情報処理学会
グループウェアとネットワークサービス研究会
2007年11月8日, 9日

遠隔ビデオコミュニケーションシステムのための仮想共有面の実装方式

平田 圭二　原田 康徳　高田 敏弘　青柳 滋己
白井 良成　山下 直美　大和 淳司　梶 克彦

NTT コミュニケーション科学基礎研究所

あらまし　本稿では，実用的な遠隔ビデオコミュニケーションを実現するための実装方式を検討する．シームレス性，実用性，柔軟性を兼ね備えるためには，多種多様な入出力デバイスの利用，入出力デバイスの自由な配置，多地点間の接続という要件を統合的に満たす必要がある．そのため，我々は入出力デバイスとそれらの間の接続を抽象化する仮想共有面という仕組みを提案する．仮想共有面を実装する際に検討すべき課題として，カメラ映像のエコー抑制，映像データの部品化，メディア情報に時空間のタグを付与するメタフォーマットなどがある．t-Room は仮想共有面に基づく遠隔ビデオコミュニケーションシステムであり，ユーザの周囲をディスプレイで囲むような円柱状の構造を持っている．現在，その実装を進めており，本稿で検討した方式の評価は今後の課題である．

キーワード　同室感，入出力デバイス抽象化，映像部品化，t-Room システム

1　はじめに

近年ブロードバンドが普及し，大画面ディスプレイ，高画質ビデオカメラ，小型高性能マイクやスピーカなどの入出力デバイスが低廉化してきた．このような背景から，複数の大画面を用いた遠隔コミュニケーション環境の研究開発や商品化が盛んになっている．従来のビデオ会議で用いられるディスプレイやスクリーンは，遠隔地ユーザの居る空間への覗き窓のような役目を果たしており，遠隔地ユーザとの共有面を作り出す物であると同時にユーザの視界を制限する物でもあった．近年は，その覗き窓を大型化，高精細化し，視界の制限を緩和するアプローチが多く見られる [2, 4].

我々のアプローチは，遠隔地のユーザとあたかも同じ部屋に居るような感覚を共有し，円滑な共同作業を実現することである．遠隔地のユーザと円滑な共同作業を実現するには，映像や音を介してさまざまなアウェアネスを共有することが望ましい．従来のビデオ会議の多くは，目の前にディスプレイやスクリーンがあるため，あるユーザの周囲に見える物，聞こえるものすべてが，他の遠隔地のユーザにも見えたり聞こえたりするわけではない．したがって，ディスプレイやスクリーンの向こう側に見えるユーザとは，同じ部屋にいる時に得られるような映像や音を共有することが難しい．

将来，ビデオ会議システムを初めとする遠隔ビデオコミュニケーションシステムが，さらに日常生活の中で頻繁に，あるいは常時利用されるようになるには，この “あたかも同じ部屋にいる感覚” が重要であると考える．その上で，実世界とシステム環境がシームレスに接続でき，実世界の多様な制約や需要に応えるだけの実用性と柔軟性を兼ね備えている必要があろう．シームレス性，実用性，柔軟性の向上を目指し，我々は共有面の拡張である仮想共有面を提案する．そして，仮想共有面の実装上の課題と解決法を議論する．

2　仮想共有面による同室感

2.1　同室感

同室感とは遠隔地ユーザとあたかも同じ部屋にいるような感覚のことであり [3]，同じ部屋に居るユーザが共有すべき映像や音を各ユーザに適切に提示することで得られる．質の高い同室感を得るには，ディ

スプレイやスピーカで囲まれた空間 (部屋) を作り，あるユーザの映像や音に関する周囲の状況を，一貫性をもって遠隔地で再構成することと，いずれのユーザにとっても対称的にその状況を再構成することが重要である [10].

現在の技術では，ユーザ間で共有される情報を立体的に一貫性と対称性をもって再構成することは困難なので，我々は，二次元の面に限定してユーザ間で情報を共有する．そのような面を**共有面**と呼ぶ [3]. 音響に関しても，ユーザを取り囲む複数スピーカ，複数マイクによって共有音場を構成する．

2.2 共有面による同室感

先行研究で提案した共有面の構成法 [3] では，ディスプレイとビデオカメラを 1 対 1 に対応させ，対向したカメラでそのディスプレイ画面を撮影し，お互い遠隔地に表示する (図 1). 図ではユーザを取り

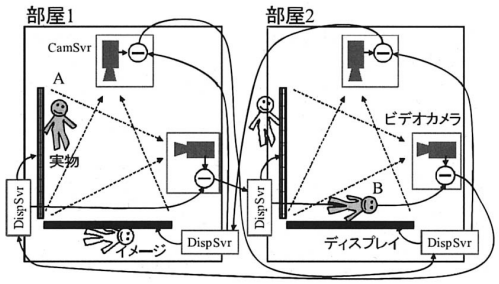

図 1: 共有面の構成法

囲むディスプレイを 1 部屋あたり 2 枚，直角になるように配置した．プロセス DispSvr (display server) がディスプレイを制御し，CamSvr (camera server) がビデオカメラを制御する．

部屋 1 の人 A は部屋 2 の対応する場所にそのイメージが表示され，部屋 2 の人 B についても同様である．この時，人 A，B はお互いに，二人が同じ部屋に居る時とほぼ同じ方向と距離の所に見えている筈である．人 A，B がディスプレイを背にしてできるだけディスプレイの表面近くに立っている限り，二人が部屋の中を自由に移動してもこの方向と距離の関係は常に成立している．ここで，図 1 のディスプレイは共有面として機能し，部屋の中に居るユーザに同室感を提供している．共有面の表面付近では一貫性と対称性の高い情報共有が可能だが，表面から離れるほど一貫性と対称性は低くなる．

2.3 エコー抑制によるカメラ映像の映像部品化

図 1 の共有面のように，ディスプレイとビデオカメラをループ状に直結すると映像のエコーが発生してしまう．そのため図中 CamSvr 内にあるように，ビデオカメラで撮影した映像からユーザの背景の映像信号を減算しなければならない．先行研究 [3] では，ディスプレイの前とビデオカメラの前に偏光フィルムを置き光学的に背景映像のエコーを抑制した [9].

CamSvr 内にあるカメラ映像のエコー抑制機能を検討する (図 2). 差分するディスプレイ表示映像は，

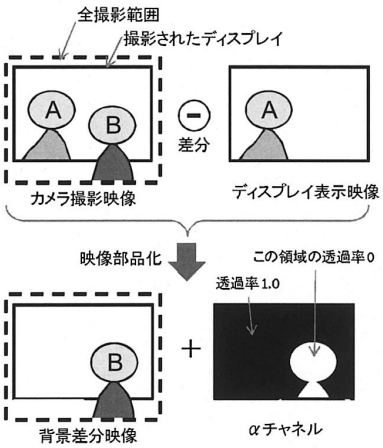

図 2: カメラ映像の部品化

対応するディスプレイが表示している共有面上の出力領域から得る．ディスプレイの前に立った人物や置かれた物のみが，α チャネル (透過率) 付き背景差分映像として計算される．このエコー抑制機能によって，カメラから得られた映像データは共有面のどこにでも幾つでも自由に α ブレンドすることができる (映像部品化).

2.4 仮想共有面

第 1 章で触れた実用的な共有面を実現するために満たすべき要件の内，多種多様な入出力デバイスの利用，入出力デバイスの自由な配置，多地点間の接続の 3 点は重要性が高いと考えられる．ここで入力デバイスや入力データには，ビデオカメラ，PC のディスプレイ出力，携帯電話，JPEG，MPEG，Flash などの画像ファイルがある．出力デバイスや出力装置には，ディスプレイ，プロジェクタ，携帯電話などがある．一般にビデオカメラは，様々な解像度やフレームレートを持ち，ディスプレイは様々なサイズや解像度を持つ．

これら 3 つの要件を統合的に満たすため，我々は**仮想共有面**を導入する (図 3)．仮想共有面とは，そ

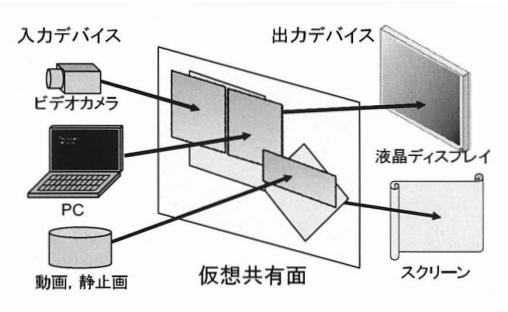

図 3: 仮想共有面による入出力デバイスの抽象化

れに対する入出力関係を抽象化・仮想化した共有面であり，映像部品化された入力映像と表示出力のマッピングのために用いられる[1]．仮想共有面を介してアクセスする抽象化された入出力デバイスは，その実世界における物理的な位置，サイズ，解像度，その他の属性とは独立なオブジェクト (入力領域，出力領域) として扱うことができる．仮想共有面の上に置かれた入出力領域の位置関係に従って入出力オブジェクトが接続される．さらに，入出力領域はその位置や形状を自由に設定でき，アフィン変換 (縦横比変更，拡大縮小，台形補正，回転など) や入力領域の重畳も可能である．

簡単な例として，ディスプレイとカメラの画角が全射影で対応していない場合を示す (図 4 (a))．ここでは DispSvr と CamSvr は省略してある．ビデ

[1]ここでは，世界全体で 1 つの仮想共有面が存在するようなモデルを考えているが，各地点ごとに別々の仮想共有面が存在するモデルを考えることも可能である．しかし，それは本稿で議論するモデルと本質的に等価である．

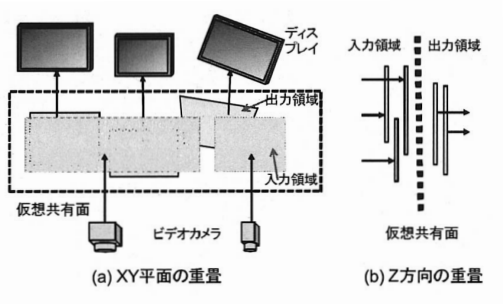

図 4: 入出力領域の重なり合い

オカメラ映像が仮想共有面のある XY 座標領域にマップされ，ディスプレイが表示を担当する XY 座標領域も仮想共有面上にマップされる．これら XY 座標領域は任意の四角形が指定できる．図 4 (a) の例では，左の広角カメラ映像の部分を異なる 2 台のディスプレイが表示している．また，左の広角カメラ映像の部分と右のカメラ映像の部分は，一番右のディスプレイが合わせて表示している．

次に，図 4 (b) は仮想共有面にマップされた入力領域と出力領域を Z 方向から見た例である．映像部品化されたビデオカメラ出力を入力領域として，Z 軸方向に任意段だけ重畳させることができる．各入力領域の透過度やそのレイヤ順で表示出力結果は変化する．一方，出力領域にも表示を担当するディスプレイがマップされるが，我々の提案する仮想共有面では，そのレイヤ順は表示出力に影響を与えないとしている[2]．つまり，表示すべき内容は 出力領域の XY 平面上の位置にのみ依存し，その Z 方向の位置には依存しない．

ある地点との接続を動的に追加したい時は，新しい入力領域を動的に重畳する処理を行えばよい．逆に，ある地点との接続を中止したいような時は，ある入力領域 (映像部品) を動的に削除するという操作を行えばよい．これより，多地点間を自由に接続・切断するビデオコミュニケーション環境の構成が可能となる．

音響に関しても同様に仮想共有面の考え方を導入することができる．抽象化された集音器 (マイク) と拡声器 (スピーカ) を共有面上に配置し，遠隔地のユーザどうしで，ユーザの周囲の音響環境を一貫性をもって遠隔地で再構成する．

[2]もう 1 つの設計解として，DispSvr 内に入力領域を重畳させる機能を持たせる方式も考えられるが，これは本稿で議論している方式と本質的に等価である．

3 遠隔コミュニケーションシステム t-Room

3.1 各プロセスの役割と諸元

現在我々は，仮想共有面に基づく遠隔コミュニケーションシステム t-Room の実装を進めている [6, 5]. 図 5 に示すように，t-Room を構成する入出力デバイスには，複数のディスプレイ，そのディスプレイを対向から撮影するビデオカメラ，ディスプレイの表面付近を録音するマイク，音場を生成するスピーカなどがある．負荷分散のため，機能ごとにサーバ/プロセスを設け，サーバ/プロセス間で直接データのやりとりをするよう設計した．インタフェースプロセスは，t-Room 間のインタフェースを管理する．出力デバイスマネジャは，DispSvr や SpkSvr (speaker server) 等の出力系サーバ群を統括し，入力デバイスマネジャは，CamSvr や MicSvr (mic server) 等の入力系サーバ群を統括する．

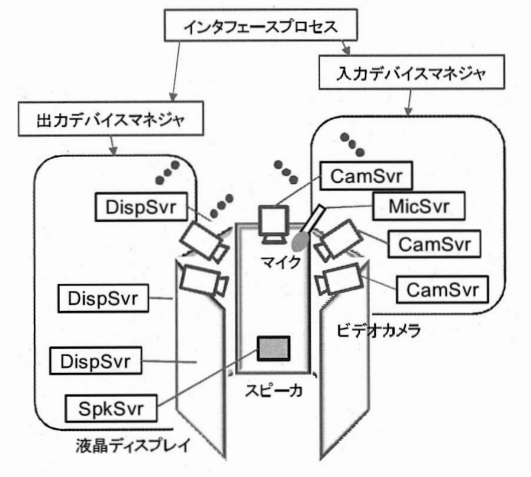

図 5: t-Room のシステム構成図

次に t-Room2.0 (図 9 右) の諸元を述べる．まずビデオ系では，SONY 製ビデオカメラ HDR-HC3 の DV 出力 (720x480) を IEEE1394 経由で CamSvr が動作している PC へ入力する．その PC の CPU は Intel Core 2 Extreme, 2.93GHz, メモリ 2GB, OS は Windows XP Professionsl SP2 である．グラフィクス処理は Qt/OpenGL を使用し，CamSvr で Motion JPEG (最大 20 fps) に変換し，TCP/IP 上の NZAM フォーマットで DispSvr に送信する．DispSvr が動作している PC は CamSvr 用 PC とほぼ同じ仕様であり，Sharp 製 65 インチ液晶ディスプレイ (1920 x 1080) に映像を表示する．次に音響系では，ワイヤレスマイク 4ch を 16bit/44.1KHz リニア PCM でエンコードし UDP/IP で SpkSvr に送信する．SpkSvr では受信した音 4 ch を 2ch ずつに集約して，2 台のスピーカから出力する．ネットワークは，現在 B-Flets ビジネス 100Mbps 1 本で厚木地区と京阪奈地区の t-Room を接続している．

3.2 時空間のタグを付けたメディア情報

ビデオカメラ等入力デバイスから得られた映像を，仮想共有面に基づく遠隔ビデオコミュニケーション環境で再利用性の高い部品として扱えるようにするため，オブジェクト指向の考え方に基づき，部品自体にその部品の属性を持たせる．そのため，我々は映像などメディア情報のためのメタフォーマット (NZAM) を提案する (図 6)．大まかに言えば，NZAM は，そのメディア情報を生成したイベントが生起した場所と時刻を含むヘッダ部と，メディア情報自身を含むデータ部から成る．ここで想定しているメディア情報は例えば MPEG, JPEG, Flash, MIDI, XML などである．ヘッダ部が時刻の情報を含むことで仮想共有面は時間軸方向に拡張され，場所だけでなく時刻の情報に基づくメディア情報の再利用が可能となる．我々はこれを時空間コンテンツ化と呼ぶが，その詳細は文献 [7] を参照のこと．

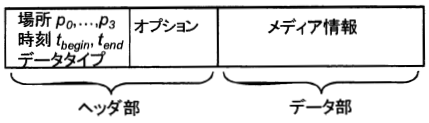

図 6: NZAM の構造

3.3 サーバ間通信

映像データ処理の一例として Motion JPEG (MJPEG) が CamSvr から仮想共有面を経由して DispSvr に送られる様子を順を追って説明する (図 7)．メディア情報が MJPEG の場合，ヘッダ部には場所と時刻の情報に加えオプション情報としてクリッピング点 $C_0 \sim C_3$ の情報が含まれ，データ

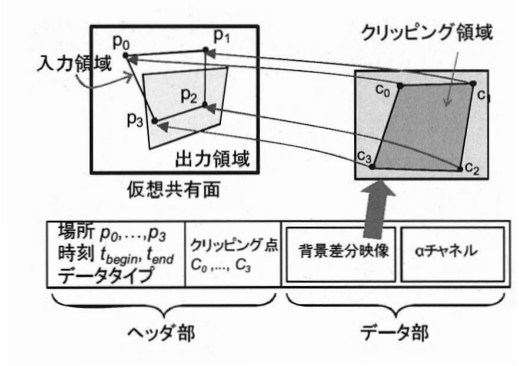

図 7: NZAM による Motion JPEG データの時空間コンテンツ化

部には背景差分画像と α チャネルの 2 つの MJPEG が含まれる.

ステップ 1: 各 CamSvr, DispSvr は初期設定ファイルから仮想共有面のどの領域を担当しているかの情報を読み込み, DispSvr はさらに自身のディスプレイサイズを読み込む. CamSvr では仮想共有面に出力する四角の領域をクリッピングするために, $C_0 \sim C_3$ の 4 頂点をあらかじめ指定しておく.

ステップ 2: CamSvr は, 仮想共有面の範囲 $p_0 \sim p_3$ (入力領域) とクリッピングの座標 $C_0 \sim C_3$ をヘッダ部に含み, 映像エコー抑制器の出力である 2 つの MJPEG をデータ部に含むような NZAM データを出力する. この時, 映像の解像度はデータ部の MJPEG 自体に含まれている.

ステップ 3: 最後に, NZAM データを受信した DispSvr は, 自身の仮想共有面上の出力領域とディスプレイサイズを加味し, NZAM に含まれる MJPEG データのクリッピングを行い, 映像の拡大・縮小, 一部切り出し等の変形を行い, 当該ディスプレイに表示する.

DispSvr が担当するディスプレイにある場所・時刻の NZAM データを表示したいとすると, その要求はインタフェースプロセスとマネジャを経由して該当する CamSvr に届く. 表示される NZAM データは, マネジャやインタフェースプロセスを経由せず, 上述したように CamSvr から要求元の DispSvr に直接返信される. 仮想共有面は, 各サーバ間で NZAM データをやりとりし, サーバ内で配置情報や解像度に関する情報を変換するための仕組みであり, 実装としてこれらサーバどうしが仮想共有面という計算資源を実際に共有しているわけではない.

3.4 モノリスの円柱状配置

ユーザに同室感を提供するため, t-Room ではディスプレイを内側に向けて円柱状に並べている. 2.2 節で述べたように共有面 (ディスプレイ) の表面と表面近傍で物体の情報が共有される. 同室感を得るためには, 2 次元である共有面を効果的に配置して, 擬似的な奥行き感・3 次元感覚を創出する必要がある. そこで, 対向する共有面が生じ, ユーザがディスプレイで囲まれるような配置が望ましいと考え円柱状の配置を選択した.

t-Room システムはモノリスというユニットで構成される (図 8). モノリスとは, カメラ 1 台, ディスプレイ 1 台, サーバ用 PC 等から成る t-Room 構成ユニットであり, t-Room ハードウェアはモノリス単位で組み上げられる.

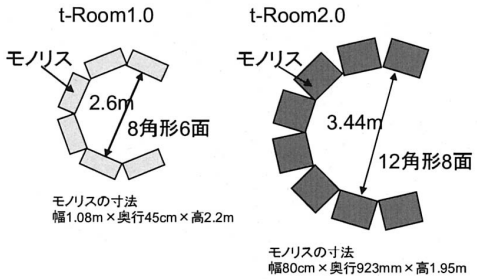

図 8: t-Room 1.0 と 2.0 を上から見た図

実機の写真を図 9 に示す. t-Room1.0 では 40 インチ液晶ディスプレイを備えるモノリスを 8 角形の 6 辺に, t-Room2.0 では縦型 65 インチ液晶ディスプレイを備えるモノリスを 12 角形の 8 辺に配置し

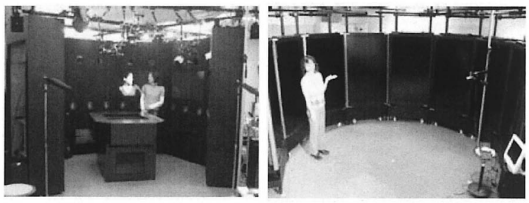

図 9: t-Room 1.0 と 2.0 の外観

た．t-Room1.0 では中央にテーブルが配置されており，テーブルの上面は仮想共有面の一部となっている．テーブル上面に組み込まれたディスプレイと，テーブルを撮影するために天井に設置されたカメラによってテーブル上に置かれた書類等のオブジェクトや指差しなどのジェスチャが共有される．このような配置は会議など資料情報を共有する際に有効であろう．一方，t-Room2.0 の大型ディスプレイによる構成は，ゴルフやダンスのレッスンなど全身の動作を遠隔地と共有する際に有効であろう．

我々はこれまで入出力デバイスの接続関係を仮想共有面として抽象化してきたが，仮想共有面では円柱のトポロジを表現することができない．モノリスを円柱状に配置した場合は，仮想円柱面で入出力デバイスの接続関係を抽象化する方が望ましい．しかし実際には，テーブルや共有ホワイトボードのような物体や面も同時に利用されることが多いと思われるので，仮想共有面と仮想円柱面が共存したモデルを考える必要があろう．

4 おわりに

多種多様な入出力デバイスの利用，入出力デバイスの自由な配置，多地点間の接続という課題を統一的に解決するために仮想共有面を導入した (2.4 節)．仮想共有面により実世界とシステム環境をシームレスに接続し，実世界の多様な制約や需要に応えるだけの実用性と柔軟性を両立させることができるか否かは，今後，さまざまな観点から評価していかねばならないと考えている．

同室感を得るにはモノリスを円柱状に配置することが望ましいと述べた (3.4 節)．実世界の制約や応用への特化を勘案し，円柱のバリエーションの配置法や，円柱以外のモノリス配置法についても検討を加える必要があろう．

本稿では紙面の制約で議論できなかったが，(1) ネットワークや CODEC 等の遅延への対処 [8]，(2) 分散したサーバの同期再生 [1]，(3) スケーラビリティと接続制御 [11]，(4) 設置やシステム設定の容易な t-Room システムの開発は大きな課題である．特に (4) は，実用性のためのみならず，遠隔ビデオコミュニケーション環境の研究開発にとっても非常に重要であると考える．簡単で柔軟に様々な形態のシステム構築が可能となれば，さまざまな実験や評価も効率良く実施できるからである．そのために，整備されたソフトウェア群 (ミドルウェア) による支援は必須である．これは，GUI 研究の長足の進歩には，ウィンドウシステムやツールキットなどのソフトウェア群の高機能化が必須であったことと同様である．今後とも，ハードウェア，ソフトウェア両面からのバランスの良い研究開発を進めていこうと考えている．

文　献

[1] 阿部博信, 福田雅裕, 山田淳, 松本佳宏, 重野寛, 岡田謙一, インターネット映像配信サービスのための映像と付加情報の同期配信方式, 情報処理学会論文誌, Vol.46, No.2, pp.525–535 (2005).

[2] Cisco, TelePresence, **http://www.cisco.com/web/JP/solution/uc/telepresence/**, 2007 年 10 月 5 日アクセス.

[3] 原田康徳, 同室感通信, インタラクティブシステムとソフトウェア *VI* —日本ソフトウェア科学会 *WISS '98*— (安村通晃 編), レクチャーノート/ソフトウェア学 21, pp.53–60, 近代科学社.

[4] Hewlett-Packard, Halo Collaboration Studio, **http://www.hp.com/halo/**, 2007 年 10 月 5 日アクセス.

[5] Keiji Hirata, Yasunori Harada, Toshihiro Takada, Shigemi Aoyagi, Yoshinari Shirai, Naomi Yamashita, and Junji Yamato, The t-Room: Toward the Future Phone, *NTT Technical Review*, Vol.4, No.12, pp.26-33 (2006).

[6] 平田圭二, 未来の電話を考える - 遠隔コミュニケーションシステム t-Room, *NTT* 技術ジャーナル Vol.19, No.6 (2007 年 6 月号) pp.10-12.

[7] 梶克彦, 平田圭二, 社会的インタラクションのコンテンツ化のためのアーキテクチャ, GNWS 2007, 情報処理学会.

[8] 小峯隆宏, 勝本道哲, 丹康雄, 多地点遠隔講義で自然なコミュニケーションを実現する DV リアルタイム処理機構の開発, 情報処理学会論文誌, Vol.46, No.2, pp.536–545 (2005).

[9] J. C. Tang and S. L. Minneman, VideoDraw: A Video Interface for Collaborative Drawing, In *Proc. of CHI '90*, pp.313-320.

[10] Naomi Yamashita, Keiji Hirata, Yasunori Harada, Toshihiro Takada, Shigemi Aoyagi, Junji Yamato, and Yoshinari Shirai, Effects of Room-sized Sharing on Remote Collaboration on Physical Tasks, *IPSJ Digital Courier*, Vol.3 (2007). To Appear.

[11] 吉内英也, 星徹, 武田幸子, SIP による集中制御型会議システムの開発, 情報処理学会論文誌, Vol.46, No.1, pp.51–59 (2005).