インタラクティブ映画第２次システムの構成

中津　良平　　土佐　尚子　　越知　武

ＡＴＲ知能映像通信研究所
〒619-0288　京都府相楽郡精華町光台２丁目２番地
Tel：(0774) 95 1400/1427/1494Fax：(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、その問題を指摘する。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。

Construction of Interactive Movie System II

Ryohei Nakatsu, Naoko Tosa, and Takeshi Ochi

ATR Media Integration & Communications Research Laboratories
2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288 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 me
that integrates various media, including telecommunications, movies, and video games. In this paper, we first explain
concept of interactive movies and briefly explain the prototype system we developed. We point out several problems in
system that need to be improved. We explain the two significant improvements introduced in the second system: the introdu
of interaction at any time and two-person participation through the use of a network. We describe the software and hard
configuration of the second system and introduce one example of an interactive story installed in this system.

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。また最近の映画はディジタル技術やコンピュータグラフィクス技術を取り入れ、新しい世代の映画へと移行しつつある。ディジタル技術やコンピュータグラフィクス技術は、従来の映画にはなかった超現実的な世界、すなわちサイバースペースを作り出す能力を与えてくれる[2][3]。しかしながら、従来の映画は、仮想の世界（サイバースペース）とそこにおけるストーリー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、サイバースペースに入り、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[4][5]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次システムの概要および問題点と、それを改良した第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

サイバースペースの中でのストーリー体験という考え方は、

テレビゲームやテーマパーク等におけるバーチャルシアターなどを含めて考えると、これまでも存在した考え方であるが、その在り方について体系的な考察が加えられた例は少なく、従来の小説などの物語展開(storytelling)を越える新しい物語が生まれる可能性があるという抽象的なレベルのものにとどまっており[6]、具体的なシステムの提示には至っていない。

我々は、誕生以来１００年を経て種々の物語展開技法が生み出され、メディアとしての位置を確立している映画[1]をその出発点として、サイバースペースの中でのストーリー体験が可能なシステムを具体的に構築することをめざしている。また、この新しいメディアをインタラクティブ映画と呼ぶことにする。インタラクティブ映画は、その運用まで含めて考えると、インタラクティブ映画システムとそれを用いて主人公となってストーリーを体験する参加者から構成される。インタラクティブ映画システムを構成するのは以下の要素である。

(１)参加者とのインタラクションに基づいて途中経過・結末などが変化するインタラクティブなストーリーを生成する機能。具体的には、ストーリーはシーンの連続として構成される。

(２)スクリーンまたはディスプレイに投影・表示される映像と、映像に関連する音声・サウンドで構成されるシーンを生成する機能。

(３)ストーリー中の主人公以外の役割を演じる単数もしくは複数のキャラクタおよびその動作を生成・制御する機能。

(４)参加者とキャラクタもしくはシーン中のオブジェクトとのインタラクションを可能にするインタラクション機能。

文献[6]では、サイバースペースの中でのストーリー体験システムの特徴をimmersion(没入感)とinteractivityであると述べているが、インタラクティブ映画ではより具体的に以下のような点を特徴とする。

(１)サイバースペースの構築と参加者のサイバースペースへの没入

CG の活用、CG と実写映像の統合、さらには立体映像の創出により、臨場感に富んだサイバースペースを構築し、参加者にあたかもその中にいるかのような感覚を与える。

(２)インタラクションを介したサイバースペースの中でのストーリー体験

サイバースペースの住人、すなわち参加者を取り巻く種々のキャラクタと音声や、身振り・手振りでインタラクションを行なうことによりストーリー展開を体験することができる。

２．２　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[4][5]。システムの内容を簡単に説明する。

(１) ソフトウェア構成

ソフトウェア構成を図１に示す。スクリプトマネージャは、インタラクティブストーリーにおけるシーン間の遷移をコントロールする。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。インタラクションマネージャは、スクリプトマネージャ、およびシーンマネージャの下にあって、各シーンにおけるインタラクションを制御する。ハンドラは各種の入出力を制御する機能を持つ。音声認識機能を制御する音声認識ハンドラ、動作の認識を制御する動作認識ハンドラ、映像の出力を制御する映像ハンドラ、サウンドの出力を制御するサウンドハンドラがある。

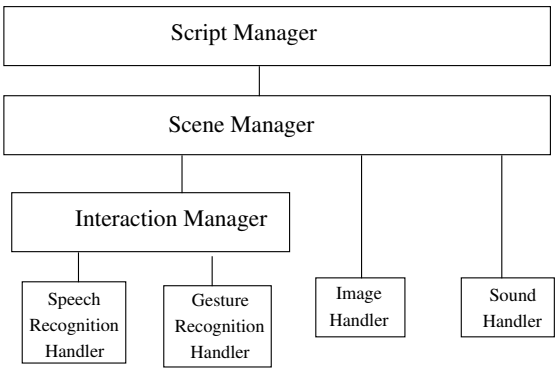

図１　第１次システムのソフトウェア構成

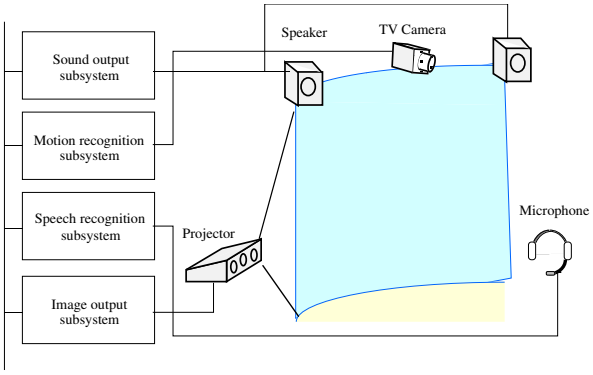

図２　第１次システムのハードウェア構成

(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 WS 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。

２．３　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者数

第１次システムでは、１人の参加者が主人公となってストーリーを体験するという形態をとっていた。しかしながら、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。

(２) インタラクションの頻度

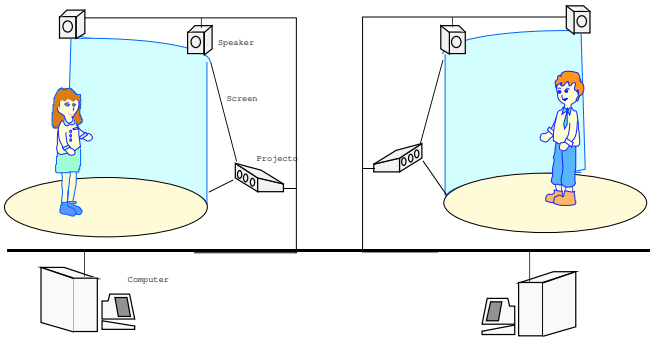

図３　第２次システムの概要図

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）複数人参加型システム

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。

（２）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。

（３）その他

参加者のアバターの表示、照明が暗い条件下でのジェスチャ認識等を可能にするため、磁気センサを用いて参加者の動きを取り込む方式を採用した。

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

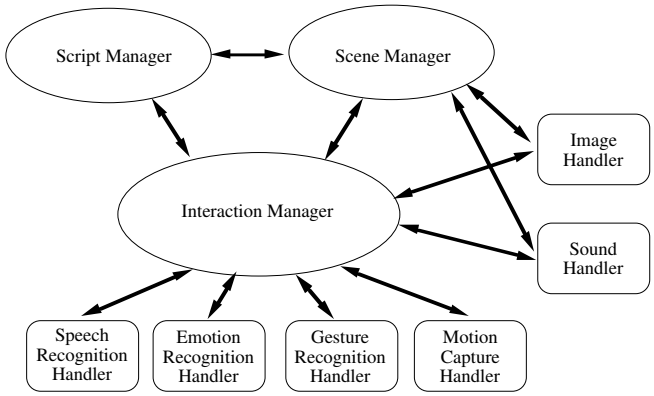

図４　第２次システムのソフトウェア構成

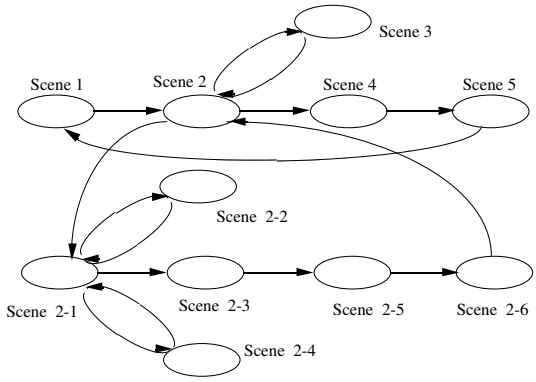

図５　シーン間遷移図

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。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 [2]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。

（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、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）

から構成されている。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に伴う action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。

（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は、感情認識とaction selection networkを組み合わせることにより実現する。具体的には、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、キャラクタ(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）キャラクタの感情状態の決定

簡単のため、キャラクタ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。

{Ep(i,T)} →{Eo(j,T+1)}

感情認識結果が入力されると、各キャラクタに活性値が送られる。

sp(i, T)→ 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キャラクタ j に送られる活性値である。

キャラクタ 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)= Σ sp(i, j, T)

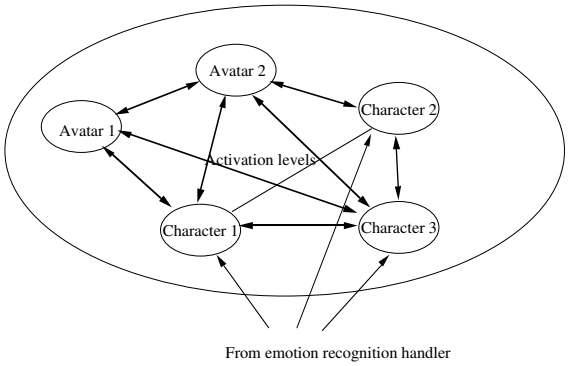

図 6　インタラクションマネージャーの構成

３）アクションの発現

活性値がいき値を越えたキャラクタは、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、他のキャラクタに活性値 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、キャラクタ同士の相互作用が発生し、感情認識結果とキャラクタのリアクションが 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 6 に示す。

３．３　ハードウェアシステム構成

図 7 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 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

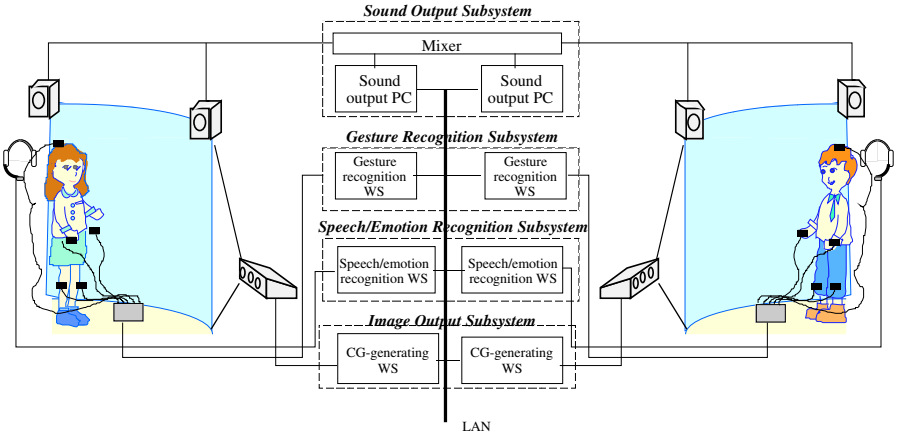

図７　第２次システムのハードウェア構成

図８　プロジェクタとアーチスクリーン

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８）。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。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]。また、感情認識はニューラルネットを用いて行なわれる[9]。

（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。ジェスチャ認識は磁気センサの出力である位置データの時間変化を特徴量とし、HMMを用いて行なっている。

（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。

４．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のでその概要を説明する。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９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。

４．１　音声特徴抽出

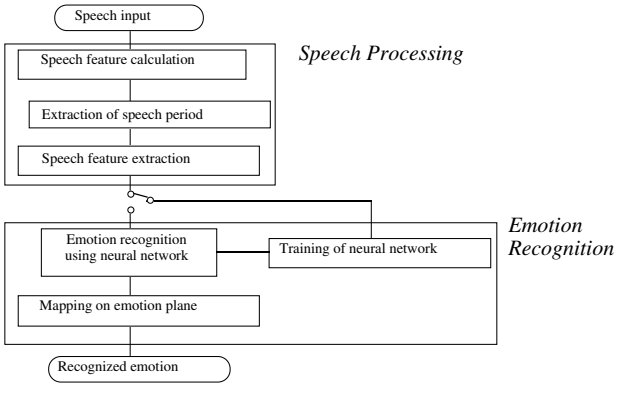

図９　感情認識のフローチャート

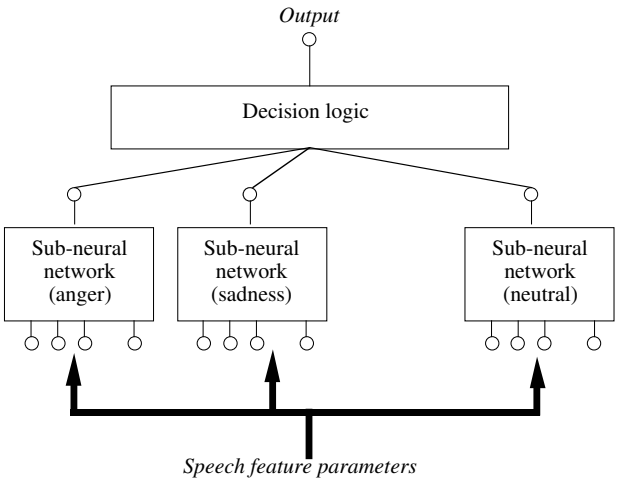

図１０　ニューラルネットによる感情認識

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。

４．２　感情認識処理

感情認識のためのニューラルネットの構造を図１０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。これらのサブネットワークは３層構造であり、入力層は音声特徴量の次元に対応した１５０個のノードよりなり、中間層は２０ないし３０のノードからなり、出力層は１個のノードから構成されている。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。

（１）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。

（２）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 つの実数値が得られる。これを、

V=(v1, v2, ..., v8)

と表現することとする。最大の出力を与える感情が認識結果である。

５．インタラクティブストーリーの構成例

５．１　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シェークスピアの「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、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。

５．２　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

図１１　「ロミオ」を演じる参加者

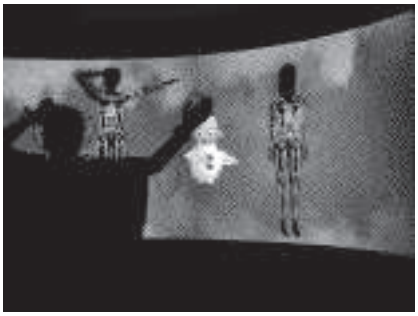

図１２　「ロミオ」がアバターの動作を制御している様子

図１３　「ロミオ」が「ジュリエットのアバターに挨拶している様子

図１４　「ロミオ」が店の品物を探している様子

図１５　「ロミオ」と恋敵「テオ」との決闘の場面

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。

図１１～１５にシステムと参加者がいくつかのシーンでインタラクションを行なっている様子を示す。

６．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新しいメディアになる可能性を持っている。本論文では、インタラクティブ映画のコンセプト、それに基づいて我々が開発した第 1 次システムの内容とその問題点について簡単に述べると共に、第 1 次システムの問題点を解決した第 2 次システムについて述べた。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。

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] Oren Jacob, "Computer Graphics Story - A Personal Overview of Computer Animation in the Movies," ACM Computer Graphics,

Vol. 31, No. 1, pp. 26-28, Feb. 1997.
[3] Graham Walters et al., "The Making of Toy Story," Course Notes of SIGGRAPH 96, Aug. 1996.
[4] 中津良平、土佐尚子、" Inter Communication Theater インタラクティブ映画の構築の試み− 、" インタラクション 97論文集、pp.67-68 (Feb. 1997).
[5] 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の基本概念と構成例− 、" 信学技報 IE96-113 (Feb. 1997).
[6] Janet H. Murray, "Hamlet on the Holodeck: The Future of Narrative in Cyberspace," Simon & Schuster, New York, 1997.
[7] Pattie Maes, "How to do the right thing," Connection Science, Vol.1, No.3, pp.291-323 (1989).
[8] 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).
[9] 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).