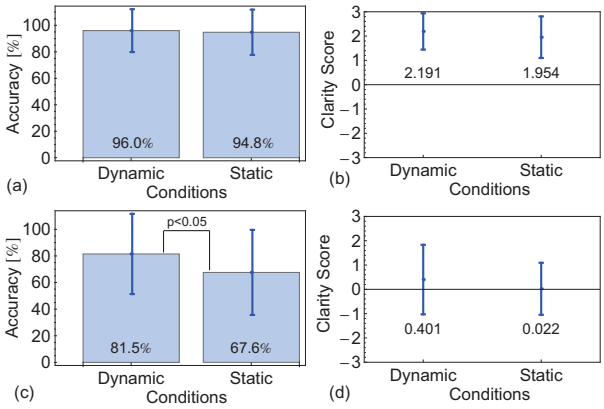

図 10 実験 2 の評価結果．(a)Q1-1: 話し手の同定の平均正答率，(b)Q1-2: 話し手同定の明瞭性の平均評定値，(c)Q2-1: 受け手の同定の平均正答率，(d)Q2-2: 受け手同定の明瞭性の平均評定値

Fig. 10 Evaluation results. (a)Q1-1: accuracy of speaker identification, (b)Q1-2: clarity of speaker identification, (c)Q2-1: accuracy of addressee identification, (d)Q2-2: clarity of addressee identification

運動を表現するという補強的な表現モダリティの提案と実装を特徴とする．この物理的なスクリーンの動きと画像上の動きとの相乗効果により，会話者の視線方向，及び，話し手の話し掛ける相手（受け手）の理解が促進されるという仮説を立て，被験者実験により，視線方向の分かりやすさ，及び，受け手の同定の正確さの点において，仮説の妥当性が示唆された．

今後，背景除去やチルト運動などの効果も検証する必要がある．例えば，背景除去の結果，会話人物の存在感が増強されることが期待される．また，スクリーンのチルト運動によって，会話人物の頷きがより強調されて表現され，観測者の注意を引きつける効果の他，頷きの機能である態度表出や発話強調を伴う社会的相互作用の理解も促進されることが期待できる．加えて，他の再生環境との比較など，より本格的な評価実験を行っていく予定である．

また，本システムは，今後，多地点間の複数人物による実時間コミュニケーションシステムへの発展が期待される．その場合，視線不一致問題と関連して，本稿提案の物理的補強表現が，観測者から見て「話し掛けられる」ことの知覚に寄与しうるものかの検証が必要である．また，カメラの最適配置や，物理系を駆動する際の遅延の問題もあわせて検討が求められる．

参 考 文 献

1) Adalgeirsson, S. O. and Breazeal, C.: MeBot: A Robotic Platform for Socially Embodied Presence, *Proc. ACM/IEEE Int. Conf. Human-robot interaction*, pp.15–22 (2010).
2) Argyle, M.: *Bodily Communication – 2nd ed.*, Routledge, London and New York (1988).
3) Bojsen-Hansen, M.: Active Contours without Edges on the GPU, Technical Report 13, Project Paper for the Course in Parallel Computing for Medical Imaging and Simulation (2010).
4) Guizzo, E.: When My Avatar Went to Work, *IEEE Spectrum*, Vol.47, pp.26–50 (2010).
5) Heider, F. and Simmel, M.: An Experimental Study of Apparent Behavior, *American Journal of Psychology*, Vol.57, No.2, pp.243–259 (1944).
6) Hirata, K., Kaji, K., Harada, Y., Yamashita, N. and Aoyagi, S.: t-Room: Remote Collaboration Apparatus Enhancing Spatio-Temporal Experiences, *Proc. CSCW'08* (2008).
7) Johansson, G.: Visual Perception of Biological Motion and a Model for its Analysis, *PERCEPTION & PSYCHOPHYSICS*, Vol. 14, No. 2, pp. 201–211 (1973).
8) Jones, A., Lang, M., Fyffe, G., Yu, X., Busch, J., McDowall, I., Bolas, M. and Debevec, P.: Achieving Eye Contact in a One-to-many 3D Video Teleconferencing system, *ACM Trans. Graph.*, Vol.28, pp. 64:1–64:8 (2009).
9) Kendon, A.: Some Functions of Gaze-Direction in Social Interaction, *Acta Psychologica*, Vol. 26, pp. 22–63 (1967).
10) Livingstone, M. and Hubel, D.: Segregation of Form, Color, Movement, and Depth: Anatomy, Physiology, and Perception, *Science*, Vol. 240, No.4853, pp.740–749 (1988).
11) Mateo Lozano, O. and Otsuka, K.: Real-time Visual Tracker by Stream Processing, *Journal of Signal Processing Systems*, Vol.57, No.2, pp.285–295 (2008).
12) Maynard, S.K.: Interactional Functions of a Nonverbal Sign: Head Movement in Japanese Dyadic Casual Conversation, *J. Pragmatics*, Vol.11, p.589–606 (1987).
13) Mori, M.: The Uncanny Valley, *Energy*, Vol. 7, No.4, pp.33–35 (1970).
14) Nakanishi, H., Kato, K. and Ishiguro, H.: Zoom Cameras and Movable Displays Enhance Social Telepresence, *Proc. CHI '11*, pp.63–72 (2011).
15) Otsuka, K.: Conversation Scene Analysis, *IEEE Signal Processing Magazine*, Vol.28, No.4, pp.127–131 (2011).
16) Otsuka, K., Araki, S., Ishizuka, K., Fujimoto, M., Heinrich, M. and Yamato, J.: A Realtime Multimodal System for Analyzing Group Meetings by Combining Face Pose Tracking and Speaker Diarization, *Proc. ACM ICMI'08*, pp.257–264 (2008).
17) Pham, V., Vo, P., Hung, V.T. and Bac, L.H.: GPU Implementation of Extended Gaussian Mixture Model for Background Subtraction, *Proc. IEEE RIVF*, pp.1–4 (2010).
18) Sakamoto, D., Kanda, T., Ono, T., Ishiguro, H. and Hagita, N.: Android as a Telecommunication Medium with a Human-like Presence, *Proc. ACM/IEEE HRI*, pp.193–200 (2007).
19) Sellen, A.J.: Speech Patterns in Video-Mediated Conversations, *Proc. CHI*, pp.49–59 (1992).