

# 言語・非言語情報の統合による直感的な仮想物体操作が可能な没入型インタラクションシステム

An immersive interaction system that can manipulate the virtual object intuitively by integrating verbal and non-verbal information

秋友 伸也\*<sup>1</sup>      大本 義正\*<sup>1</sup>      西田 豊明\*<sup>1</sup>  
Shinya Akitomo      Yoshimasa Ohmoto      Toyoaki Nishida

\*<sup>1</sup>京都大学 情報学研究科

Kyoto University, Graduate School of Informatics

In collaborative work in virtual space, to handle virtual objects you need to manipulate both the entities and concepts. In order to perform two types of operations at the same time, it is required to use some input device. In this paper, we propose a unified interface to be resolved easily by integrating verbal and non-verbal information. It was confirmed that the operations can be resolved by using the prototype system, to change the position and size of the virtual object.

## 1. はじめに

拡張現実感 (AR) により実空間と仮想空間を融合することが可能となり、博物館でのガイドや部屋の模様替えシミュレーションなど様々な利用されている。AR や VR では物理的な制約を受けない作業が可能であり、例えば複数人で AR による展示物のガイドや機材のマニュアルなどのコンテンツを作成する協調作業を考えると、他の作業者とのインタラクションを通じて展示や配置の試行が容易に行えるなど、利点が多い。本研究では、そのような他者との協調作業を通じたスムーズな AR コンテンツの作成を支援するシステムを目指す。

久木元ら [久木元 07] は仮想空間上での協調作業においてタブレット端末を用いて手書きした注釈を仮想空間に付与するシステムを開発し、注釈を共有することによる協調作業の効率化を示している。ここで付与できる情報は手書きによる注釈のみであるが、より広範な形でユーザの意図を示す情報を仮想空間や AR 空間に付与することでスムーズな協調作業が可能になると考えられる。付与する情報を仮想物体として捉えるとき、その操作のためには他者との協調作業やインタラクションを阻害しないよう、端末などを用いないよりコストの低い操作方法を考える必要がある。

仮想物体には実物体と異なり物理的な制約がないため、その概念と実体を分離して操作することができる。例えば手の届かない位置にある仮想物体も離れた位置から扱うことができるなど、実物体よりも多様な操作が可能である。仮想物体の操作方法として Mine ら [Mine 96] に代表されるものは、手で触れることのできない距離にある仮想物体を如何に手元にあるように扱うかという手法であるが、操作自体は実物体と同様に手で触れることで行うため、実物体に近い制約を課すことになり操作が冗長なものになる。一方で概念的な変数を直接操作するには「+X 軸方向に 3 移動する」のように記述するのが一般的であり、CAD などの 3D ソフトではそれらの変数を個別に指定して仮想物体を操作することになるが、AR 空間内で協調作業中に指定するには煩雑であり、インタラクションを阻害することになる。友添ら [友添 04] はデータグローブを用いたジェスチャの認識によって概念・実体の指定と自身の移動を統一に行う操作インターフェースを開発し、仮想物体を簡易に操作で

きることを示した。本研究では AR 空間での仮想物体の操作に必要な概念・実体の指定を発話とジェスチャにより行うことで、機材を身に着けることなく直感的な操作が可能になると考え、目指すコンテンツ作成支援システムの基盤としてその環境の構築を行う。

発話とジェスチャを用いることができる環境の中で、どのような振る舞いを実際に仮想物体の操作に用いるべきかは自明ではなく、本研究の目的を達成するためには、操作に用いる自然な発話とジェスチャを確認する必要がある。そのため観察実験として、二人のユーザ間でインタラクションを行いながら仮想空間上に家具を配置するタスクを行った。このタスクでは実空間にいるユーザが家具の種類や大きさ、配置に対する指示を出し、その指示を仮想空間の操作者が解釈し仮想空間上に家具オブジェクトを配置していく。次に観察実験を通して観察された指示者の操作指示を分析し、言語・非言語情報の統合指示を解釈して仮想物体を操作するシステムを構築した。最後にシステムにより簡易に仮想物体を操作できることを確認する評価実験を行い、結果を分析した。

## 2. 観察実験

仮想空間に簡易に意図した情報を付与するため、発話とジェスチャにより実空間から仮想物体を操作するインターフェースを目指し、前段階として観察実験を行った。観察実験は仮想物体の配置指示を出す指示者と実際に仮想物体を配置する操作者の二人で行い、主に指示者の指示方法を観察する。指示者の指示方法を認識し仮想物体を操作するシステムを構築することで、目指すインターフェースを構築できると考える。

### 2.1 実験環境

#### 2.1.1 使用機材

指示者に非言語行動を阻害することなく仮想空間内での視界を提供するため、出力装置に Wrap920AR<sup>TM</sup> (Vuzix 社製) を使用した。入力装置は頭部方向センサとして地磁気・ジャイロセンサ、Microsoft Kinect<sup>TM</sup> (Microsoft 社製)、マイクを用いた。Kinect の RGB カメラ、深度画像カメラにより、観測範囲内にいる人間の骨格情報を推定、取得する。これにより指示者の実空間での位置、行動を取得し、頭部方向センサと併せ指示者の視点を決定する。実験環境は実空間側、仮想空間側の二つに分かれる。実空間は縦 3.5 メートル、横 3.5 メー

連絡先: 秋友 伸也, 京都大学 情報学研究科, mail: akitomo@ii.ist.i.kyoto-u.ac.jp

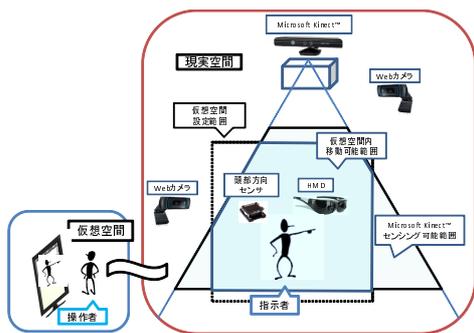


図 1: システム構成図

トル, 高さ 3 メートルの空間として設定され, 図 1 のように Kinect により観測できる範囲を自由に移動することができる. 仮想空間は図 3 のように同様の縮尺で設定された範囲の床面と, その床面のうち二面に L 字型になるように設置された壁面により構成される. これは Kinect の観測範囲を最大限に活用するための設定であり, 主に壁面側に向かって作業を行うものとする. 実空間から得られた空間座標は同一の縮尺で仮想空間に統合され, ヘッドマウントディスプレイを通じて実際に仮想空間内にユーザが存在するような視界を提供する.

## 2.2 実験設定

実験では仮想空間に家具を配置するタスクを行った. 指示者は図 1 の実空間側環境下で行動し, 操作者に対し家具を配置する指示を出してもらった. 操作者は指示者の様子を仮想空間と実空間の映像の両方から観察し, 指示者の言語指示と非言語指示から仮想空間に家具を配置する操作を行う.

観察実験でのシステムの処理の流れを図 2 に示す.

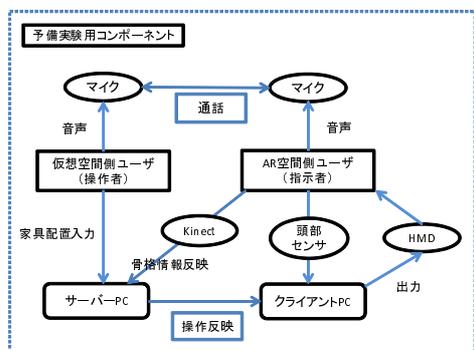


図 2: 観察実験のコンポーネント図

配置する家具の種類と位置は予め決まっており, 図 3 のようになっている. 配置は予め指示者に伝えられ, この通りに家具を配置するよう操作者へ指示を出してもらった.



図 3: 観察実験の流れ

操作者と指示者は, それぞれが持つマイクを入力とし通話を行うことができる. 操作者は指示者の様子をカメラ映像と仮想空間の情報から観察することができるが, 指示者には操作者の様子はわからない.

## 2.3 結果

実験中に観察された家具に対する操作指示は位置指定, 移動, 拡大縮小, 回転, 家具種類の指定の五種類に大別することができた. 全体的に言語のみを用いた指示が共通して多く見られた一方で, 位置・方向指定においては言語のみでは説明し難いものもあり, 手の動きにより向きを指定するなどの非言語指示が見られた. 非言語のみの指示は見られず, 動作の意図を示す言語指示が同時に見られたため, 統一的指示として捉えることができると考えられる.

統一的指示として特に位置指定のための指差し, 家具の移動のための調整ジェスチャが頻繁に見られたため, この結果から指差しと調整ジェスチャに焦点を当て, 言語指示と統合して仮想物体の操作を行うシステムの実装を目指す. ここで調整ジェスチャとは腕を前方に伸ばし, 手を小刻みに往復させる動作を示す「もう少し右側へ」といった方向を示す言語指示と共に用いられる.

## 2.4 考察

実空間にいるユーザが仮想物体の操作指示に対して用いる言語・非言語情報を観察するという目的のため, 言語指示のみでは意図を伝え難くなるよう家具配置を設定した. しかしいずれの参加者についても言語情報による指示が多くを占め, 指差し以外の非言語行動の頻度は参加者により大きく差が見られた.

参加者にはできるだけ参考図通りになるよう家具を配置することを念頭に置いて指示を出してもらったが, 実験後の配置の状態は人により様々であった. 指示の内容が細かくなると, 言語・非言語情報のいずれを用いても意図を伝え難くなり, 操作者としてもその解釈が難しくなるため, ある程度似ていると思われる配置で妥協するという様子がよく見られた. そのため, 微調整のためには指示者, 操作者間の通話を介さず, 指示者が直接家具を操作できるようなシステムが必要と感じられた.

## 3. システム構築

観察実験で得られた指針を元に, 言語情報, 非言語情報から家具の操作を行うシステムを構築した. 節 2.3 に挙げた中で特に数多く見られ, 特徴的であった指差しと調整ジェスチャに焦点を当て, それらを言語, 非言語の情報を用いて認識するシステムを構築した. 観察実験で見られた家具配置指示の多くは, まず指差しにより家具の位置を指定する, とりあえず家具を出すよう指示し, 出てきた家具に対し大きさや向きを指示する, という流れであったため, このシステムでは基本的にその流れに基づいて家具の操作を行う. 即ち, 指差しにより位置を指定して家具を出し, その後調整ジェスチャにより微調整を行うという流れである.

### 3.1 システム概要

このシステムでは, 観察実験で見られた言語情報と, 指差し, 調整ジェスチャを認識し, 家具の操作系列に変換する.

非言語行動に注目すると, 手の動きだけでその意味を一意に定めるのは大変困難である. 例えば調整ジェスチャについて考えると, 手が左右に振られた際に右を示しているのか左を示しているのかの判断は至難である.

また言語情報のみでは空間位置の指定は困難であるし、操作の変化量の微調整は各単語の微妙なニュアンスに頼らざるを得ず、意図した通りの操作を行うことはできない。

そこで言語情報、非言語情報を組み合わせて意味の特定、量の調整を行うことで、それぞれの情報の誤検出を補いつつ意図通りの操作が可能になると考えられる。システムは大きく音声認識部、行動認識部、統合部の三つからなる。図4に大まかな処理の流れを示し、以下で各要素について説明を加える。

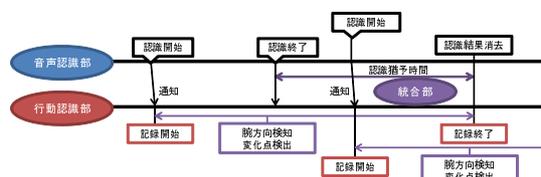


図4: 処理のタイミングチャート

### 3.2 音声認識部・言語情報の認識

言語情報の認識には入力装置として無線マイクを、音声認識エンジンとして Julius[河原 05] を用いた。Julius からは音声の入力状態、認識結果などを取得できる。先の実験で得られた、位置、サイズ、回転の指示に用いられた言語情報を軸に、自動化のためその他に必要な操作を示す言葉を追加し、それらを認識する単語の辞書ファイルを作成した。

### 3.3 行動認識部・非言語情報の認識

非言語情報の認識には Kinect により得られる骨格情報を用いた。操作は全て右手で行われるものとし、Julius に音声入力が行われた時点から、音声入力データに対応する右手位置データとして記録する。記録データに対し SWAB[?] を簡略化した処理を用いて近似し、右手動作の変化点の検出を行う。これにより、調整ジェスチャが行われた方向とその回数を得ることができる。

### 3.4 統合部

#### 3.4.1 仮想物体の操作系列

システムで認識する操作は節 2.3 に示した五種類である。

- 位置指定: 「ここに」「あのあたりに」などの指示語と指差しを認識し、指差し先を仮想空間にマークする。
- 家具指定: 実験に用いる家具の種類と自動化で用いるその他の操作(設置, 消去, 選択等)を認識する
- 移動・拡大縮小・回転: 配置した家具の微調整を行う

移動・拡大縮小・回転の三種類の操作では共通した認識処理を行う。操作種別と大まかな変化量の指定のために音声認識結果を、実際の操作と微調整のために調整ジェスチャを用いる。調整ジェスチャは元々移動操作に特化したジェスチャであるが、ここでは簡易な自動化のために回転・拡大縮小操作についても用いた。言語情報では移動を示す「右へ」拡大縮小を示す「大きく」回転を示す「回して」などの言葉と「もっと」「もう少し」といった変化量の多寡を表す言葉を組み合わせて認識する。

#### 3.4.2 言語・非言語情報の統合

各操作では認識する非言語行動の種類が決まっている。位置指定では右手が静止状態である場合のみ認識する。左右への移動操作、回転操作は左右方向の調整ジェスチャのみ、上下、前後への移動操作は上下方向の調整ジェスチャのみを操作に反映

する。拡大縮小操作ではどちらの方向の調整ジェスチャも認識することとした。特に方向を指定する操作において、言語・非言語情報間に齟齬が発生する場合は誤認識として棄却することで誤操作を防ぐことができると考える。一度の操作では、「基本量 \* 変化量 \* 調整ジェスチャの検出回数」だけ選択された仮想物体にその操作を反映する。例えば「もっと右へ」という発話とともに左右方向の調整ジェスチャが二回見られた場合、「基本量 \* 1.5 (もっと) \* 2 (ジェスチャ回数)」となり、大きく移動することになる。

## 4. 評価実験

構築したシステムを用いて、言語・非言語の統合により簡易な仮想物体の操作ができることを確認するため、評価実験を行った。意図通り操作が行えていること、操作の誤認識が起きていないことを確認する。

### 4.1 実験設定

評価実験では図1の実空間側の環境のみを利用して実験を行った。観察実験との違いはシステムが全て自動化されていることであり、操作者は存在しない。操作者が担当していた作業を、各認識部を通じ実空間側から直接行う。評価実験でのシステムの処理の流れを図5に示す。実空間、仮想空間の設定は観

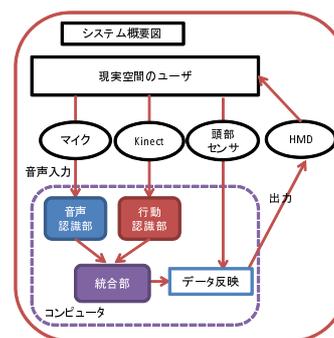


図5: 評価実験のコンポーネント図

察実験のものと同じだが、配置する家具を椅子、机、ベッド、窓の四つに減らし、タスクを単純化した。いずれの家具も拡大縮小、回転、移動の操作が必要となるよう設定しており、システムの動作を見るのに適した設定であると考えられる。



図6: 左:実空間の様子 右:参加者の視界

### 4.2 結果

実験を通して行われた操作の検出結果を図7に示す。言語により指定した通りの操作がジェスチャの検出により実行されたものを認識成功とした。連続して操作が行われる場合がほとんどでありジェスチャの認識区間が重複するため、細かな調整ジェスチャの認識回数については不問とした。言語指示があり、ジェスチャが行われたにもかかわらず、何の操作にも反映

されなかったものが認識不成功、言語を認識し、ジェスチャを検出した結果、言語による指定と実際に反映された操作が異なったものが誤認識である。

操作対象	認識成功	認識不成功	誤検出
指差し	21/24 (87.5%)	3/24 (12.5%)	0/24 (0%)
窓	38/44 (86.4%)	6/44 (13.6%)	0/44 (0%)
ベッド	117/182 (64.3%)	62/182 (34.1%)	3/182 (1.6%)
机	57/82 (68.7%)	25/83 (30.1%)	0/82 (0%)
イス	44/57 (77.2%)	13/57 (22.8%)	0/57 (0%)

図 7: 実験結果

表からもわかるとおり、誤認識がほぼ 0 である一方、認識不成功の率は高くなっている。言語情報、非言語情報の認識それ単体ではある程度の誤認識が見られ、その結果操作を行ったはずであるのにそれが反映されない、という場面が多く見られたためである。言語・非言語情報の統合による誤操作の抑制は大きく働いていることがわかり、認識不成功の場合でも一度の操作にかかる負担は極めて小さく、操作を繰り返すことですぐ正常に認識される状態に至った。誤認識の 3 例はいずれも同一参加者のもので、特に音声認識部の認識結果が大きく個人に依存し、誤操作の元となることがわかった。

実験全体で操作が困難だったものは机とベッドであった。これはジェスチャを認識する区間を右手が体から離れている間としているため、近場で作業を行うことになるものに対し、手が離れていないものと認識されてしまったためと考えられる。基本となったのは「少し」「もう少し」といった少量の指定とそれに伴う微調整の操作であり、慎重に家具を操作する様子が多くの場合で見取れた。左右・前後方向への移動は操作者の位置、向きに依存するため、操作者自身が家具を操作をしやすい位置に移動し、視点を変えながら微調整する場面も見られた。一方で「もっと、もっと、もっと」と発話を連続で行い、その間に手を振り続けることで連続して大きく家具を移動、回転させる場面も確認でき、システムが細かな操作と大まかな操作のそれぞれに対応できていることも確認できた。

#### 4.3 考察と議論

##### 4.3.1 言語情報について

観察実験において「部屋の角」「ソファの隣」といった空間内の他の要素の情報と比較する指示が多く見られており、そのような周囲環境の情報を認識して利用することや、大きく操作を行いたい場合には発話が早くなり逆に小さく操作したい場合には遅くなるといった声の抑揚を取り入れることで、より直感的に操作できると考えられる。

##### 4.3.2 非言語情報について

このシステムでは指差しと上下左右の調整ジェスチャのみを操作に反映させている。先にも述べたように仮想物体の移動操作に対しては調整ジェスチャは直感的な操作といえるが、上下・左右方向のみ検出しており前後移動操作に対しては直感的とはいえず、実験中では前後移動操作の表現として手の押し出し、手招きといった動作が見られた。同様に回転の操作に対して、手を回す、手で円を描くといった動作も見られ、検出できるジェスチャの種類を増やし、仮想物体の操作種類に合った直感的なジェスチャを検出する必要性が感じられた。また今回の実験ではユーザの行動の取得に用いた Kinect は一台のみであり、正面方向を向いたユーザの情報は精度よく取得できるが、ユーザがそれ以外の方向を向いている場合に大きく外れたデータを取得することがある。複数台の Kinect を設置

し、ユーザの体方向から信頼できる Kinect を判別するなどしてデータを統合することで、方向に依存することなくデータを取得できるようになると考えられる。

##### 4.3.3 実空間との位置合わせ

今回の実験は全て仮想空間上での仮想物体の作業であったが、本研究は AR 空間での協調作業の支援を目標としており、仮想空間の情報を実空間にマッピングする必要がある。そのためには実時間・高精度での実空間との位置合わせの機能は不可欠であり、画像特徴量を利用してマーカレスで三次元位置を検出する手法である PTAM[Klein 07] などを利用して実装を進める予定である。

## 5. おわりに

本研究では協調作業によるコンテンツ作成を支援するシステムの構築を目指し、その一部として言語・非言語情報を統合して仮想物体の操作を行うシステムを構築した。まず仮想空間上での家具配置という作業を通じてユーザの言語・非言語指示を観察する実験を行い、その分析を行った。次に観察実験の結果からシステムに用いる言語・非言語情報を決定し、それにより仮想物体を操作するシステムを実装、評価実験を行い、その有効性を確認した。

今後の課題として、コンテンツ作成支援システムのために必要な他の要素を示す。本研究の実験では事前に定めた家具を配置する仮想物体として用いたが、支援システムとして特定の家具に対応するのではなく、言語情報によるタグ付けなど、ユーザの意図を表す情報を簡易に記録し、蓄積するというシステムの構築を目指している。そのためにユーザのジェスチャから適切に仮想物体を提示する機能やタグ付けのためにより柔軟に言語を認識する機能が必要であり、ジェスチャ認識、音声認識において課題が存在する。

## 参考文献

- [久木元 07] 久木元伸如, 江原康生, 古川雅人, 小山田耕二: 没入型共有 VR 空間での遠隔協調作業における手書き注釈付与を用いた思考支援に関する実験的検証, 情報処理学会論文誌, Vol. 48, No. 6, pp. 2153.2163 (2007).
- [Mine 96] Mine, M.: Working in a virtual world: Interaction techniques used in the chapel hill immersive modeling program, University of North Carolina(1996)
- [友近 04] 友添雄亮, 町田貴史, 清川清, 竹村治雄: 遠隔仮想物体操作と仮想空間ナビゲーションのためのジェスチャを用いた統一的操作手法, 電子情報通信学会技術研究報告, ITS2003-56 (2004).
- [河原 05] 河原達也, 李晃伸: 連続音声認識ソフトウェア Julius, 人工知能学会誌, Vol. 20, No. 1, pp. 41.49 (2005).
- [Keogh 01] Keogh, E. J., Chu, S., Hart, D. and Pazzani, M. J.: An Online Algorithm for Segmenting Time Series, Proceedings of the 2001 IEEE International Conference on Data Mining, ICDM '01, Washington, DC, USA, IEEE Computer Society, pp. 289.296 (2001).
- [Klein 07] Klein, G. and Murray, D.: Parallel tracking and mapping for small AR workspaces, Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on pp. 225.234 (2007)