複数人参加型インタラクティブ映画システムの構成

中津　良平　　土佐　尚子　　越知　武

ＡＴＲ知能映像通信研究所
〒619-02　京都府相楽郡精華町光台２丁目２番地
Tel：(0774) 95 1400/1427/1494　Fax：(0774) 95 1408
{nakatsu, tosa, ochi} @mic.atr.co.jp

あらまし：映画、通信、ゲームなどのメディアを組み合わせることにより作り出される新しいメディアとしてインタラクティブ映画がある。ここでは、まず我々の開発した第１次システムについて簡単にふれると共に、それが、（1）インタラクションの頻度が少ない、（2）参加者が１人に限定される、という問題を持つことを示す。我々はこの問題を解決するために第２次システムを作成中である。本システムは、２人の人がネットワークを介して参加可能であると共に、任意の時点でのインタラクション（anytime interaction）が可能であるという特徴を持つ。システムのソフトウェア構成、ハードウェア構成の詳細を述べると共に作成した作品例を紹介する。

キーワード：インタラクティブ映画、映画、通信、ゲーム、バーチャルリアリティ、インタラクション

Construction of an Interactive Movie System for Multi-person Participation

Ryohei Nakatsu, Naoko Tosa, and Takeshi Ochi

ATR Media Integration & Communications Research Laboratories
2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-02 Japan
Tel : +81 774 95 1400/1427/1494, Fax : +81 774 95 1408
E-mail : {nakatsu, tosa, ochi} @mic.atr.co.jp

Abstract : Interactive movies, in which interaction capabilities are introduced into movies, is considered to be a new type of media that integrates various media, including telecommunications, movies, and video games. In this paper, we first explain the concept of interactive movies and briefly explain the prototype system we developed. We point out several problems in the system that need to be improved : the lack of frequent interactions and the participation limited to a single person. We are currently developing a second system. We explain the two significant improvements : the introduction of interaction at any time and two-person participation through the use of a network. We describe the software and hardware configuration of the second system and introduce one example of an interactive story installed in this system.

Key word : interactive movies, movies, telecommunications, games, virtual reality, interaction

1．まえがき

１９世紀末にルミエールらによって発明されて以来 [1]、映画は技術およびコンテンツ制作の面で進化を繰り返し、現在では、アートからエンターテインメントまでの広い領域を含む複合芸術としての地位を確立している。映画は、ストーリーテリングの技術をベースとして映像・音声による仮想世界（サイバースペース）の構築技術を加味することにより、人を仮想世界に引き込みそこにおけるストーリーを体験させるという力をもつメディアである。それと同時に、映画はそれ以上の可能性を秘めている。それは映画にインタラクション技術を導入することである。従来の映画は、あらかじめ定められた状況、ストーリー設定を観客に一方向的に与えるという形態をとっていた。これに対し、インタラクション技術を取り入れると、観客自身が主人公となって、主体的にストーリーを体験することが可能になる。

このような観点から、筆者らは従来の映画にインタラクション技術を導入したインタラクティブ映画システムの検討を行なっている。すでに、その第１次システム [2] を試作した。第１次システムをベースとして、現在改良版である第２次システムを構築中である。本報告では、第１次システムの簡単な内容を述べると共に、その問題点、改良すべき点をまとめる。また、このような考察に基づいて現在構築中の第２次システムの構成を述べる。

2．インタラクティブ映画の概要

2．1　コンセプト

インタラクティブ映画は、現在の映画との関連からすると、「観客参加・体験型の映画」といえる。インタラクティブ映画は、以下の要素で構成されている。

(1)インタラクションによってストーリー展開が変化するインタラクティブストーリー

(2)主人公となってインタラクティブストーリーの世界を体験する参加者

(3)主人公とインタラクションを行ないつつストーリー展開に参加するキャラクタ

2．2　他のメディアとの比較

ここでは、インタラクティブ映画と他のメディアとの比較を行なう。

（1）通信

遠隔地の人物および周囲環境を立体映像で再現することにより、face-to-faceでコミュニケーションを行なっているかのような臨場感に富んだ通信を行なうという研究が行なわれている。その代表例がATRで行なわれた臨場感通信会議 [3] である。これはサイバースペースを介したコミュニケーションであるが、あくまでテレビ会議の高度化版との位置付けであり、ストーリーの概念は含まれていない。

（2）映画

映画は、迫力ある映像とサウンドにより、人間を架空の世界（サイバースペース）に引き込もうとしてきた。特に最近はCG技術の急速な発展により、架空の世界、実際にはありそうもない出来事をリアルな映像として作成することが可能になってきた。この映画にインタラクションの機能を付加しようという考えはあったが、これまでの試みは、複数個のストーリー展開を用意しておき、観客にいずれかを選ばせるというプリミティブなものにとどまっていた。

（3）テレビゲーム

テレビゲーム、特にロールプレイングゲーム（RPG）は、小説の世界をゲームに仕立てたものということが出来る。基本的なストーリーが設定してあり、人はゲームの主人公を操ってストーリー展開をコントロールすることが出来る。この意味でテレビゲームはインタラクティブ映画との類似点が多いが、インタラクションがボタン操作で行なわれ、人間同士のインタラクションと異なるという本質的な相違点がある。

（3）その他のメディア

サイバースペースを構築し、その世界やキャラクタと観客とのインタラクションを行なわせようとする試みは種々行なわれている。人間との行動レベル、感情レベルでのインタラクション可能なコンピュータキャラクタの生成 [4] [5] や、インタラクティブアート [6] [7] などがそれにあたる。しかしながら、これらのインタラクションは短時間的なものであり、ストーリーの導入はなされていない。

これらのメディアのうち映画・小説、およびテレビゲームとインタラクティブ映画との詳細比較を表１に示す。

2．3　第１次システム構成

上記の考え方の基に第１次システムを構築した[2]。システムの内容を簡単に説明する。

2．3．1　ソフトウェア構成

ソフトウェア構成を図１に示す。

表１　インタラクティブ映画と他のメディアの比較

	目的	観客、プレイヤー	ストーリー	インタラクション
映画	・ストーリーを観客に疑似体験させ精神的快楽を与える	・ストーリー展開の傍観者 ・たくみなストーリー展開により強い感情移入を体験 ・精神的満足感を味わう	・質を決める主要要素 ・観客に次の展開を予測・期待させそれとのずれにより観客を引き込む	・なし
テレビゲーム	・プレイヤーに身体的快楽を与える	・ゲームの主体的参加者もしくは「人形使い」としての役割 ・身体的スキルを高めることに快感をおぼえる	・スキルを高めるための努力をサポートする役割 ・基本的には補助的役割	・ボタン入力 ・最初の設定は大変 ・慣れれば身体的快感につながる
インタラクティブ映画	・観客に主体性をもったストーリー体験を味わわせる ・精神的インタラクションをベースとして観客に強い精神的満足度を与える	・ストーリーを体験する主役 ・「主体的な体験」が映画等と異なる ・精神的満足感を得るという点でゲームと異なる	・観客を引き込むための主要要素 ・精神的満足感を与えるストーリー展開が必要	・音声、ジェスチャによる入力 ・人間にとって自然なインタラクション手段

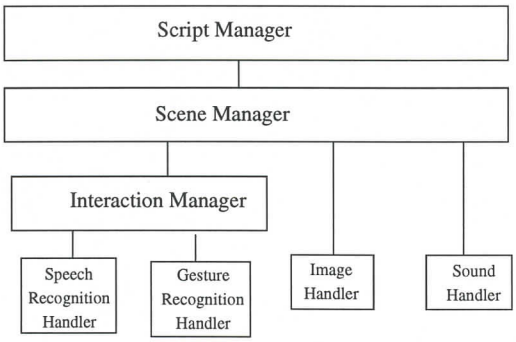

図１　第１次システムのソフトウェア構成

（1）スクリプトマネージャ

インタラクティブストーリーは、種々のシーンの集合およびシーン間の状態遷移図で表現される。スクリプトマネージャは、この状態遷移図を記憶しておき、シーン間の遷移をコントロールする。

（2）シーンマネージャー

あらかじめ各シーンの内容記述はデータとして蓄えてある。シーンマネージャは、スクリプトマネージャから指示されたシーンの記述データを参照し、各シーンの生成を行なう。

（3）インタラクションマネージャー

スクリプトマネージャー、およびシーンマネージャーの下にあって、各シーンにおけるインタラクションを制御する。音声認識機能、動作認識機能に基づいたインタラクションが行なわれる。

（4）各種のハンドラ

シーンマネージャ、もしくはインタラクションマネージャの下にあって、各種の入出力を制御する機能を持つ。音声認識ハンドラは、音声認識機能を制御するハンドラである。音声認識はHMMに基づく不特定話者、連続音声認識機能を持つ[8]。動作認識ハンドラは、動作の認識を制御するハンドラである。動作の認識は、カメラから取り込んだ映像から人物の頭、両手などのの特徴点を抽出する機能を持っている[9]。映像ハンドラは、背景、キャラクタアニメーションなどの映像の出力を制御するハンドラである。サウンドハンドラは、背景音楽、効果音、キャラクタの台詞などの出力を制御するハンドラである。

2．3．2　ハードウェア構成

図2にハードウェア構成を示す。映像出力サブシステム、音声認識サブシステム、動作認識サブシステム、サウンド出力サブシステムより構成される。

映像出力サブシステムでは、CG生成用高速WSを映像出力のために用いている。音声認識サブシステムおよび動作認識サブシステムは各々1台のより構成されている。サウンド出力サブシステムは複数台のPCより構成されており、サウンドの同時出力を行なう。

2．4　第1次システムの評価と問題点

本システムは、開発以来約半年にわたって所内の研究者や当所への見学者など約50名に体験してもらった。これらの体験者の感想などに基づき第1次システムの問題点をまとめると以下のようになる。

（1）サイバースペースへの参加

a）参加者数

第1次システムでは、参加者は1人であり、主人公となってストーリーを体験するというのが基本コンセプトであった。しかしながら、サイバースペースでのストーリー体験という基本コンセプトからすると、第1次システムはその一部の機能しか満たしていない。サイバースペースはネットワーク上に構築されるわけであるから、サイバースペースには1人ではなく複数人が同時に参加してストーリー展開を体験できることが望ましい。

b）アバターの有無

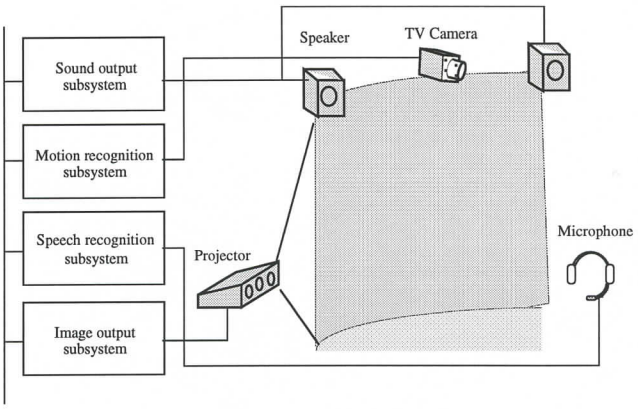

図2　第1次システムのハードウェア構成

参加者がサイバースペースのストーリーに参画する場合、本人の化身（アバター）を表示するか否かという問題がある。第1次システムでは、あくまで本人が主役を演じるというコンセプトの基にアバターは表示しない方式をとった。しかしながら、この方式では、本人の服装などがサイバースペースの中の状況（過去のある時代であったり、未知の惑星であったりする場合）とマッチしないことが多いため、本人の参画意識が高まらない。これに対し、本人の姿形をサイバースペースにマッチした姿形（例えば歌舞伎役者など）に変形してアバターとして表示する「バーチャル歌舞伎システム」[9]などの研究が行なわれており、アバター利用の有効性が示されている。

（2）インタラクション

a）インタラクションの頻度

インタラクションが原則的にはストーリーの変化時点のみに限られていた。そのため、それ以外の部分ではストーリーは映画と同様あらかじめ決められたリニアな時間軸に沿って進んでいた。このため、参加者が「観客」になってしまい、インタラクションの必要な場面で参加者としてインタラクションに加わってくれにくくなるという欠点を持っている。さらに、参加者から見ても、インタラクションの頻度が少ないと自分が参加しているという意識が薄くなり映画との違いが明確でなくなるという点も欠点である。

b）インタラクションの種類

第1次システムでは音声認識技術、ジェスチャ認識技術をインタラクションの技術として利用した。しかしながら、システム実演時には照明を落とし暗くするため、細かなジェスチャ認識が出来ず、実質的にはほぼ音声認識だけで動くシステムとなっていた。そのため、参加者からすると利用出来るモダリティが限定されインタラクションが単純になるという問題点を持っていた。

3．第2次システムの概要

3．1　改良点

上記の点を考慮して第2次システムでは、以下の点を改良することとした。

（1）サイバースペースへの参加

a）複数人参加型システム

複数人参加型システムへの最初の取り組みとして、参加者2人がサイバースペースのストーリー展開に参加できるシステムとした。このシステムの将来の狙いは、ネットワークを介した複数人参加システムであるが、今回はその手始めとして、LANで接続された2システム間での複数人参加型システムを試作した。

b）アバターの表示

本人の分身であるアバターをスクリーン上に表示する方式を採用した。アバターを用いると、その表現形式を種々変えることにより参加者とアバターの関係、また、アバターと映画中の他のキャラクタとの関連性を種々コントロールすることが出来る。またアバターの動きに自律性を付与することによって、アバターの動きは自律性と参加者の動きの複合された複雑な動きをとることが可能となる。

（2）インタラクション

a）任意の時点でのインタラクション（anytime interaction）の実現

参加者とシステムのインタラクションの頻度を上げるため、任意の時点で本人とサイバースペースのキャラクタとのインタラクションが可能な仕組みを取り入れることとした。原則として、このインタラクションは即興的なインタラクションであって、ストーリー展開に影響を及ぼさない。このようなインタラクションをstory unconscious interaction（SUI）と呼ぶことにする。これに対し、ストーリーの分岐点におけるインタラクションであって、ここでの結果がそれ以降のストーリー展開を決定するものをstory conscious interaction（SCI）と呼ぶことにする。

b）マルチモーダルインタラクションの導入

音声認識を中心としたインタラクション機能に加え以下のインタラクション機能を付加することとした。

・感情認識機能：anytime interactionを実現するため感情認識機能を取り入れる。すなわち、任意の時点で、参加者がキャラクタに話しかけると、声に含まれる感情の認識結果に応じて、キャラクタが声やCGによって即興的に対応する。このメカニズムにより、参加者が常にストーリーと関わっているという感覚を与えることが出来る。

・モーションキャプチャ：参加者の動きをアバターの動きに反映させるため、磁気センサを体の適切な部分に装着するいわゆるモーションキャプチャ方式を採用した。これにより、本人がアバターの動作をコントロールしている感覚を生成できる。これもanytime interactionの実現に寄与している。

・ジェスチャ認識：照明の暗い条件下でのジェスチャ、3次元的なジェスチャ、および細かなジェスチャの認識を可能にするため、磁気センサによるモーションキャプチャを行ない、そのデータをHMMを用いて認識することによりジェスチャの認識を行なうこととした。

SCIは、音声認識、ジェスチャ認識の結果を用いる。これらの機能を取り入れた第2次システムの概観を図3に示す。

3．2　ソフトウェアシステム構成

第2次システムのソフトウェア構成を図4に示す。

（1）システム構成のコンセプト

第1次システムではストーリー展開に重点がおかれていたが、第2次システムではanytime interactionの考え方を取り入れることによってストーリー展開と即興的なインタラクションとをバランス良く制御する必要が生じる。そこでトップダウン的なシステム構成から分散制御的なシステム構成へ移行することとした。このような場合、全体としてのストーリー進行の制御が困難になる可能性があるが、anytime interactionの機能を優先させることとした。分散制御システムのアーキテクチャとして種々のものがあるが、ここではaction selection network [10] を採用した。これは複数個のエージェント間で活性値の送受が行なわれ、活性値がいき値を越えたエージェントが活性化し、そのエージェントに付随したプロセスが動作するというものである。

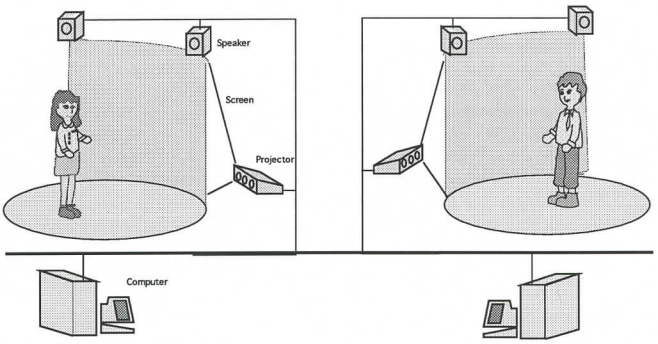

図3　第2次システムの概要図

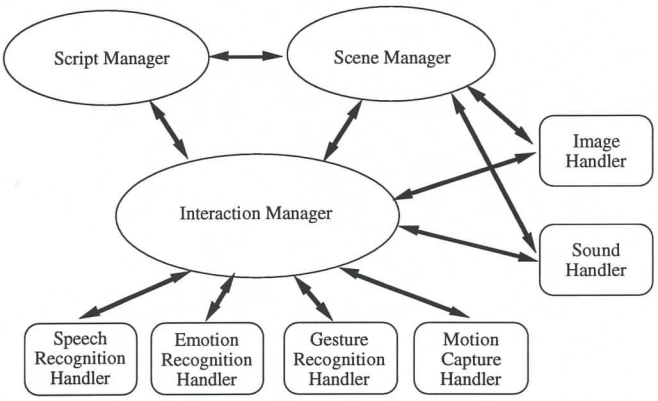

図4　第2次システムのソフトウェア構成

（2）スクリプトマネージャ

スクリプトマネージャの役割は第1次システムと同様であり、シーン間の遷移を制御する。インタラクティブストーリーは、種々のシーンの集合およびシーン間の状態遷移図で表現される（図5）。スクリプトマネージャは、この状態遷移図を記憶しておき、シーンマネージャから送られてくるインタラクション（SCI）の結果に応じて、シーン間の遷移をコントロールする。

（3）シーンマネージャ

シーンマネージャはシーンの記述およびシーン内のストーリー進行を管理する。シーン内のストーリー進行に関係ある出来事をeventと呼び、シーンマネージャはevent遷移の制御を行なう。あらかじめ各シーンの内容記述はevent networkとして蓄えてある。シーンマネージャは、スクリプトマネージャから指示されたシーンの記述データを参照し、各シーンの生成を行なう。シーン毎のeventは以下の要素から構成されている。

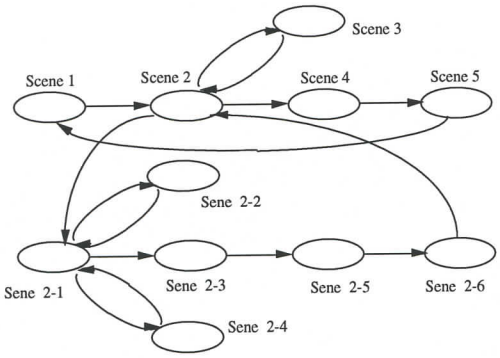

図５　シーン間遷移図

ａ）シーン映像
ｂ）背景音楽
ｂ）効果音
ｃ）キャラクタのアニメーションおよび台詞
ｄ）参加者とキャラクタのインタラクション（SUI）

第１次システムでは、シーンマネージャーは、これらを出力する時間管理をを行なっていた。しかしながら、第２次システムでは、anytime interaction の考え方を取り入れているため絶対的な時間の管理は出来ない。可能なのは相対的な時間の管理もしくは時間順序の管理である。そこでここでは action selection network の考え方を取り入れることとした。動作の概要は以下の通りである。

１）event には他の event および外部と活性値の送受を行なう。
２）活性値の累積値がいき値を越えると event が活性化する。
３）event の活性化に伴い event が発現する。また、他の event に活性値が送られると共に、その event の活性値はリセットされる。

活性値の送られる方向､強さなどを定めておくことにより、event の発生順序をあらかじめ定めておくことが出来ると同時に、event の発生順序にゆらぎやあいまいさを導入することも可能である。シーンを記述する action selectgion network の例を図６に示す。

（４）インタラクションマネージャ

anytime interaction の実現には、インタラクションマネージャーが最も重要な働きをする。anytime interaction を支えるベースとして、各キャラクタ（参加者のアバターもキャラクタの１つと考える）に感情状態を割り当て、参加者とのインタラクション、およびキャラクタ相互のインタラクションがキャラクタの感情状態を決定し、それに応じて各キャラクタの反応が決まるという構造を考える。

１）感情状態の定義

参加者 (i=1,2,...) の時刻 T における感情状態および、その強度を以下のように定義する。

Ep (i,T) , sp (i.T)　where sp (i, T) =0 or 1
（０は入力がない場合、１はある場合をさす。）

同様に、オブジェクト (i=1,2,...) の時刻 T における感情状態、および、その強度を以下のように定義する。

Eo (i,T) , so (i,T)

２）オブジェクトの感情状態の決定

簡単のため、オブジェクトの感情状態は参加者からの感情認識結果が得られた場合、その感情状態によって決定することとする。

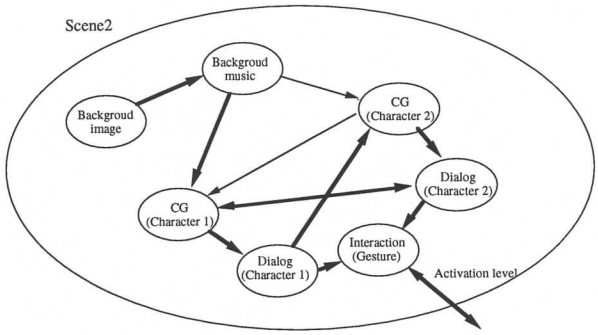

図６　シーンデータの例

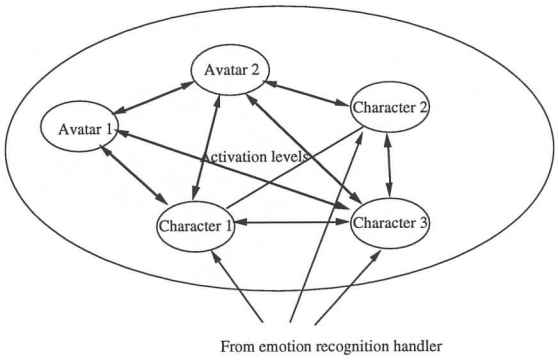

図７　インタラクションマネージャーの構成

$$\{Ep\ (i,T)\} \rightarrow \{Eo\ (j,T+1)\}$$

感情認識結果が入力されると、各オブジェクトに活性値が送られる。

$$sp\ (i, T) \rightarrow sp\ (i, j, T)$$

sp (i, j, T) は、参加者 i の感情認識結果に基づいてオブジェクト j に送られる活性値である。

オブジェクト j の活性値は送られてくる活性値の総和となる。

$$so\ (j, T+1) = \Sigma\ sp\ (i, j, T)$$

３）アクションの発現

活性値がいき値を越えたオブジェクトは、アクション Ao (i, T) をおこす。アクションは感情状態によって定まる。具体的にアクションとは、参加者の感情に応じた、キャラクタの動作、台詞によるリアクションをさす。同時に、他のオブジェクトに活性値 so (i, j, T) が送られる。

if so (i, T) ＞THi
then Eo (i,T) → Ao (i, T) , Eo (i, T) → so (i, j, T)
so (j, T+1) = Σ so (i, j, T)

この仕組みにより、オブジェクト同士の相互作用が発生し、感情認識結果とオブジェクトのリアクションが１対１に対応する単純なインタラクションに比較して多様なインタラクションが可能となる。action selection network で表現されたインタラクションマネージャーの構成を図７に示す。

３．３　ハードウェアシステム構成

図８に第２次システムのハードウェア構成を示す。映像出力サブシステム、音声・感情認識サブシステム、動作認識サブシステム、サウンド出力サブシステムより構成される。

（１）映像出力サブシステム

CG 生成用高速 WS ２台（Onyx Infinite Reality および Indigo2

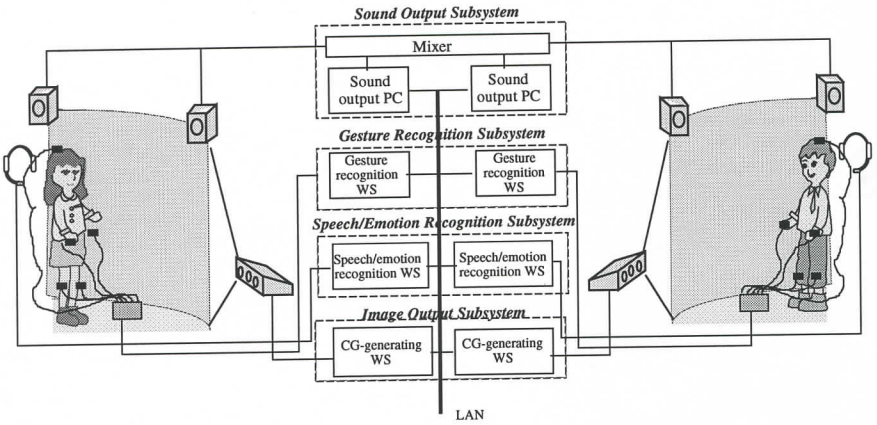

図８　第２次システムのハードウェア構成

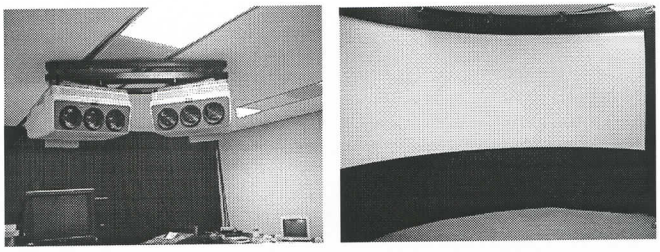

図９　プロジェクタとアーチスクリーン

IMPACT）を映像出力のためのWSとして用いている。Onyx上には、スクリプトマネージャ、シーンマネージャ、インタラクションマネージャ、映像出力ハンドラの各ソフトウェアがインプリメントされている。キャラクタの映像はCGアニメーションデータとしてあらかじめ蓄えておき、リアルタイムでCGが生成される。背景のCG映像もディジタルデータとして蓄えておき、リアルタイムで背景映像が生成される。背景の一部は実写映像を用いており、これは外付けのLDに蓄えておく。これら複数キャラクタのCG、背景のCG、背景の実写映像はOnyxおよびIndigo2のビデオボードでオーバーラップ処理が行なわれる。

臨場感に富んだ映像生成のために、CG映像は立体表示させる。また、参加者を映像で取り囲みインタラクティブ映画の世界に没入させるため、アーチスクリーンを採用した。あらかじめ、左眼用、右眼用の２種の映像データをWSで作成しておき、立体視コントローラを介してこれらを混合すると共に、２台のプロジェクタによりアーチ型のスクリーンに投影する（図９）。ただし、Indigo2側は処理速度等の問題のため立体視は採用しておらず、また映像出力も通常の大型ディスプレイで行なう。

（２）音声・感情認識サブシステム

音声および感情認識は２台のWS（SUN SS20×２）で実行される。これらのWS上には音声認識ハンドラ、感情認識ハンドラもインプリメントされている。マイクから入力された音声はSUN内蔵の音声ボードによりAD変換され、音声認識ソフト、感情認識ソフトにより音声認識、感情認識が実行される。２台のWSが２人の参加者それぞれの音声入力を処理する。

（３）動作認識サブシステム

動作認識は２台のSGI Indyで実行される。Indy上には動作認識ハンドラもインプリメントされている。それぞれのWSは、２人の参加者の体に装着された磁気センサからの出力を取り込み、アバター制御用のデータとして用いると共に、ジェスチャの認識も行なう。

（３）サウンド出力サブシステム

サウンド出力サブシステムは複数台のPCより構成されている。同時に出力する必要があるサウンドは、背景音楽、効果音、キャラクタ毎の台詞である。効果音、キャラクタの台詞はディジタルデータとして蓄えておき、必要に応じてDAを行なう。これらのサウンドの同時出力をサポートするため、複数チャンネルの同時DAが可能なように複数台のPCを用意してある。また、背景音楽はあらかじめ外付けのCDに記録しておき、その出力制御をPCから行なう。これら複数のチャンネルより出力されたサウンドはコンピュータ制御可能なミキサ（ヤマハO2R）でミキシングされ出力される。

３．４　インタラクション技術

３．４．１　ジェスチャ認識

ジェスチャ認識については種々の研究が行なわれている。それらは体全体の大きな動きの認識 [9] と手先の細かな動きの認識 [11] の２種類に大別される。しかしながら、日常のコミュニケーションで生じる身振り・手振りはこの中間的なものがが多い。また、このような動作により非言語的情報が表出・認識される。映画の世界に没入するためには非言語コミュニケーションが重要であり、そのためにもジェスチャの認識は必要である。以上のような考察の基にHMMを用いたジェスチャー認識を取り入れることとした。

（１）ジェスチャーのデータ取得

オクルージョンの問題や、ジェスチャーの持つ３次元空間的特徴からすると、カメラ入力を用いることはデータ取得段階でかなりの困難さを持つ。したがってここでは、磁気センサーを体の適切な位置に装着し、複数の磁気センサーからの入力情報をジェスチャーに関するデータとして用いた。本センサでは、３次元の位置座標およびセンサーの回転角度の情報が得られる。これらのうち、今回は３次元位置座標を用いることとした。

磁気センサiの時間tにおける３次元位置座標を (xi (t) , yi (t) , zi (t)) とする。N個の磁気センサ（1,2,...,N）を用いることとした場合、以下のデータを時間tにおける特徴パラメータとして用いた。

$$v(t) = (vx1(t), vy1(t), vz1(t), vx2(t), vy2(t), vz2(t), \ldots\ldots, vxN(t), vyN(t), vzN(t))$$

（ただし、$vxi(t) = xi(t) - xi(t-1)$　vy, vz も同様）

入力の特徴量はv (t) の時系列として与えられる。

（２）HMMによる認識と学習

認識処理ではHMMを用いて特徴ベクトルの生成確率を算出する。HMMの構成は図１０に示したようにleft-to-right型とする。

HMMは以下で定義できる。

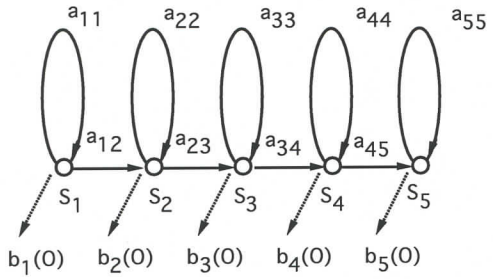

図１０ left-to-right HMMの構成

a）状態： s= {s1, s2, ..., sN}
b）出力シンボル：O1, O2, ..., OM
c）状態遷移確率：aij （状態 si から状態 sj に遷移する確率）
d）出力確率：bj (O) 状態 sj においてベクトル O を出力する確率

このとき、あらかじめ O の学習データが与えられたとき、最適な状態遷移確率、および出力確率を求めておく、認識時には、出力されたシンボルの系列が与えられたとき、これを出力する最適の状態遷移およびその場合の出力確率が得られる。したがって、カテゴリー毎に学習した HMM を用意しておくと、観測されたシンボル確率が与えられたとき、それを出力する最適の HMM が決定される。

（３）学習データ

あらかじめ、６種の感情（喜び、怒り、驚き、悲しみ、軽蔑、恐れ）に対応したジェスチャを定めておき、これを学習データとして用いる。学習データは５人の被験者から一人、各ジェスチャあたり１０サンプルずつ取得し、上記の学習アルゴリズムを用いて学習を行なった。

３．４．２ 感情認識

anytime interaction を実現するための鍵として感情認識を採用した。通常の音声認識では、入力できるのは基本的には認識対象の語彙に限定されるため、任意の時点で任意の音声入力を行なわせようとする anytime interaction の考え方には適合しない。そこで、不特定話者、内容不依存タイプの感情認識を用いることによりこの問題の解決を図った。認識アルゴリズムとしてはニューラルネットを採用することとし、大量の学習データを用いることにより、安定した性能を得ることを狙った。認識対象の感情は、「喜び、怒り、驚き、悲しみ、軽蔑、からかい、恐れ、普通」の８種類である。

図１１は処理の流れのブロック図である。音声特徴抽出部、感情認識部から構成されている。

（１）音声特徴抽出

感情認識のために、音韻特徴を表わすパラメータと韻律特徴を表わすパラメータを用いる。音韻特徴パラメータとしてはLPCパラメータを用いる。韻律特徴としては、エネルギー、音韻特徴の時間変化、およびピッチを用いる。入力音声ディジタル化された後、適当な長さのフレーム毎にLPC分析が行われ、以下の特徴パラメータが求められる。

Ft = (Pt, pt, dt, c1t, c2t, ..., c12t)

ただし、Pt、pt、dt、(c1t,c2t,..c12t) は、t フレームに関する音声パワー、ピッチ、時間変化パラメータ、LPC パラメータである。次に、このパラメータの時系列から音声パワーを用いて音声区間が抽出される。抽出された音声区間の全体から、等間隔になるように配置された１０フレームを取り出す。これらの１０フレームの特徴パラメータをまとめることにより、音声の特徴量は１５０次元（１５×１０）の特徴ベクトル、

FV = (F1, F2, ..., F10) ,

として表現される。ここで、Fi は、i フレームの特徴パラメータである。FV は、感情認識部への入力として用いられる。

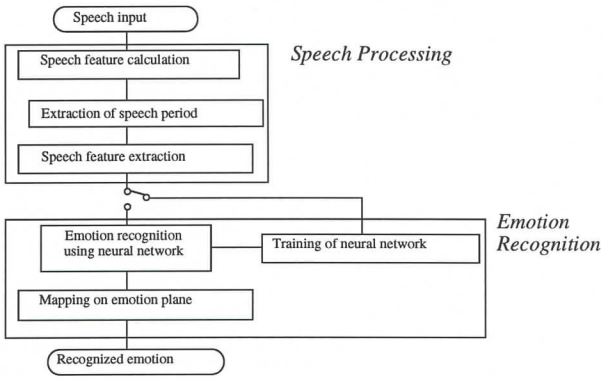

図１１ 感情認識のフローチャート

（２）感情認識

感情認識のためのニューラルネットの構造を図１２に示す。このネットワークは８つのサブネットワークの集合とそれらのサブネットワークの出力を統合する論理部から構成されている。８つの各々のサブネットワークは８つの感情（怒り、悲しみ、喜び、恐れ、驚き、愛想をつかす、からかい、および普通）のそれぞれにあらかじめチューンしてある。感情認識の困難さは、認識すべき感情によって大きく異なっているため、１個のニューラルネットを用意するより、各々の感情に対応したニューラルネットを用意しておいて、これらをそれぞれの感情にチューンした方がいいと考えられるため、このような構造を採用した。

（３）ニューラルネットの学習

感情認識を行うためには上に述べたニューラルネットをあらかじめ学習させておく必要がある。我々の目標は不特定話者、コンテキスト独立型の感情認識であるため、以下のような音声サンプルを学習データとして用意した。

単語：１００個の音韻バランスがとれた単語
話者：１００名（５０名の男性と５０名の女性）
感情：普通、怒り、悲しみ、喜び、恐れ、驚き、愛想をつかす、からかい
音声サンプル１：各々の話者が８つの感情で１００個の単語を発声する。
音声サンプル２：各々の話者が各感情毎に各母音を５回づつ発声

この学習データを用いて予備実験を行った。その結果、男性、女性をまとめたニューラルネットを用意するより、男性、女性それぞれにチューンしたネットワークを用意する方が学習、認識共に有利であることがわかった。

（４）ニューラルネットによる感情認識

感情認識の際には、音声特徴抽出部で得られた音声特徴量が、上に述べた方法で学習が行われた８つのサブネットワークに入力されれる。その結果として、８つの出力が得られる。

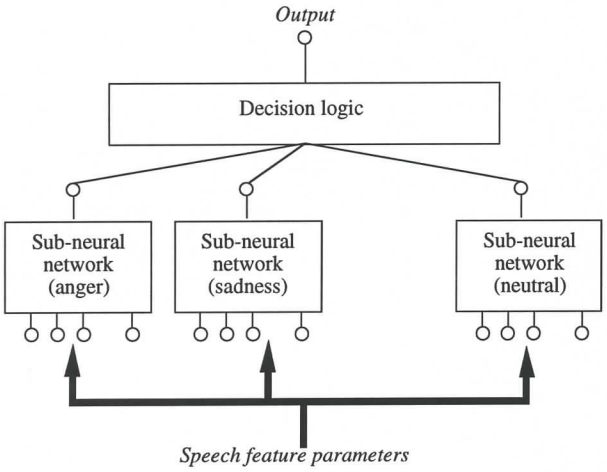

図１２　ニューラルネットによる感情認識

最大の出力を与える感情が認識結果である。

４．インタラクティブストーリーの構成例

４．１　インタラクティブストーリー

以上述べたシステムの上に具体的なインタラクティブストーリーを構築した。ストーリーのベースとしてジェークスピアの「ロミオとジュリエット」を採用した。これは以下の理由による。

（１）主人公が二人（ロミオ、ジュリエット）いるため、複数人参加型のシステムに適したストーリーである。

（２）誰でも良く知っているストーリーであり、ロミオもしくはジュリエットになりたいと感じる人は多いため、参加者が容易にインタラクティブストーリーの中に没入し、相手や、ストーリー中のキャラクタとのインタラクションを行なってくれる。

核となるプロットは以下の通りである－ロミオとジュリエットは、彼等の悲劇的な死の後「黄泉の国」に魂と記憶を失った状態で送られる。黄泉の国のキャラクタのガイドやアドバイスに助けられながら、彼等は徐々に自分自身を再び見い出し、最後に魂を再び手に入れて現世へと帰っていく。

４．２　インタラクション

本システムは二人の人が同時に参加できる。上のストーリーの例では一人がロミオをもう一人がジュリエットの役割を演じる。図３に示した２台のシステムは別の場所に設置されており、LANを介して接続されている。(同じ場所に設置することも可能である。この場合には、参加者＝パフォーマーによって演じられるストーリー進行の全体を観客が楽しむというパフォーマンスとして上演することも可能である。)各々の参加者はスクリーンの前に磁気センサーとマイクを装着した状態で立つ。ロミオ役は3Dの液晶シャッター眼鏡をつけ3Dの映像を楽しむことが出来る。基本的にはストーリー進行はシステムが制御するが、先に述べたようにanytime interactionの機能を採用しており、参加者は自由にキャラクタとインタラクションを行なうことが出来る。参加者のインタラクションの頻度によって本システムはストーリーの進行を楽しむストーリーベースのシステムにも、即興的なインタラクションを楽しむインタラクションシステムにもなる特徴を持っている。

５．まとめ

映画とインタラクション機能を統合することによって実現されるインタラクティブ映画は通信、映画、テレビゲームなどを統合した新しいメディアになる可能性を持っている。本論文ではまずインタラクティブ映画のコンセプトとそれに基づいて我々が開発した第１次システムの内容について簡単に説明した。また、第１次システムを試用することによってわかった問題点を指摘した。我々はこれらの問題点を解決した第２次システムを構築中である。本システムでは二人の人がネットワークを介して、インタラクティブストーリーの進行に参加できる。また、いつでもストーリー中のキャラクタとインタラクション出来る機能を持たせている。システムの評価については別の機会に報告する

文　　献

[1] C.W.ツェーラム（月尾嘉男訳）、”映画の考古学、”フィルムアート社 (1977) .

[2] 中津良平、土佐尚子、”インタラクティブ映画構築に向けて －Inter Communication Theaterの基本構成と概念例－、”信学技報 IE96-113 (1997) .

[3] Haruo Noma, et al., "Multi-Point Virtual Space Teleconference System," IEICE Trans. Commun., Vol. E78-B, No. 7 (July 1996) .

[4] Pattie Maes et al., "The ALIVE system : Full-body Interaction with Autonomous Agents," Proceedings of the Computer Animation '95 Conference (1995) .

[5] Naoko Tosa and Ryohei Nakatsu, "Life-like Communication Agent - Emotion Sensing Character 'MIC' and Feeling Session Character 'MUSE' -," Proceedings of the International Conference on Multi-media Computing and Systems, pp.12-19 (June 1996) .

[6] Machiko Kusahara, Christa Sommerer, and Laurent Mignonneau, " Art as Living System," システム／制御／情報、 Vol.40, No.8, pp.344-351 (Aug. 1996) .

[7] Tetsu Shimizu, et al., "Spontaneous Dialogue Speech Recognition Using Cross-Word Context Constrained Word Graph," Proceedings of ICASSP'96, Vol. 1, pp. 145-148 (April 1996) .

[8] Christopher Richard Wren, et al., "Pfinder : Real-Time Tracking of the Human Body," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, pp.780-785 (July 1997) .

[9] Jun Ohya, et at., "Virtual Kabuki Theater : Towards the Realization of Human Metamorphosis Systems," Proceedings of 5th IEEE Workshop on ROBOT AND HUMAN COMMUNICATION (RO-MAN'96) , pp.416-421 (Nov. 1996) .

[10] Pattie Maes, "How to do the right thing," Connection Science, Vol.1, No.3, pp.291-323, 1989.

[11] Vladimir I. Pavlovic, et al., "Visual Interpretation of Hand Gestures for Human-Computer Interaction : A Review," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7. pp.677-695 (July 1997) .