

高精細型インタースペースによる会話環境の検討

菅原昌平、松浦宣彦、箕浦大祐、松本敏宏、正木茂樹
NTT ヒューマンインタフェース研究所

あらまし：

3次元共有仮想空間を用いたコミュニケーション環境「インタースペース」の研究を行っている。従来のインタースペースは、通常のPC、インターネットアクセス環境で使えるように、仮想環境の表示精度、仮想空間での利用者の分身（アバタ）の動作などを最適化していた。本稿では、PCおよびネットワーク環境の制約を取り除いた形で、より親しみやすいコミュニケーション環境を実現し、あらたな仮想空間サービスの検討を行っている。

A Study of Communication Environment based on "Hi-Quality InterSpace"

Shohei SUGAWARA, Norihiko MATSUURA, Daisuke MINOURA, Toshihiro MATSUMOTO and
Shigeki MASAKI
NTT Human Interface Laboratories

Abstract:

This paper describes the extension of "InterSpace" in order to realize more realistic space for natural way of communication among users. InterSpace is a platform architecture for constructing networked virtual environments using 3D computer graphics. One of the most important features of InterSpace is that InterSpace integrates the real-time facial video and voice communication capabilities. However, the current InterSpace exhibits shortcomings that limit the potential in computer-mediated communication environment. Especially, we identify the following two areas for necessary improvements:

1. Avatar Appearances

This involves rich expressions of avatars including human-like avatar animations and so on.

2. Rich Visual and Audio Communication

This involves high quality real-time facial video and audio transmission, more effective facial video mapping on avatars, rich voice communication, speaker detection method and so on.

1. はじめに

パソコンの処理能力、特に3次元CGの表示能力は、CPUおよびグラフィック・チップの性能向上により、飛躍的に進歩している。一方、インターネットの利用の普及に伴い、多数のユーザがネットワーク上の仮想社会の住人となりつつある。このような状況の中で、従来からパソコン通信などで提供されている文字ベースのオンラインチャットに代わって、WEB上でのコミュニティサービス（個人ホームページのリンク集とフォーラム、チャットを組み合わせたようなサービス）、さらには3次元仮想空間に多数のユーザが参加してコミュニティを形成しチャットや情報交換を楽しむサービスが出現してきた。3次元仮想空間を用いたコミュニティサービスでは、利用者のアバタ

(分身)を仮想空間内に表示し、各利用者は空間内をアバタの姿で移動し他の利用者のアバタと出会い、チャットや情報交換を行っている。現在提供されているサービスは主にエンタテインメント指向のものが多く、アバタもアニメ風のキャラクタを用い、利用者は好きなキャラクタに変身して他の参加者とのチャットを楽しんでいる。

2. インタースペースの概要と問題点

2. 1. インタースペースのねらい

インタースペースは仮想空間を媒介として、実世界のユーザ同士のコミュニケーションを活性化することを狙ったシステムである。インタースペースの世界では基本的にユーザ同士が相手の顔を見、声を聞くことで実世界の相手を意識したコミュニケーションを行う。インタースペースの世界で提供される種々のサービスの中では、ゲームなどの特定のサービスを除けば、実際の人間とのコミュニケーションをつねに前面に押し出しているところが、通常の3次元チャットシステムとの大きな違いである。アバタはその顔の部分にビデオ映像を表示する機能を持ち、基本的に利用者は自分の実際の顔を他人に見せ、音声でチャットを行う(図1)。

2. 2. インタースペースのアーキテクチャ

インタースペースシステムは、クライアント-サーバ構成をとり、各端末上のインタースペースブラウザ(クライアント)からサーバに対しそれぞれのアバタの現在位置・方向、音声データ、顔映像を送信し、サーバがそれらのデータを他のクライアントに対して配信する。通常のインターネット接続では28.8Kbpsから36.6Kbpsのモデムが良く使われることから、クライアント・サーバ間の伝送速度28.8Kbpsで使用に耐えるようシステムを構成した。すなわち、音声データに関しては、多人数のグループ会話を実現するためミキシング処理が必要となるが、サーバでミキシングを行うことで双方向1本の音声チャネルによるグループ会話を提供している。また、仮想世界の各コンテンツについて、そのモデルデータ、イメージデータ、サービス定義ファイル等を管理し、クライアントの要求に応じダウンロードする。

2. 3. インタースペースの問題点

先に述べたように、インタースペースの中ではアバタを介したコミュニケーションであっても、実際の相手を意識させることを重視している。このため、実際の人間と仮想世界でのアバタの姿とのギャップが目立ってしまうことにもつながる。現在の一般的なPCの処理能力、ネットワークの伝送能力でのサービスでは、現状のアバタや世界モデル、映像・音声品質が限界であるが、インタースペースの利用者からの意見としても、アバタが人間らしくないという指摘がしばしばあった。また、

- 1) アバタの外観が似通っているため、離れたところから見ると誰だかわからない
 - 2) 音声の方向感・距離感が掴めない為、どこから話しかけられているのかわからない、突然声が聞こえて驚く
 - 3) 並んで会話しているとき誰が話しているのかわかりにくい
 - 4) アバタの動作がぎこちない
- などの問題点もある。

3. 高精細型インタースペース

3. 1. 高精細型インタースペースのねらい

高精細型インタースペースでは、ハイエンドの PC と広帯域のネットワークを前提として、より人間らしいアバタを実現して、現実世界のコミュニケーション環境に近づけることをねらいとした。アバタが人間らしく見えるために、

- 1) モデルが本人に似ている
- 2) モデルが人間のような動作をする

の 2 点が必須であろうと想定し、これらを実現するとともに、音声・顔画像の品質を高めた。

3. 2. 高精細型インタースペースのアーキテクチャ

高精細型インタースペースのプロトタイプ開発にあたっては、現在のインタースペースのサーバおよびクライアントモジュールを可能な限り流用し、高精細化に伴って変更を行う部分のみ新たに開発することで、効率化を図った。

1) アバタを人間らしくするための機能

以下の機能を、主にインタースペースクライアントの基本部に対する機能追加で実現した。

アバタのモデル高精細化（ポリゴン数で従来の 40 倍程度）

アバタの動作モデルのポーズ数増加（2 倍程度）

首のみを動かす動作の提供

2) 顔画像の高精細化

以下の機能を、新たな顔画像処理部で実現した。

顔画像フレームレートの向上（従来の 3 倍程度）

顔画像解像度の向上（従来の 4 倍）

横顔・正面顔の画像の切り替え

クロマキー合成による実写顔映像のアバタへのシームレスな統合

3) 音声の高品質化

以下の機能を、新たな音声処理部で実現した。

簡易 3 次元音声

圧縮方式、サンプリングレートの変更による音声品質向上

表 1. で網掛けをした部分を追加している。

基本部 (ISC)	GUI の提供 アバタアニメーションの制御
通信モジュール (UCMC)	サーバとの通信
顔画像処理部 (HQVT)	顔映像の送受信
音声処理部 (HQAT)	音声の送受信

表 1. 高精細型インタースペース・クライアント構成

高精細インタースペース・サーバ構成

サーバに関しては従来のものをそのまま利用している。ただし従来のサーバは顔映像および音声の多重化処理を行っていたが、今回はこれらの機能は使っていない。これは、100Mbps 程度のネットワークを前提に、マルチキャスト通信により端末で多重化を行うことで、サーバの改造を避けシステム開発を効率化したためである。

3. 3. プロトタイプシステムの概要

図2. に今回実現した実写を元にしたアバタの顔を示す。ア)は通常の顔、イ)は口を開いている顔である。他利用者からの音声データを受信した場合、アバタの顔モデルを取り替えることで口の開閉を表現している。図3. にアバタのポーズ例を示す。他利用者の位置情報の変化から移動を検出すると該当利用者のアバタに対し歩行ポーズを表示する。離着席、お辞儀、ドア開閉などは利用者の動作指示を受信した場合に、該当利用者のアバタに指定ポーズを取らせている。

図4. に端末使用中のユーザの様子を示す。利用者を3方向から撮影し、仮想空間内での利用者間の位置・方向の関係により映像を切りかえることで、仮想空間内で横を向いている利用者の顔映像を表示する際には横顔映像を用いる。横顔映像表示中の画面を図5. に示す。

4. 会話環境に与える影響について

4. 1. 高精細型インタースペースの利用感

高精細型インタースペースのプロトタイプによるヒューマンインタフェースの詳細な評価はこれから着手するところである。現在プロトタイプを利用しながら、利用者の反応を見ることで、定性的な評価およびシステムの改良を行っている。

4. 2. アバタ高度化による効果

アバタ高度化による効果としては、体の形態がふくらみを持った実際の人間の体つきに近くなったことによる効果と、顔を実際の利用者本人の顔モデルにしたことによる効果に分けて考えられる。

◎人間の体型に近くなったことによる効果

- 1) アバタの違和感が少なくなる
- 2) アバタの向きがわかりやすくなる
- 3) 2) と関連して特に顔の向きが明確になる

などがあげられる。3) の項目は当初想定していなかった効果である。顔の形を3次的に再現することで、顔の向き（視線方向）の判定が容易になった。これは、複数利用者による会議形態の会話や遠方の利用者に対する注意喚起などの際に有効に作用すると思われる。また、口の開閉を行ったことで、発話者の同定が改善された。これは後に述べる3次元音場との相乗効果がある。

◎アバタ動作の高度化による効果

アバタの動作を人間の動作に近くしたことにより、従来ロボットの動作のような印象を与えていたアバタの振る舞いの違和感が少なくなった。また、首の動きを独立させたことで、前項とあわせて、視線方向の変化を再現させることができた。



図1. 従来のインタースペース

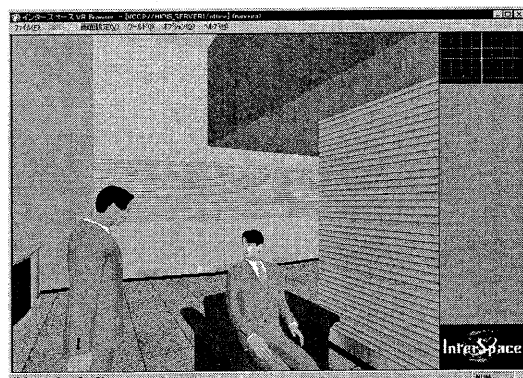


図3. アバタのポーズ例



図2. アバタの顔
ア) 口を閉じたところ



図4. 端末利用中のようす



図2. イ) 口を開いたところ

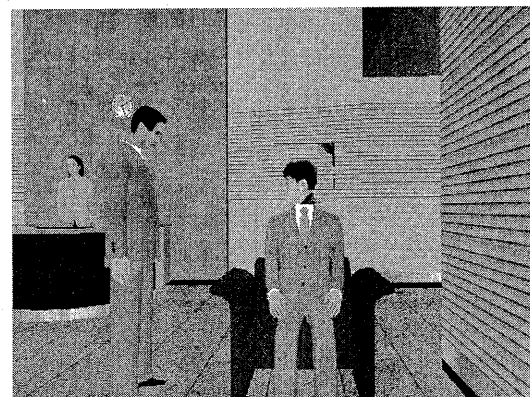


図5. 横顔映像の表示

◎本人の顔に近くなったことによる効果

- 1) 視認性がよくなる
- 2) 違和感が増す

1) については、特に遠距離にいる場合などでも、相手の判別がしやすくなる。しかし、相手の顔に近くなるほど、実際の顔との違いが強く感じられ、「気持ち悪い」という感想を抱く利用者もある。アバタの体のモデルに関してはそれほど指摘を受けなかったのは、「制服」的に捉えられているためと思われる。これは、従来のインタースペースでは見られなかった感想であり、心理的な要因分析を行う必要がある。

4. 3. 顔画像表示による効果

クロマキー合成により顔画像を切り取り顔の位置に表示した効果としては、従来のインタースペースで生じていた、相手の背景映像の混在が抑制され、表示が見やすくなったことが挙げられる。しかし現時点では顔の表示位置精度などに問題があり、さらに改良を行っている。

4. 4. 3 次元音場の効果

3次元音場により他の利用者の音声を制御することで、音源の距離および方向が識別できることを確認した。これにより、複数利用者の音声を同時に聞いている場合でも話者の判定が容易になった。また、相手との距離がごく近い場合には音源は中央に集中するが、この場合は口開閉機能により話者識別を助けることができる。

5. 今後の検討課題

5. 1. 評価項目

高精細型インタースペースのプロトタイプを使用した評価項目としては、前記のような利用感が得られることから、

- 1) コミュニケーション環境
- 2) アウェアネス

の両面からの評価を行う。

コミュニケーション環境としては、①モデルが実際の相手と似たことにより仮想空間で会話することへの心理的障壁が減少したかどうか、②顔映像の表示を行うことで会話の伝達能力が増強したか、③3Dのモデルと顔映像を接合することで違和感が増大したかどうか、などを検証する必要がある。アウェアネスの評価としては、3Dモデルおよび顔映像それぞれでの視線方向の識別能力を検証する必要がある。

5. 2. システムの課題

現在のところ顔映像と3Dモデルの接合処理、ポーズの種類、動作指示のユーザインタフェースなどに改善を要すると考えている。また、利用者の顔モデルおよびポーズデータの動的生成技術もさらに研究を必要とする。

参考文献：菅原他、「多人数参加型環境を実現した三次元サイバースペース～インタースペース TM のアーキテクチャ」仮想都市研究会第1回シンポジウム、日本VR学会、1997年7月