

# CAVEにおける音声を用いた 複合的インターフェースの検討

藤城 幹山 高橋 裕樹 中嶋 正之

東京工業大学 情報理工学研究科

近年のコンピュータ技術の飛躍的な向上によって物理的には存在しない3次元世界、仮想世界をコンピュータを用いて構成し、表示することが可能になってきている。またその仮想世界に対して実在する世界のようにインタラクションを行うVR (Virtual Reality) は医療、通信、娯楽などの様々な分野で利用されている。VRはユーザがアプリケーションを扱ったり、他のある人とコミュニケーションをとったり、共同作業などを行なうときに有効な手段であると考えられる。本稿ではVRシステムにおいてより使いやすいインターフェースを目指し、音声入力と仮想の棒による指示動作を利用したインターフェースの検討を行なった。

## Investigation of Multiple Interface Using Speech Recognition in CAVE

Kanzan FUJISHIRO Hiroki TAKAHASHI Masayuki NAKAJIMA

Graduate School of Information Science & Engineering,  
Tokyo Institute of Technology

With the recent advances in computer technology, we are able to create 3D world that does not exist and virtual world in the computer and to display them accordingly. The ability to interact with virtual world ( virtual reality ) implies that it can find application in many fields such as medical, communication and entertainment etc. We believe that VR can be a useful mean when a user has to deal with an application or when he needs to communicate and work together with other users. This paper aims to provide a user friendly interface for the VR system through the use of speech and the movement of a virtual stick.

## 1 はじめに

近年のコンピュータ技術の飛躍的な向上によって物理的には存在しない3次元世界、仮想世界をコンピュータを用いて構成し、表示することが可能になってきている。またその仮想世界に対して実在する世界のようにインタラクションを行うVR(Virtual Reality)は医療、通信、娯楽などの様々な分野で利用されている。

VRシステムの1つであるCAVE(CAVE Automatic Virtual Environment)[2]では体験者を複数の大型スクリーンによって囲むことで視野を覆い、時分割方式による立体視を用いることによって、より没入感の高い仮想空間の表示が可能である。このCAVEシステムに限らずシステムを有効に利用するためにもインタフェースは重要である。また今後コンピューター、CG(Computer Graphics)、VRなどの専門的知識の少ない人や慣れていない人にもこのような仮想空間を利用する必要性が高まることが考えられるので、より使いやすい、簡単なインタフェースが望まれている。

本稿ではこのような観点からCAVEという没入感の高いVRのシステムで音声を用いたインタフェースに関して検討を行なった。

まず第2節で背景について述べる。第3節で今回使用したハードウェア構成を述べる。第4節で今回実装したシステムの概要を述べる。第5節でまとめおよび今後の課題について述べる。

## 2 背景

VRはユーザがアプリケーションを扱ったり、他のある人とコミュニケーションをとったり、共同作業などを行なうときに有効な手段であると考えられる。このVRシステムの一つである高精細立体映像表示装置CAVEは、直方体の3側面をスクリーンで覆ったものと、床面をスクリーンとする箱型のVR体験装置である。このような大型投影装置を用いることで、体験者の視野を完全に覆うこと、また、体験者と等身大の映像を表示することが可能になるため、現実感の高い仮想空間を体験することが可能になる。CAVEには、次のような特徴がある。

- (1) 多人数参加型のVR空間
- (2) 多面高精細大型映像システムの利用
- (3) 実時間3D立体空間の創成
- (4) 6軸位置入力システムによる高性能対話性
- (5) 3次元オーディオ環境

また音声認識技術の向上によって実時間で音声認識可能で十分に精度の高い音声認識のソフトも手軽に入手できるようになってきている。音声入力の利点としては次のようなことがあげられる。

- (a) 人間にとって自然で簡単な入力方法である
- (b) キーボードなどのような他の入力装置と比較して熟練を必要としない
- (c) 手や足などで別の作業をしながらでも並列的、相補的に操作することが可能である

ジェスチャやマウスなどの他の様々な入力方法と音声入力をあわせたマルチモーダルインタフェース[3][5]においてこの音声認識技術は大変重要な位置を占めており、これらを組み合わせることによって人間にとってより自然なコンピュータの操作が実現できる。

本稿は以上のような観点から音声と仮想空間中の物体の指示動作に使用する仮想の棒を用いたインタフェースを検討した。

## 3 ハードウェア構成

ハードウェア構成を図1に示す。SGI ONYXが仮想環境全体の管理を行なっている。ヘッドセットマイクとジョイスティックが繋がっているPCとONYXとはTCP/IPイーサネットで通信を行なっている。赤外線エミッタ、コントロールユニットはONYXのシリアルポートにつながっている。

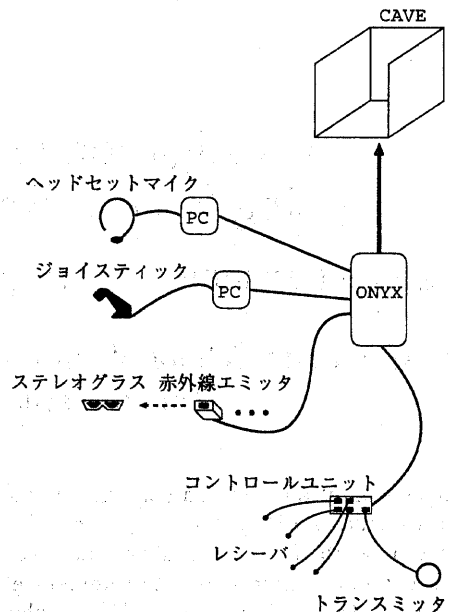


図1: システムのハードウェア構成

### (1) SGI ONYX/Reality Engine2

描画処理を行なう。その他にも仮想環境全体の管理を行なっている。開発環境はC/C++でグ

ラフィックスライブラリには IRIS Performer を用いた。

#### (2) 音声認識用 PC

音声認識は計算量等コンピュータに負担がかかるため別のコンピュータで音声認識を行ない、その結果を仮想環境の管理をしている ONYX へ通信する。音声認識時間や通信による時間の遅れは現在 1~2 秒程度である。CPU は Intel Pentium Processor 133MHz。メモリーは 48MB。音声認識ソフトは IBM Voice Type Simply Speaking Gold で Windows95 上で稼働している。

#### (3) Polhemus 社 3SPACE FASTRAK

3次元位置情報を検出するために使用した位置センサは Polhemus 社の磁気センサ 3SPACE FASTRAK を使用した。トランスミッタで磁界を発生し、レシーバでその磁界を感知して位置を計測する。3次元位置 (X, Y, Z) 座標と回転 (Pitch, Yaw, Roll) の 6 自由度の情報が入りできる。最大 4 つのレシーバを同時使用可能であり、今回は 3 つを使用した。レシーバを 3 つ使用したときのデータレートは各レシーバにつき 40 ポイント/秒である。

#### (4) Stereo Graphics 社 CrystalEYE2 赤外線エミッタ

CrystalEYES2 は液晶シャッターグラスでありこれによって時分割方式の立体視を行なう。赤外線エミッタが同期信号を発生し CrystalEYES2 がこれを受け取る。

## 4 システム概要

今回実装したものは仮想空間に表示したメニューを選択する方式と音声によるコマンド入力方式によって仮想空間での操作を決定する。また仮想空間中の対象の指示動作は仮想空間中に表示されている仮想の棒によって選択を行なう。

メニュー選択方式ははじめて使うような人にはわかりやすい。しかし操作が多いような場合には一度に表示するメニューの数が多く階層が深いため選択するのに時間がかかってしまう。そのためコマンド入力方式のインタフェースも必要である。コマンド入力方式のインタフェースとして代表的なものはキーボードによるコマンド入力であるが CAVE のような体験者が自由に仮想空間中を歩けるようなシステムではキーボードを使用することは体験者を拘束してしまい没入感を損なう恐れがある。音声による操作では装着物がヘッドセットマイクのみですむので拘束感も少ない。そこで音声による操作を採用した。

## 4.1 仮想空間における装着物

仮想空間中では図 2 に示す装着物を図 3 のような状態に装着してインタラククションを行なう。

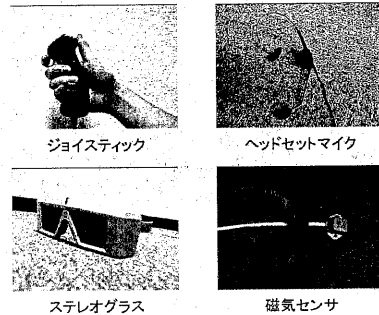


図 2: 装着物

以下に図 3 での装着している状態の説明をする。

#### (1) ステレオグラスと磁気センサ

立体視を行なうために利用する。ステレオグラスに磁気センサを付ける理由は立体視を行なうときに視点の位置が必要であるからであり、この磁気センサによって位置情報を取得する。

#### (2) ヘッドセットマイク

音声認識用に装着しているもので口元にマイクを設置することによって雑音の影響を減らし、認識率の低下を防ぐことができる。

#### (3) ジョイスティックと磁気センサ

仮想空間中に表示されているメニューの選択やオブジェクトの選択は仮想空間中に表示する仮想の棒によって選択を行なう。この仮想の棒は図 3 右手のジョイスティックに位置が対応していて必要になる位置情報はこのジョイスティックにつけた磁気センサの値を用いている。十字キー 1 つとボタンが 4 つありこの十字キーやボタンで操作を行なう。

#### (4) 磁気センサ

図 3 左手の磁気センサは仮想空間中に表示するメニューの位置にあわせてある。前述 (3) のジョイスティックとあわせてメニュー選択を行なう。図 4 はメニュー選択を行なっている様子を示している。

## 4.2 操作の流れ

この仮想環境での操作の流れは以下の通りである。

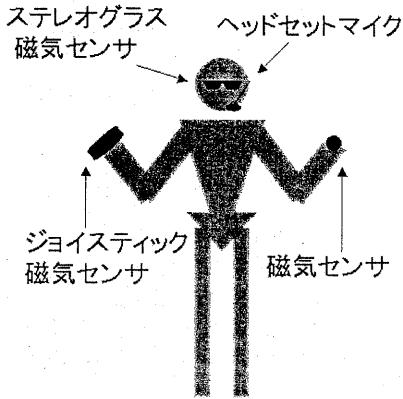
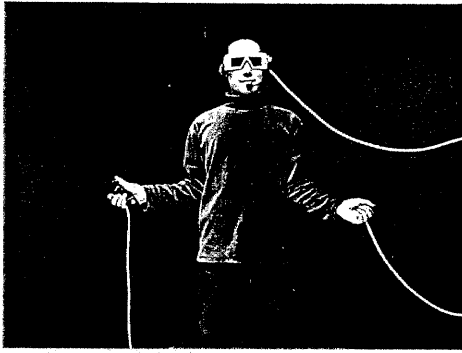


図 3: 装着した状態

まずコマンドを選ぶ。コマンドはメニュー選択または音声入力によって決定する。そしてそのコマンドが対象を必要とするコマンドならばその対象を仮想空間中から仮想の棒で指して選択する、また次に必要に応じて出てきたダイアログを操作する。

コマンドを選択  
(メニュー選択か音声入力)

↓

選択した操作に仮想空間中の対象  
が必要なら仮想の棒によって選択する

↓

必要に応じて出てきたダイアログで操作

(1) メニュー選択

図 4 にメニュー選択を行なっているところを示

す。仮想空間中のメニューは図 3 で左手に持っている磁気センサの位置と対応して左手の位置に表示される。それを仮想の棒で指し、ジョイスティックのボタンを押していくことでメニューを選択していく。

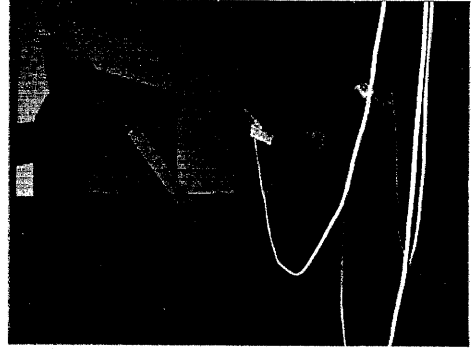


図 4: メニュー選択の例

(2) 音声入力

ある音声に対してはひとつのコマンドが割り当てられているのでそのコマンドに対応した音声を発声すればあるコマンドが選ばれる。そのためメニュー選択よりも高速にインタラクションを行なうことが可能である。

図 5 に音声処理の流れを示す。体験者の発した音声は図 3 にあるようにヘッドセットマイクで拾われ音声認識用のコンピュータに送られる。そこで音声認識を行ない、その結果が仮想環境の管理を行なっている ONYX に送られる。

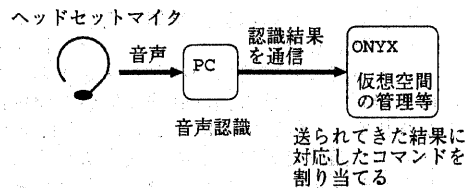


図 5: 音声部分の処理

(3) 仮想の棒による対象の選択

選択したコマンドに対象が必要な場合には仮想の棒を仮想環境中の対象にあわせジョイスティックのボタンでコマンドの対象の選択を行なう。仮想空間中に仮想の棒を直接表示してその仮想の棒で対象を指し示すことで指示動作の曖昧さが減る。またこの仮想の棒は伸縮させることができる。

#### (4) 仮想空間中の移動

仮想空間の移動方法は以下の3通りである。

一つめはある平面上での前進、後退、右移動、左移動、右回転、左回転とその平面に垂直に上、下へ移動、見上げるように動く後向き回転と見下ろすように動く前向き回転をする方法がある。この平面ははじめは水平面で、後向き回転と前向き回転をしたときに傾く。壁や床の衝突判定とあわせて壁や床に沿って移動する方法とあわせると建物などのウォークスルーに向いている。

二つめは仮想の棒の向いている方向に前進、後退と仮想の棒に関して右まわりに回転、左まわりに回転、CAVEの正面スクリーンに仮想の棒の向いている方向を回転させる移動方法がある。一つめの移動方法より自由度の高い3次元の移動が行なえる。

三つめは仮想空間のマップを表示してそのマップ上を仮想の棒で指すことによってそこへ瞬間的に移動する方法である。行きたい場所にすぐに移動することができる。

この三つの移動方法をジョイスティックの十字キーとボタン、または音声とメニューによる指示で操作をする。

### 4.3 アプリケーション

今回次に示したアプリケーションを作成した。

#### (1) 3次元仮想世界のウォークスルー

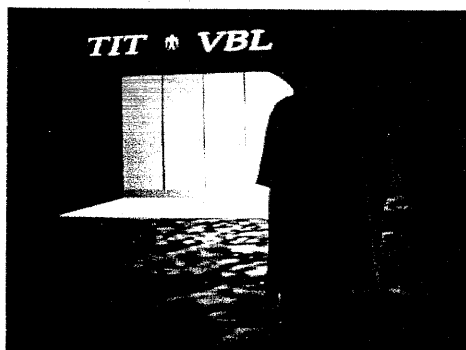
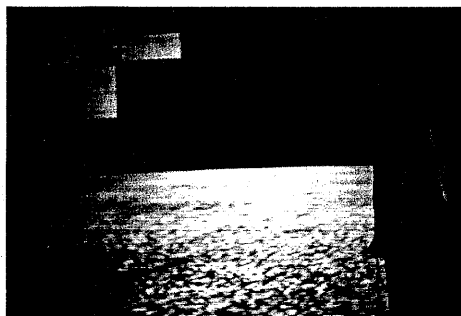


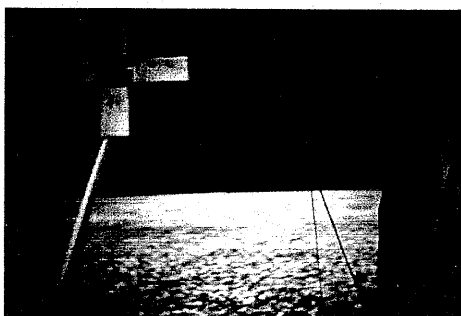
図 6: ウォークスルー

作成された仮想世界を歩いて体験することができる。図6にウォークスルーを行なっている様子を示す。

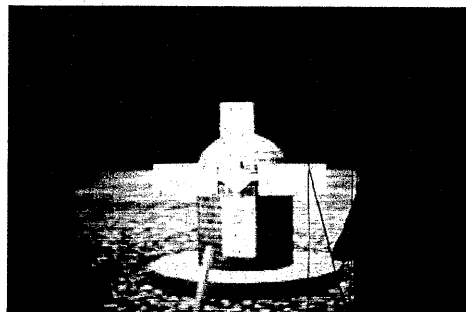
移動方法は4.2(4)に書いた3通りの移動方法をジョイスティックの十字キーとボタンで操作するか音声による指示で操作を行なう。しかし音声で実際の進む方向指定しながら移動するような、例えば「前進する」「右に回転する」などの場合には音声認識の時間と通信の遅延や認識誤



音声入力かメニュー選択によって操作を選ぶ  
(ここでは音声を発している)



ジョイスティックの位置と連動した仮想の棒で  
3D オブジェクトを選択



ジョイスティックのボタンを押しながら3D オブ  
ジェクトを移動させる

図 7: 3D object の移動の例

りなどで望んだ方向に進めない場合もありイライラを感じることもある。

### (2) 3次元仮想世界の作成, 編集

非常に高い没入感があり, 等身大のモデルを表示できる CAVE での仮想世界の作成, 編集は非常に有用であり, バーチャルスタジオ等の作成にも有効であると考えられる。図7に3Dオブジェクトの移動を行なっている様子を示す。

### (3) 簡単な物理シミュレーション

図8で示しているのは電界のシミュレーションである。球が電荷であり矢印がその点での電界の向きを表している。電荷に適切な質量を与えてその電界での動きのシミュレーションを行なえる。



図8: 電界のシミュレーション

## 5 まとめ

本稿では CAVE という VR システムにおいての音声を用いたインタフェースについての検討を行なった。CAVE では体験者が自由に歩きまわれるので拘束感のあるような入力デバイスは不向きである。したがってキーボードなどよりも音声入力による操作の方が有効であると考えられる。また仮想空間中の物体の指示動作には仮想世界中に表示した仮想の棒を用いたがこれによって指示物体の曖昧性も減った。

しかし音声入力による操作では音声認識を行ない, そのあと認識結果を通信するので発声から実行までの間に若干の遅れが生じる。そのため即時性を要求するような操作には問題が残った。

今後は音声入力部分を自然言語へ拡張することや新しい入力デバイスの開発, また人間への出力を画像のみでなく音声を加えていく予定である。

## 参考文献

- [1] Stuart K. Card, Thomas P. Moran and Allen Newell, "The Psychology of Human-Computer Interaction", Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1983.
- [2] Masayuki NAKAJIMA, Hiroki TAKAHASHI, "Multi-Screen Virtual Reality System: VROOM - Hi-Resolution and four-screen Stereo Image Projection System -", Proc. IWNVMT, 16-17 January, pp.95-100, 1997.
- [3] "Speech-based Multimodal Interaction in Virtual Environments: Research at the Thomson-CSF Corporate Research Laboratories", PRESENCE, TELEOPERATORS AND VIRTUAL ENVIRONMENTS, Volume 6, Number 6, pp.687-699, December 1997.
- [4] 藤城 幹山, 高橋 裕樹, 中嶋 正之, "CAVE における音声を用いた複合的インタフェース", 電子情報通信学会 98 年 春季大会 A-16-17, 1998.
- [5] 中嶋 秀治, 加藤 恒昭, "WWW ブラウザとのマルチモーダルインタラクションをクリックを使わないマウスの動きと音声を入力とするインタフェース", 情報処理学会論文誌, Vol.39 No.4, Apr 1998.