

# 音声から画像へのメディア変換を用いた サイバースペース上での多人数コミュニケーションシステム

四倉 達夫                      小林 智典                      藤井 英史                      森島 繁生  
Tatsuo YOTSUKURA Tomonori KOBAYASHI Eishi FUJII Shigeo MORISHIMA  
{tatsuo, koba, eishi, shigeo}@ee.seikei.ac.jp

成蹊大学工学部 電気電子工学科 情報通信研究室  
〒180 東京都武蔵野市吉祥寺北町3-3-1  
Tel. 0422-37-3742, Fax. 0422-37-3871

## 1. はじめに

コンピュータの発展に伴い、近年コンピュータ上にサイバースペースを作り出し、あたかもその環境にいるような感覚を生み出させるバーチャルリアリティ技術が急速に進歩してきた。これは仮想空間をより現実感を持った空間に近づけ、そこで人間同士のコミュニケーションをリアルに表現することが要求される。

そこで自分自身を投影した人間同様の顔を持つ分身としてのエージェント(avatar: アバタ)を仮想空間上に置き、マイクから入力された自然音声の分析から会話時の口の形状、表情をリアルタイムに合成を行う実時間メディア変換システム<sup>[1]</sup>を基にし、サイバースペース上での多人数コミュニケーションシステムを実現した。

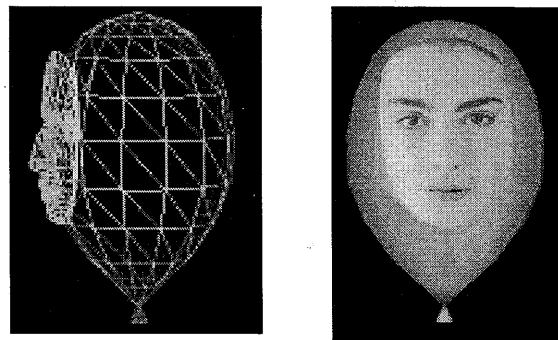


図1-3 アバタ本体

## 2. アバタのモデル化

リアルな顔画像をもつアバタを生成するため、対象人物の正面顔画像に三角形ポリゴンで構成させる顔モデルを整合する(図1-1)。このモデルにテクスチャマッピングを施すことで合成画像を作成する。また、アバタ本体は風船形状の3次元モデルを用いた(図1-2)。このモデルを用いることによりアバタの動き、各感情表現を風船の形状変化、動きにより対話相手にわかり易く伝えることができる。顔画像の合成モデルと風船形状モデルを合体させ、アバタを図1-3の様に構築した。

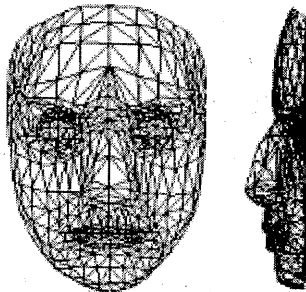


図1-1 3次元顔モデル

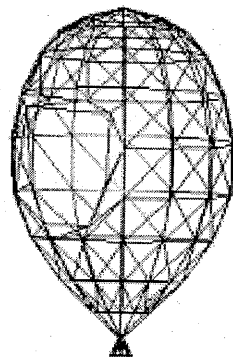


図1-2 風船モデル

## 3. システム構成

図2にサイバースペース上での3人でのコミュニケーションシステムのシステム構成を示した。各プロセスの働きは以下の様になる。

### (1) 音声入力プロセス (Client)

自Playerの自然音声をもとに16KHz、16BitでA/D変換を行い、ネットワーク制御プロセスにより、Ethernetを通じてServerに送出させる。

### (2) 音声分析プロセス・パラメータ変換プロセス (Server)

各Clientから送られてきた自然音声をもとに音声分析プロセスで線形予測分析(LPC分析)を使い、各PlayerのLPCケプストラムを算出し、また感情推定のためにピッチ分析も併せて行う。次にパラメータ変換プロセスでニューラルネットワークを用いて、発話時の口の形状を規定する口領域の変形パラメータ(以下、口形パラメータ)に変換し、ピッチ情報は6基本感情である喜び、怒り、嫌悪、恐れ、悲しみ、驚きの表情変形パラメータ(以下、表情パラメータ)へと変形される。音声分析プロセス・パラメータ変換プロセスは各Playerごとに別プロセスで実行される。

### (3) 位置情報入力プロセス (Client)

マウスによって決定された自Playerの位置情報を獲得しServerに送出させる。

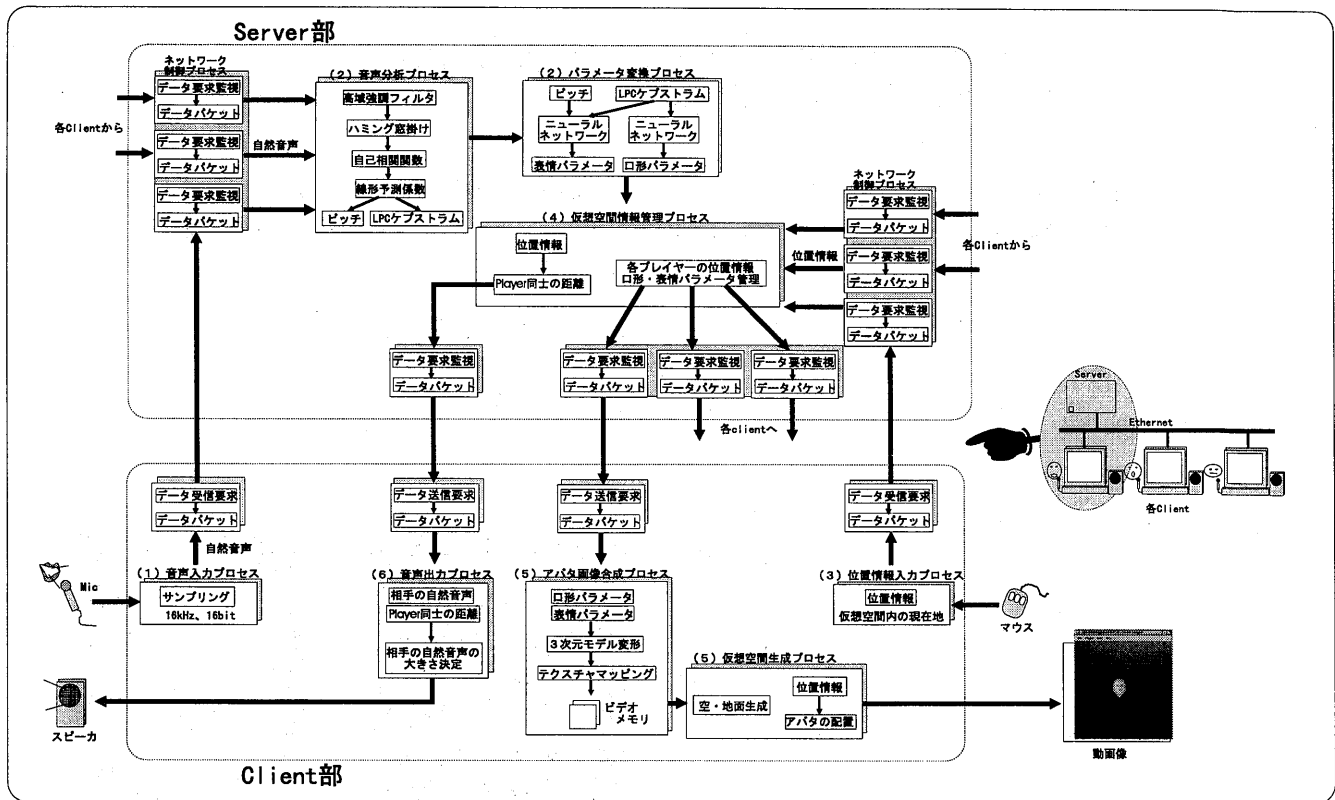


図2 システム構成

**(4) 仮想空間情報管理プロセス (Server)**

A: 送られてきた各Playerの位置情報と(2)で求めた口形・表情パラメータをそれぞれのClientに送出する。

B: 位置情報を元に仮想空間内でのPlayer同士の距離を算出し、送り先以外の自然音声と共にClientへ送出する。

**(5) アバタ画像合成・仮想空間生成プロセス (Client)**

(4): Aから受け取った各データをアバタ画像合成プロセスで顔モデルを変形・移動し、それにテクスチャマッピングを行い、風船も表情パラメータによって動作させPlayerごとの動画像を生成する。仮想空間生成プロセスにおいて空間内に空、地面を生成、そしてPlayer達の位置情報から各アバタを配置していく。Displayの表出方法は2種類あり、1つは自分のアバタから見た視界(図3-1)、もう1つは仮想環境内の上空から見た視界(図3-2)を用意した。各視界はウインドウ内のメニューバーによって選択できる。

**(6) 音声出力プロセス (Client)**

(4): Bから受け取った距離より、相手が今どのくらいの場所にいるかが分かる。その情報から相手の自然音声に強弱をつけ、音声に遠近感を持た

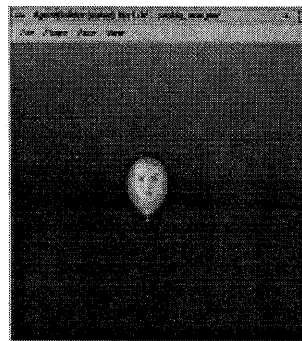


図3-1 アバタから見た視線

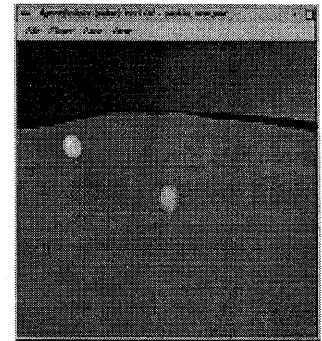


図3-2 上空から見た視線

せ、スピーカに出力する。

**4. まとめ**

自然音声から口形・表情パラメータを求めリアルティのあるアバタを用いた仮想空間上での多人数によるコミュニケーションシステムを構築した。現時点は3人間、LAN上という閉鎖的空間内でのみ動作可能である。今後はより多い人数でかつ、通信での肥大化した情報量を軽減させInternet上での運用を目指す。

**参考文献**

[1] 宮下、岩澤、坂口、森島「仮想人物とユーザとの対話を実現するための音声から画像への実時間メディア変換システム」信学会総合大会 p403(1996)