

人体の動作を利用した擬人化エージェントに対する考察 — モジュール型人体動作生成とその応用について —

呂山, 吉坂 主旬, 神谷 俊之, 一色 敬, 宮井 均

NEC 関西 C&C 研究所

本稿では、ノンバーバルコミュニケーションの一部である身体の動作を中心とした擬人化 IF の考察、及び実際の人体動作を分析し、抽出した動作パターンを基にしたモジュール型の人体動作生成システムについて述べる。また、アプリケーションとして、本システムにより得られた人体動作などを利用して擬人化エージェント — 電子図書館における CG 司書を試作した結果について説明し、擬人化エージェントの有効性を検証する。試作した CG 司書は、主に胴体、頭部の動き、及び音声による対話などの擬人的要素を備えたものである。

A Study of Anthropomorphous Agent with Human Motion — Human Motion Generation System and its Application —

Shan Lu, Shujun Yoshizaka, Toshiyuki Kamiya, Takashi Isshiki, Hitoshi Miyai

NEC Kansai C&C Research Laboratory

In this paper, we discuss anthropomorphous agent technology and the required factors to design anthropomorphous interfaces by use of it. Then we also describe a human motion generation system based on some motion modules, which are obtained by analyzing actually the human motion. Finally, we report the results applying the anthropomorphous agent to CG-librarian as the user-interface of the electronic library. The CG-librarian has some human-like motion features such as body-motion and verbal-dialogue, etc.

1 はじめに

近年、グラフィックスユーザインタフェース (GUI) に次ぐ次世代ユーザインタフェース (UI) として、人間社会で通常行われているコミュニケーションの形式に近づける UI の研究が盛んに行われている [1, 3]。この中で、バーバルコミュニケーションはもちろんのこと、ノンバーバル (非言語) コミュニケーション、つまり、人間の表情や動作などによるコミュニケーションも非常に重要視されつつある。こうした音声や表情、動作など、複数の対話チャンネルを用いて、コンピュータとのインタラクションの円滑化を図るものとして、マルチモーダルコミュニケーション、あるいは、擬人化エージェント、擬人化インタフェースといったユーザインタフェースが多く提案されている [2, 3, 4]。

本稿では、ノンバーバルコミュニケーションの一部である身体の動作を中心とした擬人化 IF と、それに基づき試作した電子図書館の UI について述べる。以下の章では、まず、擬人化 IF の考え方、UI の中に用いられる人体動作に対する分析、そして、開発したモジュール型の人体動作生成システム、最後に現在試作中の電子図書館システム [5] の UI の一部である CG 司書について説明する。

2 擬人化エージェントについて

コンピュータエージェントの擬人化については、二つの観点がある。一つは、コンピュータそのものを一個の人格として見ることで、つまり、「コンピュータの人格化」である。もう一つは、「自然言語による対話のしやすさや、エージェントに対する直感的な理解など」を特徴とするインタフェースとしてとらえるものである。前者は機械を人格化するのに対して、後者は、擬人化の視覚性を巧みに利用している。

筆者らは後者の視点から擬人化エージェントをとらえ、研究を進めている。一般に、擬人化 IF は、コンピュータの内部構造や動作をユーザから隠し、また、人間の顔表情や動作、及び音声などの人間社会で普段行われているコミュニケーション手段を採り入れ、より柔軟、自然なインタフェースを実現する可能性が存在する。その効果として、以下のものが期待されている。

- 人間の日常的コミュニケーションに近い形の

擬人化 IF は、素人でも簡単に操作できるという意味で、アイコンベースの GUI を超える可能性がある。

- ビジュアル的な情報 (動作、表情) 提示により、計算機に対する直感的な認識が可能であり、人間とコンピュータとの円滑な情報交換が可能となる。
- ノンバーバルコミュニケーションを含めた各種のインタラクション情報を活用することができる。
- 専門的な知識を持たせた知的なエージェントにより、機械の代行として、ユーザの問い合わせなどに柔軟に対処することができる。

擬人化したエージェントを設計する際、B. Laurel 氏は、次の二つの基本的条件を特に強調している。 [6]。

- (1) エージェントを人間ではなく、キャラクターであると考えよ。
- (2) エージェントは、信頼が置け、容易にアクセスできるものでなければならない。

つまり、擬人化エージェントにとっては、実際の人間の振舞いを模倣するのではなく、人間にとって、分かりやすいキャラクターの演出が重要となる。

このような可能性を有する擬人化 IF を実際に構築する際、現在利用可能な音声認識技術や CG 技術を考慮して、擬人化エージェントに対して、以下の基本的な要件が考えられる。

- (1) ユーザとエージェントの対話は、音声によって行われる。完全な自然言語の音声認識と理解はまだ実現していないが、ある程度、限定された範囲での音声による入力が可能である。
- (2) エージェントのキャラクター性を高めるため、動作などをある程度リアリティのある視覚的な表現とする。
- (3) さらに、ユーザにエージェントの状態などを容易に認識させるため、表情変化を持たせる。

本稿では、上述した擬人化エージェントの (1) と (2) の目標を中心に開発したシステムについて述べる。

3 人体動作に対する考察

前述したように、筆者らは、擬人化エージェントの演出性を表現するため、エージェントに人体モデルを持たせ、そのモデルに動作や表情などの擬人化要素を持たせることに着目した。当対象とするエージェントは、人間との対話に用いることを想定しているため、人体の頭部と胴体を含めた上半身の動きが中心であると思われる。そこで、この章では、その基本となる人間の上半身動作の分析を行う。まず、会議中の演説者や聴講者などの全身の動きと上半身の動きの比較を行い、上半身運動の特徴を分析した結果について簡単に説明する。さらに、実際のテレビ放送におけるニュースキャスタや人形劇などを素材として、人物の典型的な動作を分析した結果を述べる。

3.1 上半身動作の特徴

上半身と全身の動作との比較にあたって、会議などのコミュニケーション場における人間の動きを題材とした。特に座って話す人の動きと歩きながら話をする人の動きに注目して、実際の会議などの状況を撮影したビデオ画像を分析して、表1のような違いを抽出した。

表 1: 上半身動作の特徴

項目	上半身	全身
周期性	ない	ある
動作の種類	少ない	多い
各部位の同時動作	少ない	多い
胴体動作の重要度	大きい	小さい

(1) 周期性

歩行、走行や跳ぶなどの全身に関わる動作を行う時、足や手などの部位は同じ動きを繰り返すことが多く、周期的な動作特徴が見られる。それに対して、上半身の動きは、体軸(背骨)を中心に、左右、前後の方向へ一方向の動きが多くて、繰り返しの動きがほとんどない。

(2) 動作の種類

手、足の動きを含めた全身運動では、手や足の動き自由度はかなり大きくて、動作の種類

は多い。それに対して、上半身の動きでは、脊椎が中心的な動きをする。脊椎の動作範囲は狭くて、上半身の動作としてはあまり多くの種類がない。

(3) 各部位の同時動作

ここでの同時動作というのは、各部位が協調して動いて、最後に一つの動作を完成するものを意味する。全身運動の場合、身体全体のバランスを取るため、足を動かす時と同時に、手や胴体も動かさなければならない。座っている状態での動作では、上半身の各部位がかなり安定かつ独立に動くことが多い。

(4) 胴体動作の重要度

全身運動の場合、胴体が単に一つの部分にすぎず、胴体の細かな動きがさほど重要ではない。それに対して、上半身の場合、胴体が上半身の主要な部分を占めているので、リアリティをあげるためには、忠実にその動きを模倣する必要がある。

以上のような上半身の動作の特徴から、人体の動きモデルに基づいて、動作の生成を行う際、以下の点に注目する必要があると考えられる。

(1) 上半身の動きの種類は比較的小さいため、動きのパターンを抽象化した表現が可能である。

(2) 人体の全身動作を考える際、胴体を一つ、あるいは、2、3のセグメントとして処理する機会が多い。それに対して、上半身の動きには脊椎骨が中心的な役割を果たしているため、比較的精密に脊椎骨の動きを生成する必要がある。

(3) 上半身の動きの場合、体全体の協調性がそれほど強くなく、ある程度、各部分の独立な動き表現が可能である。

上述した内容から、人間がコミュニケーションする際、上半身の動きを一定のパターンの組み合わせで表せばよいことが分かる。具体的には上半身のそれぞれの部位の動きをモジュールとして定義し、複数のモジュールの組み合わせ、合成によって、動作パターンの生成が可能となる。

以下、動作パターンというのは、一定の意味を含む人間の動作を指すものであり、例えば、”挨拶

拶”, ”同意” や, ”否定” などの意思を表すときの動きを意味するものとする。また, 動作モジュールは, 身体各部分の動きを記述するいろいろな運動パラメータをまとめたものであり, 基本的には, 人体の構造に対応して分類したものであり, 例えば, 上半身では, 頭部, 首, 胸, 腰の動きなどである。

3.2 動作パターンの解析

人物の動きは通常, 個人差があつて, 詳細に観察すると, 完全に一致することがないが, ノンバーバルコミュニケーションにおける動作の意味分析の研究 [7] によると, 共通の動作パターンが存在すると報告されている。これと同様の観点から, 人間のコミュニケーション時の人体上半身の動きについて分析, 分類を試みた。具体的には, ニュース, 映画, 人形劇などの映像を VTR に録画して, 人間 (人形) の目, 頭, 胴体の動きを観察して, これらの動作が行われた状況や意味を分析したものである。

まず, 人体のそれぞれの部分, すなわち, 目, 頭部や胴体の単純な動作を部分動作として分析した。その主な結果については, まず, 目には左右, 前後及び開閉の運動があり, 頭部は, 主に前後, 左右へ傾ける運動になっている。胴体の動作はさらに, 首部, 胸部及ぶ腰部の動きに分解することができ, 脊椎全体としては, 前方と後方に曲げる動きが主である。

表 2: 連続動作パターンの例

連続動作	説明	コード
うなづく	同意する, 言葉の区切り	MP-101
頭を下げる	考える, 打ち明ける	MP-102
顎を上げる	考える, おこる	MP-201
頭を振る	否定する	MP-401
頭をひねる	疑問に思う	MP-402
胸をはる	言い放つ	MP-204
顔を出す	同意する	MP-302

このような部分動作の組み合わせから合成される連続動作の部分パターンを表 2 に示す。これらの連続動作のパターンは, 頭部, 首もしくは胴体の動きから合成される。表 2 中の MP-xxx は, シ

ステムの中に使われる動作パターンを識別するコードで, 特別な意味を持たない。

表 3: 感情表現と動作パターン

感情表現	動作パターンコード
同意	MP-101, MP-302
考える	MP-102, MP-104, MP-402, MP-403
強調	MP-103, MP-108, MP-303, MP-305
挨拶	MP-105, MP-107
謝る	MP-105
質問	MP-105, MP-301, MP-304, MP-305, MP-306
否定	MP-401
怒る	MP-201
言葉区切り	MP-105, MP-201
言い放つ	MP-204, MP-404
打ち明ける	MP-203

表 3 には, これらに基づいて, 会話や感情表現時の動作表現を, 動作パターンコードの組み合わせで示している。表から分かるように, 同じような感情を表すときでも, いろいろな動作表現方式があり, 逆に同じ動作パターンで, 異なる感情を表現することもある。表 3 中の MP-xxx は表 2 中の動作パターンのコードである。

4 モジュール型の動作生成システム

この章では, 動作モジュールから前述の動作パターンを生成するシステムを説明する [8]。

人体動作の生成にあたって, 人体モデルは必須であり, 本研究では, モデルとして人体の骨格構造を反映したスケルトンモデルを用いた。上述した様に, 上半身の動きの中では胴体の動きが中心的であるので, モデルは, 解剖学の実測データを用いて, その動きを支配する脊椎の構造を詳細にモデリングした。

4.1 動作生成システムの構成

動作モジュールから動作パターンを生成するシステムの構成を図 1 に示す。このシステムは, 大きく分けて, 動作パターン解析ユニット, 動作モ

ジュール制御ユニット、フレームデータ生成ユニット、及びフレーム描画ユニットの四つの部分から成る。前述した動作パターン名を入力として、最後にパターンの動作シーケンスデータが出力される。以下、それぞれのユニットについて説明する。

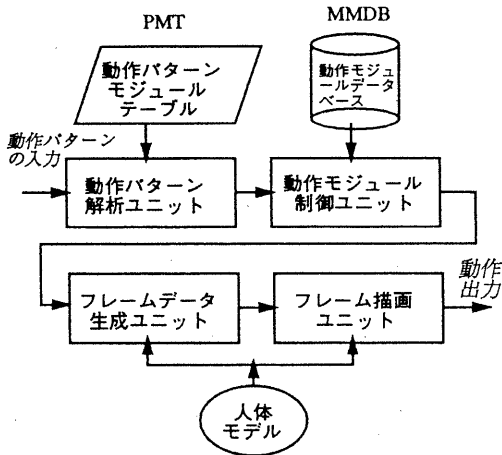


図 1: 動作生成システムの構成

(1) 動作パターン解析ユニット

本ユニットは、外部から指定された動作パターンを解析して、そのパターンを生成するための必要な動作モジュールの情報を出力する。動作パターンの解析の際、動作パターンモジュールテーブル (PMT と呼ぶ) を参照する。PMT は、動作パターンを構成する動作モジュールの名称及び動作モジュールの動作範囲を表すデータから成る。したがって、動作パターン名を用いて、PMT をアクセスすると、対応する動作モジュールの種類及び動作の範囲を得ることができる。

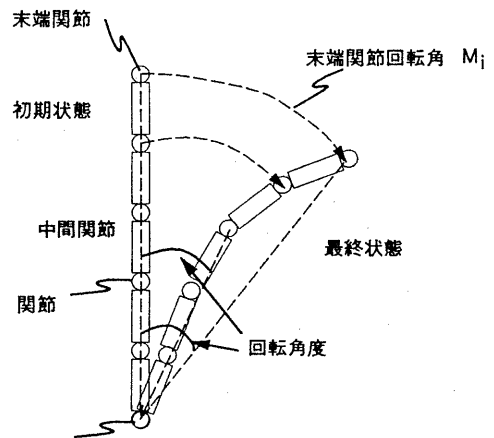
(2) 動作モジュール制御ユニット

このユニットでは、前のユニットから出力された複数の動作モジュールを実際にどのように動作させるかを決定する。まず、図 1 に示す動作モジュールデータベース (MMDB と呼ぶ) を参照して、動作モジュールのデータを引き出す。そのデータに基づいて、各動作モジュール間の時間と空間の制御情報を生成する。

この動作モジュールについて詳細に説明する。

前述したように、動作モジュールは、人体の部分関節の動きをまとめたものである。部分関節の

動きは、一番下に位置する関節を回転の中心として、他の関節の一定範囲の回転運動によって合成される (図 2)。一番先端の関節の回転の角度 (末端関節回転角度という) は、この部分関節の大まかな動きをきめる。PMT 中のモジュールの動作範囲は、この末端関節回転角 (M_i と呼び、インデックス i は動作モジュールの番号を示す) として定義される。しかし、同じ末端関節の最終状態に達成するための中間関節の可能な動きの種類は、膨大な数にのぼる。図 3 には、同じ末端関節の最終状態に対して、4 種類の中間関節の回転タイプを示している。このような多くの種類の回転タイプの中から一意的に中間関節の回転を決めるため、動作モジュールの動き情報として末端関節の回転角度に対する中間関節の動きを表すパラメータ (中間関節回転係数 $w_{i,j}$ と呼ぶ。 i は動作モジュールの番号、 j は中間関節の番号を示す) を用いる。



回転の固定点

図 2: 部分関節の動きを記述する動作モジュール

この係数の値は 0.0 から 1.0 の間にある。 $w_{i,j} = 0.0$ は、 i 番目の関節の動作に対する寄与はゼロ、つまり動かないことを表す。 $w_{i,j}$ が大きくなるほど、全体の動きの中で、その関節の動きが大きくなり、 $w_{i,j} = 1$ は、その関節の動きが最大限であることを示す。動作モジュールの種類に対応して、異なる中間関節回転係数 $w_{i,j}$ が設定される。

モデルの関節の初期状態を P_i とすると、動作パターン解析ユニットより得られたモジュールの動作範囲 M_i と中間関節回転係数 $w_{i,j}$ を用いて、次式で、関節の最終状態が計算される。

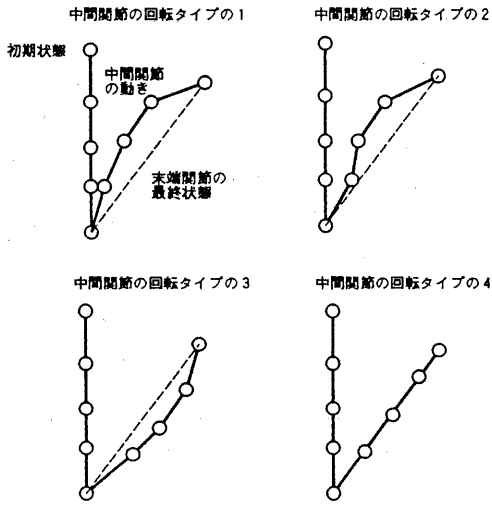


図 3: 中間関節の回転タイプ

$$P_i^* = P_i + w_{i,j} M_i$$

こうして、動作モジュールの初期状態 P_i から最終状態 P_i^* を計算することができる。

また、この部分関節の動作速度を表す実行時間は動作モジュールの中に含まれる。このように、MMDB の中に、各動作モジュールの中間関節回転係数 $w_{i,j}$ 、及びモジュールの動作に必要な時間を格納する。このモジュールの実行時間から、複数の動作モジュールのタイミング制御データを生成する。

(3) フレームデータ生成ユニット

このユニットは、動作モジュール制御ユニットから得た関節の状態データと時間制御データを用いて、すべてのフレームにわたる関節の状態を計算する。

C_i^0 と C_i^* を、初期状態と最終状態のモジュールの関節の座標とすると、その間の任意時刻における関節の座標値 C_i^t は、線形補間法

$$C_i^t = C_i^0 + t * (C_i^* - C_i^0) / T$$

で計算する。式中の t と T は、それぞれ、ある時刻の値とそのモジュールの実行総時間である。

(4) フレームデータ描画ユニット

フレームデータ生成ユニットで生成された各フレームの関節位置 C_i^t 、及び人体モデルからの外形データを用いて、連続的に動作のシーケンスをディスプレイに描画する。

こうして、動作パターン名を入力として、動作のシーケンス映像を生成することができる。

5 「Virtual Library(仮想図書館)」における CG 司書

前述した人体動作システム的应用として、開発中の電子図書館システム「Virtual Library」のユーザインタフェース部に、人体動作や表情、音声などを持つ擬人化エージェント (CG 司書と名付ける) の試作を試みた [9]。

以下、「Virtual Library」の UI の構成、及び CG 司書の実現方法について述べる。

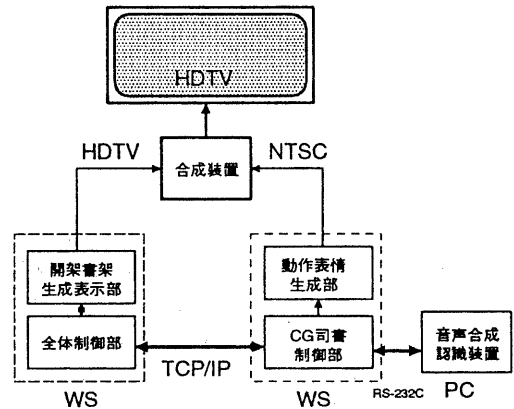


図 4: 「Virtual Library」の UI 部の構成

5.1 「Virtual Library」UI 部のシステム構成

「Virtual Library」の UI 部の構成を図 4 に示す。システムは、全体制御部、仮想開架書架表示部、CG 司書制御部及び音声合成認識装置から構成されている。

ここで「仮想開架書架」というのは、CG により画面上にあたかも開架式の図書館のような画像を生成表示するものである。出力端末としてハイビジョンを用いている。「仮想開架書架」の生成は書架位置情報などの基本データと、書籍の形状

データ、表紙画像データ等をもとに行われる。

CG 司書は、「Virtual Library」の中でシステムのヘルプ、システムへの入力受け付け及び検索の補助を行うユーザ・システムの仲介役として位置づけられる。従来の研究では図書館の中を利用者がガイドなしで自由に検索を行う手法が提案されている。しかし、この手法は利用者がシステムに不慣れな場合の不安感や、自分の現在位置／状態が分からなくなるといった状況が発生することが問題となる。CG 司書はメニュー階層の各時点で可能な入力についてのガイダンスを行い、利用者からの求めに応じてヘルプのメッセージの応答を行うことを想定し、さらに、CG 司書に実際の司書のような知識を持たせたエキスパートシステムを連携しユーザからの検索要求を解釈し効率的な検索を行う場合の補助を行うことを考えている。

現在のシステムでは、身体の動作を伴う音声出力による「Virtual Library」の案内、及びユーザの音声入力による書架の検索を行う CG 司書の部分機能を実装している。

開架書架と CG 司書の出力画面例を図 5 に示す。



CG 司書

仮想開架書架

図 5: CG 司書と開架書架の画面例

5.2 CG 司書の構成

前述した機能を実現するための CG 司書のシステム構成を図 6 に示す。このシステムは、大きく分けて、以下の部分から構成される。

通信サーバは、ユーザに対する入力要求や、ユーザ入力の音声認識の結果を全体制御部への通信機能を果たす。対話制御部は、CG 司書とユーザとの対話をコントロールする。全体制御部から、書架や検索などのユーザに対する入力を要求する場合、対話制御部が、音声認識部に命令を送り、音

声認識入力待ちの状態に入る。また、検索の結果や案内などが送られてくるとき、これらのメッセージを音声合成部に送り、音声合成の出力を行う。メッセージ生成部は、CG 司書とユーザとの対話を行うとき、必要なテキストメッセージを生成する。

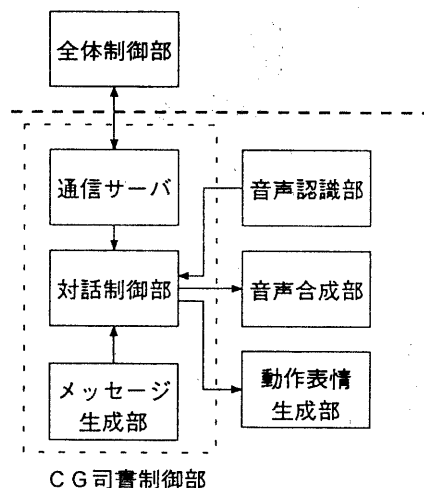


図 6: CG 司書のシステム構成

図 5 に示された CG 司書が、案内の文書を読み上げたり、ユーザの検索要求を応答したりすることによって、システムのガイド機能を実現する。ユーザが図書分類名、例えば、「歴史」、「工学」といった名前を音声により指示することによって書架の選択を行う。以下、CG 司書がユーザとの対話の一例を示す。

司書：いらっしゃいませ。「Virtual Library」についてご説明致しますでしょうか？

ユーザ：はい。

司書：本電子図書館では、仮想書架を探索することにより、...

それでは、書架へご案内致します。書架名をおっしゃって下さい。

ユーザ：計算機関連。

司書：計算機関連ですね？

ユーザ：はい。

司書：それでは、計算機関連の書架へご案内致します。

.....

図 7 に、CG 司書のイメージを示す。

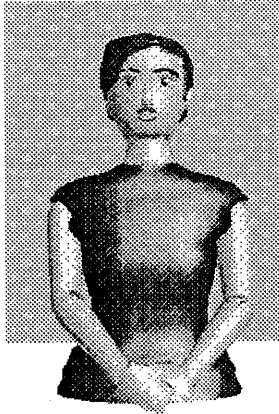


図 7: CG 司書の画面イメージ

6 おわり

本稿では、人体動作の生成システム、電子図書館における CG 司書について述べた。人体動作生成システムについて、2 章で述べた頭部と胴体の部分動作を動作モジュールとして、いろいろな動作パターンを生成することができ、人間が普通に対話するときの動作、例えば、頷き、挨拶、否定などの動きに、ほぼ対応できることが確認できた。このシステムに新たに動作モジュールと動作パターンを追加すれば、新しい動作の生成を容易に行うことができることになる。また、擬人化エージェントの応用例として試作した CG 司書は、胴体、頭部の動き、単純な口の動きの開閉、及び音声対話を行うような、単純な擬人化 IF により実現されている。検索や案内を行う場面において、単調なテキストメッセージの入出力をしている場合に比べて、より”親しみ易く”、”優しい”インタフェースになる効果があると思われる。

最後に、今回の試作した人体の動作生成システム及び擬人化エージェントの試みである CG 司書の問題点と今後の課題について述べる。

- (1) 3 章の動作生成システムは、個人差を表現できない。つまり、動作モジュールや動作パターンを、ほとんど複数の人の動作を分析した平均値から定義したので、個人差を反映する手段を提供していない。
- (2) 試作した CG 司書は、胴体と頭部の動きに限定しているが、手の動きや、顔の表情も非常に重要な感情、意思を表現する要素であり、

今後の課題である。

- (3) CG 司書の状態を、例えば、司書の顔の表情を変化させることによってユーザに提示することが有効だと考えている。
- (4) CG 司書は、ビジュアル的表示だけではなく、知的エージェントとして、効率的な検索とガイダンスなどのサービスを提供する機能も必要である。

参考文献

- [1] 末永康人, 他. 人物像と音声によるインタフェースの枠組 (human reader). *NTT R&D*, Vol. 42, No. 4, pp. 445-454, 1993 年.
- [2] 瀬戸重宣, 他. 実時間音声対話システムにおけるマルチモーダル応答. In *8th Symposium on Human Interface*, pp. 693-698, 1992 年.
- [3] 長谷川修, 他. 自然感の高いビジュアルヒューマンインタフェースの実現のための人物動画像の実時間並列協調的認識. *信学会論文誌 D-II*, Vol. J77-D-II, No. 1, pp. 108-118, 1994.
- [4] 長尾確, 竹内章一. コンピュータとの自然な対話のための新しいモダリティ 人工知能研究会資料 (SIG-SLUD-9204), pp. 9-16, 1993.
- [5] 一色敬, 他. 電子図書館システム「virtual library」の概要. 情処第 48 回全国大会, pp. 1.341-342, 1994.
- [6] Brenda Laurel. 劇場としてのコンピュータ, 第 5 章, pp. 167-177. トップラン, 1992.
- [7] Gary Bente. Facilities for the graphical computer simulation of head and body movements. *Beh. Res. Meth., Ins., & Comp.*, Vol. 21, No. 4, pp. 455-462, 1989.
- [8] 呂山, 他. 人体動作生成システムの提案. 情処第 47 回全国大会, pp. 2.345-346, 1993.
- [9] 神谷俊之, 他. 電子図書館システム「virtual library」のユーザインタフェース. 情処第 48 回全国大会, pp. 1.343-344, 1994.