

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

梶 克彦† 平田 圭二†

†NTT コミュニケーション科学基礎研究所
〒 619-0237 「けいはんな学研都市」精華町光台 2-4
E-mail: kaji@cslab.kecl.ntt.co.jp

あらまし 我々はこれまでに遠隔地との同期的なジェスチャーや目配せ・体の向きなどの社会的インタラクションを支援するシステムである t-Room を構築してきた。今回は、遠隔地の同期的通信に加え、過去のインタラクションを利用した非同期インタラクションを実現するため、社会的インタラクションのコンテンツ化と配信のためのアーキテクチャを構築する。まず社会的インタラクションをコンテンツ化のために、記録装置と再生装置の同一化と、時間によるインデキシングによってインタラクションの構造化をすることを挙げる。また t-Room において実装した社会的インタラクションのコンテンツ化・配信アーキテクチャの実装について述べる。さらに本アーキテクチャを用い、過去の再現や引用、追体験を実現する。

キーワード 時空間コンテンツ, 時空間構造化, 追体験, 同期・非同期の融合

1 はじめに

グループを形成しお互いが協調しあうインタラクションは、人類が経済的・文化的成長を遂げて行く過程において、重要な役割を担っていたといえる [1]。現在ではインターネット技術の発達により、電子掲示板やテレビ会議システム、Wiki 等によって、遠隔地・複数人のインタラクションを容易に行うことができるようになった。

Tang ら [2] は遠隔地とのインタラクションを円滑に行うためには、社会的インタラクションが重要であると指摘している。社会的インタラクションとは、ジェスチャーや目配せ・体の向きなどの身体性を通して人と人がコミュニケーションすることであり、社会的インタラクションが対人コミュニケーションをスムーズにする効果がある。このような背景から、遠隔地との社会的インタラクションを支援する研究 [3, 4, 5, 6, 7, 8] が進められている。

しかし従来の社会的インタラクションシステムには、それを蓄積し、利用するという観点が欠けている。社会的発展に大きく寄与する社会的インタラクションを参照可能にするためにコンテンツ化を行わなければならない。

インタラクションを循環させることで、知識やコンテンツの生産性・創造性を向上させる可能性がある。社会的インタラクションの履歴を踏まえ、さらに社会的インタラクションを再生産していくことができれば、そのインタラクションがより重要性を増すことになるだろう。Synvie [9] やニコニコ動画 [10] では、動画に対するコメントや説明を共有することで、新たなコメントを引き出すことに成功している。

過去の社会的インタラクションの雰囲気を感じるためには、体験者がそのインタラクションにに入り込み、そのプロセスを体験するような感覚が必要になるだろう。角ら [11] は人と環境の間のインタラクションを記録・蓄積することでインタラクションのプロトコルの大規模辞書の構築を目標にしてるが、過去のインタラクションのプロセスの再現性が低いことが問題である。例えば多くの人の心をつかむ講演の方法を知るためには、手本となる講演で述べられた文章を提示するだけではなく、その講演での声の大きさ、ジェスチャー、観衆の反応や雰囲気などを踏まえなければならない。

これらを踏まえて、社会的インタラクションを記録・蓄積しコンテンツ化し、そのインタラクションを追体験するための要件を整理し、実際に遠隔コミュニケーションシステムである t-Room に対して社会的インタラクションのコンテンツ化を行うアーキテクチャを設計し、プロトタイピングした。

2 社会的インタラクションのコンテンツ化

本章では、遠隔地との社会的インタラクションを支援するシステムが、そのシステムを利用する際に起きるインタラクションを記録し、それをコンテンツ化・利用するための要件を述べる。

2.1 時間によるインデキシング

我々は日常的に、電子メール・掲示板等を介して非同期的な遠隔のユーザとのインタラクションを行っている。これらのシステムでは過去のインタラクションを自由に参照・引用してメッセージを返信することができる。参照・引用のためにはインタラクションの一部を適切に探し出すことが必要であるが、その際時間軸に沿って探索することが一般的である。そこで、社会的インタラクションシステムに時間軸を導入し、インタラクションに時間によるインデキシングを行うことで、昔のいつのインタラクションも探し出し利用可能にする。これによりコンテンツ化に必要な参照可能性が満たされる。

本研究では、遠隔地とのインタラクションに時間軸を導入することで時間的・空間的に広がりのあるコンテンツとなることから、社会的インタラクションのコンテンツを時空間コンテンツと名付ける。

また遠隔地との同期通信を「現在時刻における時空間コンテンツを利用する」処理であるとみなすことで、遠隔地との同期的な通信と非同期的である時空間コンテンツ利用を、統一的なアーキテクチャで実現することができるであろう。

2.2 記録装置と再生装置の同一化

社会的インタラクションの特徴である雰囲気や盛り上がり、細かなニュアンスを感じ取るためには、過去の出来事に参加していない第三者が、その出来事を後から追体験することである。環境に入り込み、ジェスチャや視線などを含めたインタラクションを追体験することで、その場の雰囲気を臨場感をもって感じ取ることができる。そこで、インタラクションを記録するシステムと再生するシステムを同一化し、追体験を可能にする。

高田ら [12] のシステムは記録と再生を同一化した。ある一つの平面をインタラクションの共有面ととらえ、その平面の記録と再生をプロジェクタとカメラを用いて同時に実現することにより、何段階にも及び引用が可能なビデオメッセージシステムを実現している。このように記録装置と再生装置を同一システム上で実現することで、過去のインタラクションの引用を何度も繰り返すことが可能になる。電子メールのやりとりのように、インタラクションは一度きりでは終わらず、何度も積み重ねられることが多い。

2.3 社会的インタラクションの構造化

社会的インタラクションを記録するデータは、ビデオ、音声、センサ情報などのデジタルメディアであるが、これらの素材同士の時間的・空間的關係を適切につなぎ合わせておかなければ、後に利用することが困難になる。Greenhalgh ら [13] は、仮想環境でおきる出来事の記録

と利用のために、「時間的属性」「空間的属性」「表現的属性」からなる Temporal Links を提案しており、この手法は遠隔コミュニケーションにおいても適用可能である。

あるメディアをコンテンツにするための条件として、検索・加工が可能である点が挙げられる。2.1 節において、時間軸を導入することが社会的インタラクションの検索の手がかりとなると述べたが、それだけでは不十分である。例えばどの時間からどの時間までがひとまとまりのインタラクションであるのか、またそのインタラクションは何を目的としているのか等の構造がインタラクション検索や加工に必要となる。

そのようなインタラクションの構造の中には、同期的なインタラクションの最中にシステムが認識可能であるものや、リアルタイム性が保証できない処理により認識されるものがある。しかし、後に時空間コンテンツを利用する際のためには、同期的なインタラクションには利用できなくても、計算量が多く処理時間のかかる解析や、人手による情報付与手法も積極的に利用することが望ましいだろう。

2.4 時空間コンテンツ操作の記録と再生

録画装置と再生装置が一体化したシステムでは、時空間コンテンツ操作を行うこと自体が時空間コンテンツとして保存されるため、過去に行った時空間コンテンツ操作をそのまま再現することが可能になる。そのため通常のビデオ操作である再生・一時停止・早送り・巻き戻しに加えて、システム操作自体を再現するかどうかについて考慮しなければならない。

例えば時空間コンテンツの操作を忠実に再現する場合、図 1 のようにシステム A の時刻 $t_0 + \delta t$ において t_0 から行われたインタラクション α の様子を再現する操作をシステム A において実行すると、 δt 後の $t_0 + 2\delta t$ には t_0 を再現する操作が再び出現し、その操作を再現するとさらに 2 段目のインタラクション α の再現が開始される。一方、システムが記録していた操作は実行せずに単純に過去の再現を行いたい場合もあるだろう。そのため時空間コンテンツ操作を行うシステムには、間接的な操作を再現するかどうかをユーザが詳細に決定する仕組みを備える必要がある。

このような時空間コンテンツの操作は複雑になるため、頻繁に利用される時空間コンテンツの操作をテンプレートとして利用可能にするべきである。

3 時空間コンテンツ処理のためのアーキテクチャとその実装

t-Room はビデオと複数画面を用いたテレコミュニケーション環境であり、多地点間の身体的コミュニケーション

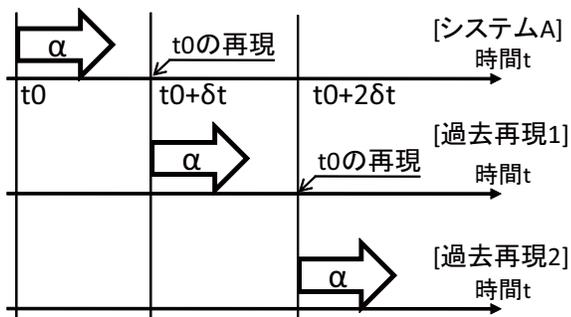


図 1: 時空間コンテンツ操作の再現



図 2: 大型ディスプレイによる全身の動作の共有

ンを実現するために仮想共有面 [8] という概念を導入している。仮想共有面とは、遠隔地との共有空間を面に限定し、その面を記録・再生するための解像度や配置の違いを吸収するための概念である。

山下ら [14] により、t-Room の等身大ジェスチャの共有が同期的遠隔コミュニケーションに有効であることが指摘されている。また t-Room では、カメラやディスプレイ等の数や配置を柔軟に構成することができる。例えば図 2 のように大型ディスプレイを配置することで、全身の動作を共有することが可能になる。

本章では t-Room における同期的遠隔コミュニケーションを時空間コンテンツ化し、それを利用するためのアーキテクチャと実装について述べる。まず t-Room におけるコミュニケーション構造の記述形式について述べ、次にその構造化と配信のアーキテクチャについて説明する。また t-Room に蓄積された時空間コンテンツを利用した、同期・非同期が融合されたインタラクションを例示する。

3.1 t-Room における時空間コンテンツ表現

メディアのコンテンツ化のためには、そのメディアの構造を表現しなければならないが、メディアとその構造情報を切り離すことで、コンテンツ検索や利用のためのアーキテクチャを単純化させることができる。そこで、

```

<done>      [実行した場所と時間]
<exec troom="A" time="yyyy/MM/dd hh:mm:ss">
  <operation>
    <time begin="yyyy/MM/dd hh:mm:ss"/> [時間情報]
    <place troom="B"/> [場所情報]
    <presentation layer="B" transparency="1"/> [表示情報]
  </operation>
  <operation>
    ...
  </operation>
</exec>
</done>

```

図 3: t-Room の操作履歴の時空間構造

ビデオカメラの映像やマイクの音声など、t-Room 内で獲得されるデータをコンテンツ素材、システムの操作履歴や解析処理の結果、人手により付与されるラベルなどの情報を時空間構造に分離する。

t-Room では、入力デバイスであるカメラやマイクのデータを NZAM と呼ばれる統一形式に変換し、コンテンツ素材化する。2.3 節で述べたように、インタラクションは時間的・空間的・表現的な情報を持った構造として記述することが必要であるため、映像・音声などのデータに対して、そのデータの生起した時間、仮想共有平面中の座標、データの種類等をヘッダに付加する。NZAM は仮想共有平面に何枚でも重ね合わせることが可能になる。これにより、例えば複数の過去を重ね合わせて再生しつつ、同時に多地点の遠隔通信を行うといったような、同期・非同期の融合したインタラクションを実現する。

t-Room では、操作履歴やラベル等のアノテーションを、柔軟な構造化形式である XML により記述する。例えば A 地点に設置された t-Room を時刻 t から B 地点の t-Room と接続した場合、その操作内容を簡素化したものを図 3 に、実際の XML を付録に示す。ユーザが t-Room の操作を行った場所と時間が exec 要素の属性として記述され、実行されるオペレーションが operation 要素以降に記述される。operation 要素内では実際の接続要求が time(時間)・place(場所)・presentation(表示)の 3 つの要素により指定されている。このような形式であれば、現在時刻における同期通信と、過去の時空間コンテンツの利用が同一形式で表現可能になる。過去の指定の場合は、operation 内の time="now" が過去の時間指定となる。t-Room は仮想共有平面を重ね合わせることで多地点の遠隔コミュニケーションを実現しているが、表示の属性として重ね合わせるレイヤに ID を持たせ、さらにそのレイヤの透過度を指定することが可能である。

また、解析処理によって得られる情報 (映像中のオブ

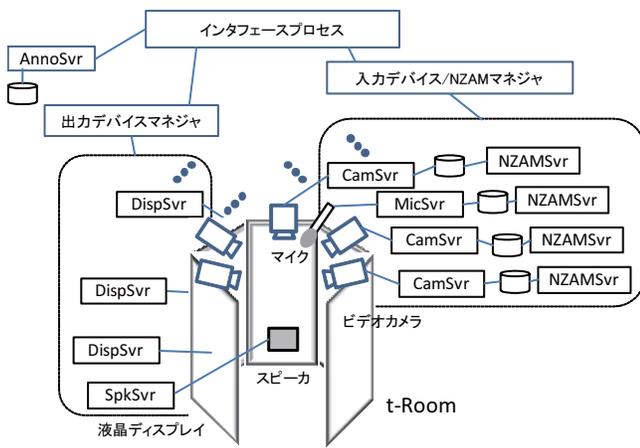


図 4: t-Room のシステム構成

ジェクトの切り出し情報等) や、人手によるインタラクションへのアノテーション (インタラクションへ付与されるラベル等) も、同様に場所・時間・表示情報に分割し XML 形式で記述する。これらの情報は時空間コンテンツの検索・加工に利用する。

3.2 t-Room インタラクションの時空間コンテンツ化

図 4 に t-Room のシステム構成を示す。本システムはユーザがシステム操作を行うインタフェースプロセス、ディスプレイサーバ (DispSvr) やスピーカサーバ (SpkSvr) などの出力系サーバの管理を行う出力デバイスマネージャ、カメラサーバ (CamSvr) やマイクサーバ (MicSvr) などの入力系サーバと NZAM サーバ (NZAMSvr) を管理する入力デバイス/NZAM マネージャ、時空間構造を管理する時空間構造化サーバ (AnnoSvr) から構成される。

t-Room システムが動作している間、随時ビデオや音声などの入力データはカメラサーバ、マイクサーバを通して NZAM へとコンテンツ素材化され、データベースへ蓄積される。

入力デバイスサーバと NZAM サーバは同一マネージャが管理する。通常の同期通信の場合は各入力デバイスサーバに、過去の時空間コンテンツを利用する場合は NZAM サーバに処理を振り分けることで、現在と過去の差を吸収し、同期・非同期の時空間コンテンツ利用を同一アーキテクチャで実現するためである。

ユーザがインタフェースプロセスを通して t-Room 間の接続などの操作を行った結果や、非同期に行われる画像中のオブジェクト認識結果や人手により付与されるラベルなどは 3.1 節で述べた XML 形式として時空間構造化サーバへ保存される。本サーバは自 t-Room や他 t-Room からの検索要求を受け付けるようにするため、t-Room システムの外に配置する。

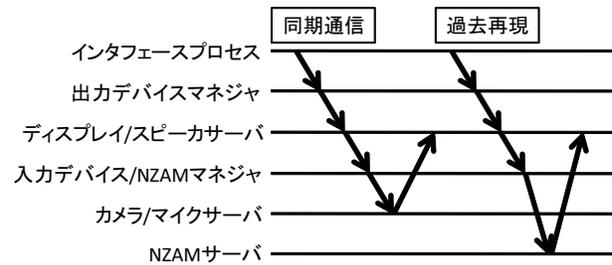


図 5: 時空間コンテンツ要求と送信の流れ

3.3 時空間コンテンツの配信

本アーキテクチャの命令伝達の流れを図 5 に示す。ユーザがインタフェースプロセスの操作により時空間コンテンツの利用要求を行うと、その要求を出力デバイスマネージャへ送信する。出力デバイスマネージャは、利用要求をディスプレイサーバやスピーカサーバなどの各出力デバイスサーバに分配する。次にディスプレイ/スピーカサーバは、利用要求に該当する遠隔の t-Room の入力デバイス/NZAM マネージャにコンテンツ素材を要求する。入力デバイス/NZAM マネージャは、要求されたコンテンツ素材の時間情報が現在の時刻の場合はカメラサーバ、マイクサーバ等の入力デバイスサーバへ配信命令を振り分ける。カメラ/マイクサーバは、対応するディスプレイ/スピーカサーバと接続し、コンテンツ素材を配信する。要求するコンテンツ素材の時間情報が過去の時刻の場合は各 NZAM サーバへ配信命令を振り分け、以下同様に NZAM サーバが各ディスプレイ/スピーカサーバへコンテンツ素材を送信する。

t-Room ではその機構上、カメラ映像に対して映像エコーを抑制する機能を持っている [8]。本機能ではカメラ画像から、ディスプレイ中に存在しているオブジェクトの背景差分を取り、実体のみを切り出しているが、このオブジェクト切り出しの情報は、特定のオブジェクトの表示・非表示を操作するなど、時空間コンテンツの加工にも有効である。そこで t-Room ではビデオカメラの映像に対するオブジェクトの切り出し情報を時空間構造として蓄積する。

時空間コンテンツは、単純に再現する以外に、それを加工してインタラクションに利用することができる。例えばインタフェースプロセスにおいて、時空間コンテンツの特定のオブジェクトを消去するという加工要求があった場合、NZAM サーバがコンテンツ素材を配信する際、該当するオブジェクトの切り出し情報を時空間構造化サーバから引き出し、それを利用してコンテンツ素材から該当オブジェクトを消去し、出力系サーバに配信する。

現在はすべての t-Room の時空間構造を 1 つの時空間構造化サーバが集中的に管理しているが、今後インタラクションの量が増加すると、管理しなければならない時

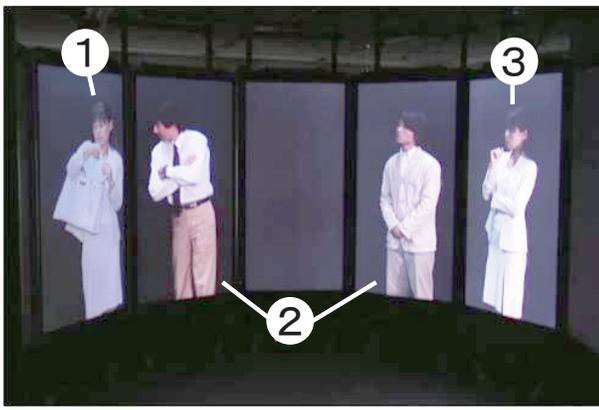


図 6: インタラクションの多段引用

空間構造情報が膨大になってしまう．その場合，時空間構造化サーバを分散化する必要がある．

3.4 t-Room における時空間コンテンツの利用例

最も一般的な時空間コンテンツの利用例として，過去の再現が挙げられる．この場合，t-Room は過去に起きたインタラクションの記録をそのまま再生することで実現される．例えば過去に行った会議の様子を確認しなおしたい時などに有効となる．

また，時空間コンテンツを利用することで，電子メールのように引用を繰り返すことができる．t-Room におけるインタラクションの多段引用の様子を図 6 に示す．この図では 1 の女性が鞆についてコメントしているところを，後に 2 の男性二人が引用しコメントを残しており，3 の女性 (1 と同一人物) がさらにその様子を閲覧している．ジェスチャや体の向き，目線などの社会的インタラクションを踏まえて引用することで，電子メール等と比べて，直接・具体的・身体的状況に関する情報を含め，意図を伝達することが可能になる．

また過去のインタラクションに追加的に参加することで，時空間コンテンツを充実させることができる．その例が図 7 である．まず 1 の男性が体操をし，次に 2 の女性がその時空間コンテンツを再生しつつその体操に参加し，以下 3 (1 と同一人物)，4 (2 と同一人物) の順でインタラクションを重ねることで，過去の人物との組み体操を実現している．このように追加的参加でのインタラクションにより時空間コンテンツを充実させることで，さらに新しいインタラクションを誘発し，時空間コンテンツを充実化させることが可能になると考えられる．

さらに，t-Room での過去のインタラクションに入り込み，追体験を行うことができる．追体験は図 7 の追加的参加と異なり，過去のインタラクションの人物を現在の人物と置き換えることで，置き換えられた人物の体験を感じることである．図 8 は，非同期のインタラクシ



図 7: 過去のインタラクションへの追加的参加



図 8: 同期・非同期の融合したインタラクション

ョンを用いた，遠隔地との同期的なインタラクションの様子を表している．この例では図 7 における 2，3 の人物が消去され，2 の人物の代わりに現在の人物が，3 の代わりに現在の遠隔地の人物が入り込み，過去の人物・遠隔地の人物とのインタラクションを行っている．このように時空間コンテンツを加工し，追体験に適用することで，同期・非同期が融合された新たなインタラクションモードが実現される．これは，2 章で述べた環境に入り込んだ追体験の例である．

3.5 ネットワーク資源・計算機性能の制限への対応

同期的な遠隔コミュニケーションシステムの場合，ネットワークの帯域制限によって遅延が起きるため，高画質のビデオを配信することが困難である．しかし高い臨場感をもった追体験のためには高画質であることが望ましいだろう．そこで t-Room では同期的なコミュニケーションを行う際には標準画質のビデオ (Standard Definition) を配信するが，同時に高画質 (High Definition) での蓄積を行う．次に，一度実現されたインタラクションは，同一の t-Room 上で再現される可能性が高いと想定し，インタラクションを行った t-Room 同士がシステムがネットワークを利用していない時間を使って高画質のビデオを相互配信し，過去の時空間コンテンツの再現等の際に

はその高画質データを用いる。

t-Roomで蓄積した高画質ビデオと、遠隔地のt-Roomから配信した高画質ビデオを合成することで、過去のインタラク션을臨場感をもって再現することが可能になるが、計算機の性能の制限のため、動的に複数の高画質の映像を合成して表示することが困難な場合がある。そこで、システムが利用されていない時間を使って、あらかじめ映像を一枚に合成しておき、それを再生することで、過去のインタラク션을高画質で再現する。

また3.4節で述べたような追体験を行うためには時空間コンテンツを加工して提示しなければならない。たとえばあるオブジェクトを消去したり、特殊なエフェクトをかけたりする必要はあるが、動的にそれらの動画像処理を施して提示することが困難な場合がある。このような場合でも、あらかじめユーザがどのような加工を施すかをシステムに入力しておくことで、複数の動画の合成と同様に、システムのアイドル時間に加工を行う。

このような工夫によって、ネットワーク資源や計算機の性能が制限されている環境でも、同期・非同期を融合させたインタラク션が可能になる。

4 おわりに

本稿では、社会的インタラク션을コンテンツ化することの意義と要件を述べ、それに従ってt-Roomに対し時空間コンテンツ処理のためのアーキテクチャを実装した。また時空間コンテンツを利用することで、過去の再現や追体験が実現できることを述べた。今後の課題として、以下の点を挙げる。

インタラク션の記録装置と再生装置が同一化され、かつ複数の時空間コンテンツを重ね合わせることが可能であるため、表示されるそれぞれのオブジェクトがどの時間のものであるかを判断することが困難である。そのため各オブジェクトの時間軸を認識させるための工夫が必要である。

今回はインタラク션の時間軸移動を過去方向に移動することを可能にしたが、あるメンバーがそろった時に、昔のインタラク션の様子を自動的に再生するなど、時間軸を未来方向にも拡張することが考えられる。

謝辞

日頃より本研究に関して熱く議論してくださったt-Roomプロジェクトの皆様へ深く感謝いたします。

文 献

- [1] D. Tapscott, A. D. Williams. ウィキノミクス-マスコラボレーションによる開発・生産の世紀へ-. 日経 BP 社, 2007.
- [2] J. C. Tang. Findings from observational studies of collaborative work. *International Journal of Man-Machine Studies*, Vol. 34, No. 2, pp. 143-160, 1991.

- [3] H. Ishii, M. Kobayashi, J. Grudin. Integration of interpersonal space and shared workspace; clearboard design and experiments. *ACM Transactions on Information Systems*, Vol. 11, No. 4, pp. 349-375, 1993.
- [4] 山下淳, 葛岡英明, 山崎敬一, 山崎晶子, 加藤浩, 鈴木栄幸, 三樹弘之. コミュニケーションにおけるフィードバックを支援した実画像通信システムの開発. *情報処理学会論文誌*, Vol. 45, No. 1, pp. 300-311, 2004.
- [5] A. J. Sellen. Speech patterns in video-mediated conversations. In *CHI*, pp. 49-59, 1992.
- [6] O. Morikawa, T. Maesako. HyperMirror: Toward Pleasant-to-use Video Mediated Communication System. In *Proceedings of CSCW*, pp. 149-158, 1998.
- [7] H. Ishii, B. Ullmer. Tangible Bits; Towards Seamless Interface between People, Bits and Atoms. In *CHI*, 1997.
- [8] 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. 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. *グループウェアとネットワークサービスワークショップ*, 2007.
- [9] 山本大介, 清水敏之, 大平茂輝, 長尾確. Synvie: ブログの仕組みを利用したマルチメディアコンテンツ配信システム. *情報処理学会第68回全国大会*, 2006.
- [10] niwango. ニコニコ動画 (rc). <http://www.nicovideo.jp/>.
- [11] 角康之, 伊藤禎宣, 松口哲也, シドニーフェルス, 間瀬健二. 協調的なインタラク션の記録と解釈. *情報処理学会論文誌*, Vol. 44, No. 11, pp. 2628-2637, 2003.
- [12] T. Takada, Y. Harada. Citation-Capable Video Messages: Overcoming The Time Differences without Losing Interactivity. In *i3 Annual Conference Proceedings*, pp. 31-38, 2000.
- [13] C. Greenhalgh, J. Purbrick, S. Benford, M. Craven, A. Drozd, I. Taylor. Temporal links: Recording and Replaying Virtual Environments. In *ACM Multimedia*, pp. 67-74, 2000.
- [14] 山下直美, 平田圭二, 高田敏弘, 原田康徳, 白井良成, 青柳滋己. Effects of room-sized sharing on remote collaboration on physical tasks. *情報処理学会論文誌*, 2007 (to appear).

A t-Room 操作のXML形式

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE dispsvr2-place>
<dispsvr2-place version="1.0">
  <layer id="kyoto">
    <object>
      <contents nzam="nzam:~di.kyoto" />
      <attribute transparency="0.8" />
    </object>
  </layer>
</dispsvr2-place>
```