

別刷

日本バーチャルリアリティ学会 論文集

日本バーチャルリアリティ学会

感情に反応する自律型バーチャルアクターと仮想世界の生成

土佐 尚子^{*1} 中津 良平^{*1}

Creating an Imaginary World with Autonomous Virtual Actors that Responds to Emotions

Naoko Tosa^{*1} Ryohei Nakatsu^{*1}

Abstract -- Here, virtual reality is discussed from the viewpoint of a media art. The works of an artist are a reflection of a given period and the consciousness of its people; virtual reality is considered to be one form of high technology that allows such expression. We seek to build a system where anyone can actively create an image they envision, allowing a sense of realism to be felt. Therefore, we need a human-like character that has a psychological make up, expresses emotions, is intelligent, and possesses its own personality. Neuro-baby, MIC and session character MUSE possess such realistic traits and communicate through a virtual reality system. We are currently researching the next step: Interactive Theater, a story that changes to correspond with emotions.

Keywords; media art, A-life art, life-like computer character, interactive cinema, communication

1. まえがき

本論文では、メディアアートの立場から、バーチャルリアリティについて述べる。メディアアートとは、メディアの特性を表現に取り込む芸術を総称して言う。筆者らは、フィルム、ビデオ、そして近年からインタラクティブとその仮想環境を構築するためにコンピュータを使用して作品を制作している。アーティストは、その時代性と人間的な意識を反映する作品を提示するので、表現を反映するものとして、バーチャルリアリティのような先端技術を用いることは、充分考えられることである。

2. 表現内容におけるアート&テクノロジー研究

筆者らが映像に求めた人工現実感とは、映像に触れるという行為と、自分が創造するイメージが頭脳を持ち能動的に自己生成し、自己判断でき対応するシステムである。コンピュータベースのインタラクティブ映像はその可能性を持っている。しかし、従来のパターン化した装置の様なインタラクティブ映像ではなく、インタラクションの機能を用いてリアルな対応のできる生物と、そのコミュニケーションを

実現したいと考えた。そして私達にとって身近で一番コミュニケーションをする生物“人間”を選び、心理状態や、感情表現、性格付けや知能、行為とコミュニケーションをテクノロジーを用いてどこまでリアルに個性を持って表現できるのかという興味のもとに研究を進めてきた。インタラクティブキャラクター、エージェントの研究は、最近多く行われているが[1][2][3]、本研究はアーティストと技術者の協力により、アーティストの感性で行われている点が特徴である。以下に具体的に紹介していく。

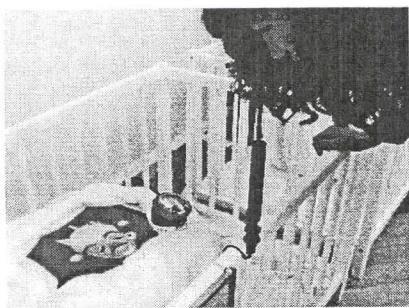
3. ニューロベイビー[4]

3.1 コンセプト

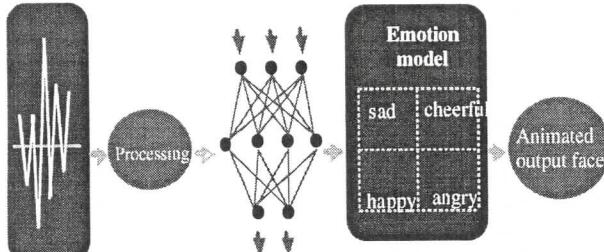
この忙しい情報化社会の中で、人間ほど気をつかわなくて良くて、ペットほど面倒を見なくてよい、ある程度自分自身で判断できる何かを、ふと求めたことはないだろうか。また人間が、老若男女問わず「人型」の物を好むのはなぜだろうか。昔から、土偶からロボットにいたるまで、人は自分と同じ形を作り、愛着を覚え感情移入している。ニューロベイビーは、自己の分身のようであり自分に最も近い他者がコンセプトである。ニューロベイビーは人間が声をかけても、その意味はわからない。しかし声の抑揚から、感情を判断し答えることができる。

*1:(株) ATR 知能映像通信研究所

*1:ATR Media Integration & Communications Research Laboratories

図 1 ニューロベイビー
Fig.1 Neuro-Baby

Voice Data

図 2 . 处理プロセス
Fig.2 Processing Flow

3.2 システムデザイン

システム設計は、富士通研究所との共同研究で行った。リアルタイム3次元CGアニメーションで作られたニューロベイビーは、ニューラルネットワークの学習アルゴリズムにより、声の抑揚、高低、強弱を認識し、泣いたり笑ったりという感情表現を行う。図1は、オーストリアのリンツで行われたアート&テクノロジーの国際会議「アルスエレクトロニカ」での展示風景である。その処理プロセスを図2に示す。ソフトウェアデザインは、次の5ステップからなる。

1. 感情タイプの設定と配置
2. 学習用データの収集
3. 音声と感情タイプの関係の学習
4. 表情の形状モデリング
5. 表情と感情タイプの対応付け（図3参照）

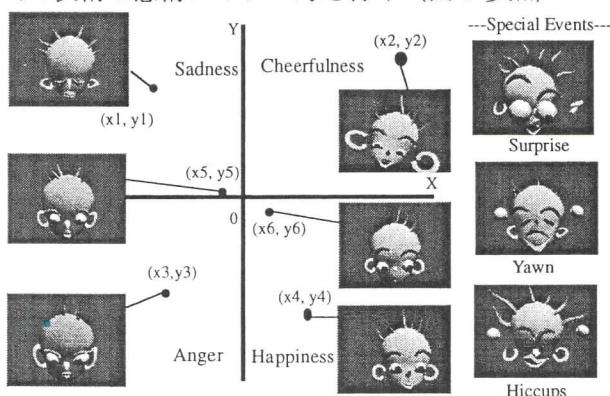


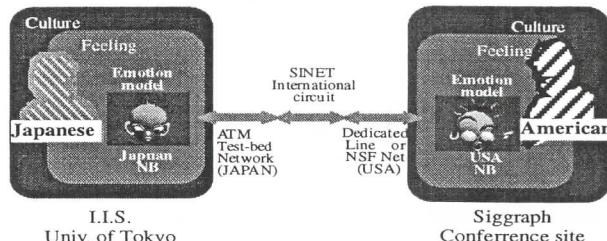
図 3 . 顔の表情と感情タイプの対応

Fig.3 Neuro-Baby's emotional expression

3.3 ネットワークニューロベイビー[5]

(1) コンセプト：ネットワークを用いて日米間を接続した両端にニューロベイビーを置き、外国人との非言語対話のインターフェース（感情を翻訳するエージェント）として用いる。国や民族ごとの感情タイプを正しく設定することにより、図4に示すように、それぞれの国の人々の感情表現が相手の国の感情に翻訳されて伝えられることにより、文化を越えた感情コミュニケーションができる。これは、文化の異なる人々の間の相互理解を深める新しいコードになる可能性を秘めている。

Network Based Neuro-Baby

図 4 . ネットワーク型ニューロベイビー
Fig.4 Networked Neuro-Baby

(2) ネットワーク型ニューロベイビーシステム：SIGGRAPH'95の会場と、共同研究先の東京大学生産技術研究所を結んで公開実験を行った。双方の会場にニューロベイビーのシステムを設置し、通信回線にてリアルタイムに感情のやりとりを行う。図5に処理プロセスを示す。本システムでは、新たにハンドシェイクマシン（握手を感知するセンサーデバイス）で、握り方の強さをデジタルデータ化して、感情の分析に用いる（図6）。又、アイトラッキングの装置は、ニューロベイビーの視覚となり、相手のいる方向に向かって顔を向けるようになる。図7にネットワークを示す。実験では、日本の利用者の感情データが米国のニューロベイビーに送られ、逆に米国の利用者のデータが日本のニューロベイビーに送られた。それぞれのニューロベイビーは、利用者の国民性に対応した反応（表情と握手）を相手に伝えた。

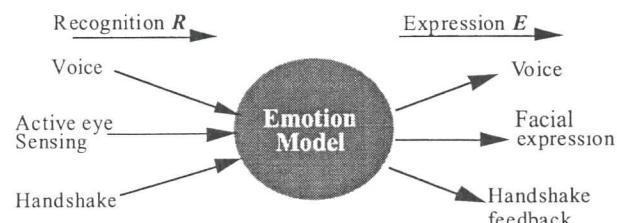


図 5 . ネットワークニューロベイビーの処理

Fig.5 Processing model of the Networked Neuro-Baby

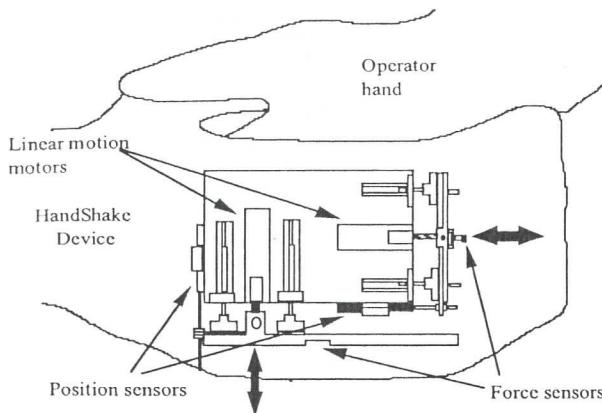


図6. 握手マシンの設計

Fig.6 Hand Shake Device representing the palm of the people

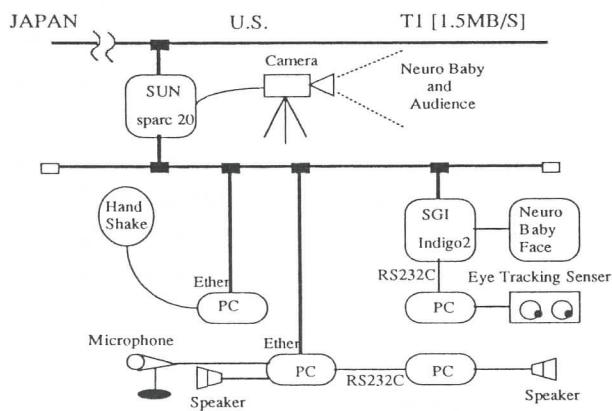


図7. ネットワークシステム

Fig.7 System configuration for Networked Neuro-Baby with robotic hand

4. マルチモーダルインタラクション

キャラクター：ミック[6]

4.1 コンセプト

筆者らが、ニューロベイビーで実現しようとしたものは、「感情によるコミュニケーション」である。感情によるコミュニケーションは、「非言語コミュニケーション」[7]という広い文脈の中でとらえることができる。筆者らは、ニューロベイビーをベースに「非言語コミュニケーション」による人格を持つコンピュータキャラクター「ミック」の研究を開始した。ミックはニューロベイビーの進化形であり、以下の点でニューロベイビーから進歩している。

(1) 非言語コミュニケーションの能力の増大：ミックはニューロベイビーと同様に感情に反応するキャラ

クタであるが、認識出来る感情を4つから8つ（喜び、怒り、驚き、悲しみ、愛想をつかす、からかう、怖れる、普通）に増やした。

(2) キャラクタの反応パターンの高度化：全身像のCGを作り、単に顔の表情による反応だけではなく、全身でのアクションにより反応である感情表現をすることにした。

(3) 感情認識機能の高度化：8つの感情の各々に対応したニューラルネットを用意し、これらを並列に設置することにより、感情を認識する方法を採用し、認識性能の向上を図ると共に、音韻バランスのとれた多数単語を多数話者が種々の感情で発声した音声データを訓練データとして用いることにより、不特定話者、内容独立型の感情認識を可能とした。

(4) 音声によるコミュニケーションの機能に加え、画像認識機能を取り入れることにより、人間の簡単なジェスチャーにミックが反応する機能を追加した。

4.2 感情

ミックは、8つの感情を声の抑揚から認識し、それに反応できる。例えば、だれも話しかけないと居眠りをし、だれかが話しかけると、機嫌が良いときは、「こんにちは」、悪い時は、「バイバイ」と返事をする。低い声でばかにすると怒り、からかうと逆立ちをする。口笛を吹いてあげると、エキサイトしてジャンプをし、人間が不機嫌な顔をして咳払いをすると、悲しくなって手で顔を覆い後ろをむいてしまう。たまに愛想をつかし人間に愚痴をこぼす。ミックの感情モデルは、図8の2次元座標に表わされる。X軸は快、不快、Y軸は感情の強弱である。具体的な8つの感情および、それに対応した入力音声（＊で示した）は以下の通りである。

- a. Joy (幸福, 満足, 楽しい, 心地よい, 笑顔)
 - * 高く強い声, 口笛など。
- b. Anger (立腹, 憤り, 不機嫌)
 - * 低く強い声で叱る。
- c. Surprise (驚く, ショック, びっくりする)
 - * 突然、大きな声で叫ぶ。
- d. Sadness (哀しい, 泣く, 寂しい)
 - * 低く弱い声, 咳払いなど。
- e. Disgust----* 愛想をつかした疲れた声。
- f. Teasing----* 高く軽い声でからかう。
- g. Fear-----* 低く低い声で、怖がらせるような声。

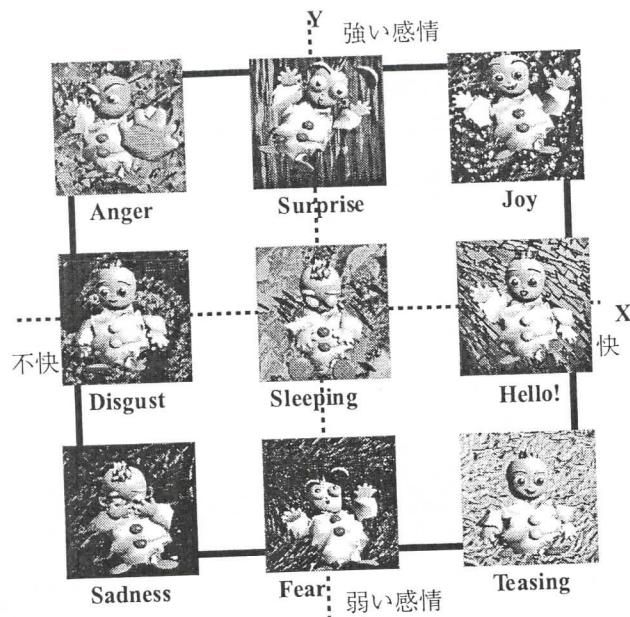


図8. ミックの代表的な感情表現

Fig.8 MIC's emotional expression

4.3 処理プロセス

図9は処理の流れのブロック図である。全体の処理は3つの部分、すなわち、音声特徴抽出部、感情認識部、および反応生成部から構成されている。音声特徴抽出部では、まず入力音声から音声特徴パラメータがリアルタイムで抽出される。次に、音声パワーを用いて音声区間の抽出が行われる。抽出された音声区間を用いて、入力音声のそれぞれに応じた音声特徴量が決定される。この音声特徴量は、感情認識部への入力として用いられる。感情認識部では、ニューラルネットを用いて感情認識が行われる。ニューラルネットの出力を論理処理して、2次元の感情平面への写像を行う。2次元の感情平面上には8つの感情をあらかじめ適切に配置しておく。2次元の感情平面上の認識感情の位置およびその動きに応じて、ミックの反応、すなわち顔の表情と体の動きが生成される。これらの顔の表情および体の動きは著者の一人であるアーティストの直感と感性と経験によって注意深く事前に定めておく。反応パターンはコンピュータグラフィックスで表示されると共に適切な音声もしくは音楽が出力される。

4.4 音声に含まれる感情の認識

4.4.1 感情認識の意味

音声には、言語的情報（意味内容）、感情情報、個人性情報（話し手が誰か）の3種類の情報が含まれている。通常のコミュニケーションでは意味内容が主役であると考えられているが、感情情報は意識ないし無意識のレベルで重要な役割を果たしている。特に赤ん坊の場合は、音声の意味内容を認識できる以前に感情の認識が出来るようになることが心理学的に確かめられている。コンピュータで音声を処理する場合、これまでビジネスなどへの応用を目的として意味内容を自動的に認識することに努力が注がれてきた。しかしながら、意味内容の認識のみを追及すると、コンピュータと人間の対話が感情を排した味気ないものになると共に、感情情報、個人性情報を無視しているため、意味内容の認識精度そのものにも限界があるという問題点を持つ。したがって、人間とコンピュータの自然な対話を実現するためには、音声に含まれる感情を認識することは避けて通れない課題である。従来感情音声の分析研究は行われているが[8][9]、認識を試みた例は少ない。

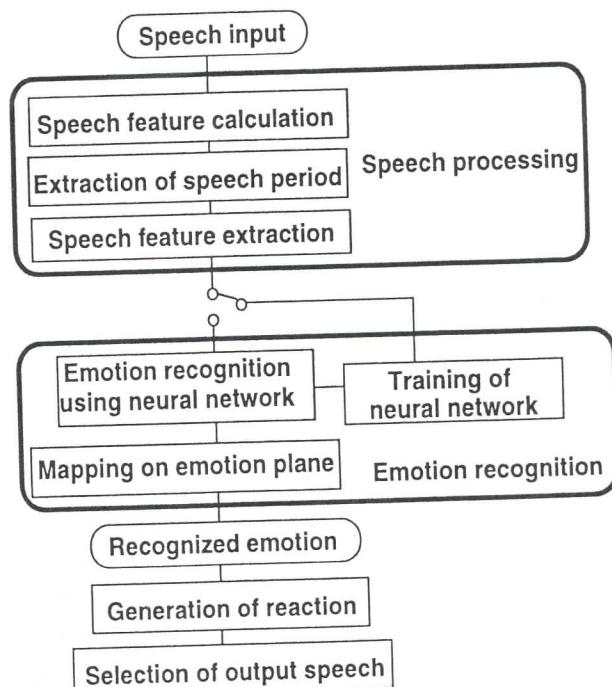


図9. ミックの処理プロセス

Fig. 9 Blockdiagram of the processing flow

象徴である。ミューズは、人間と音楽のメロディとリズムでコミュニケーションをとる。ミックは言葉で記号的な感情表現のコミュニケーションを行うが、ミューズは感情表現する言葉に満たない、もつと微妙な気分のコミュニケーションができる。彼女はピアノで我々に話しかけるので、下記の楽典に基づいて彼女と即興的なフィーリングセッションを楽しむことができる。

- a. Joy --- 上行のグリッサンドを演奏する
- b. Anger--- 低音域で同じ音を数回演奏する
- c. Surprise--- 強く複数音で数回演奏する
- d. Sadness --- 下行のグリッサンドを演奏する
- e. Disgust--- 高音域で不協和音を演奏する
- f. Teasing--- 高音域でトリルを演奏する
- g. Fear--- 低音域でトリルを演奏する

5.2 処理プロセス

具体的な処理アルゴリズムはミックと同一であるが、入力を音声ではなくシロホンの音にしているため、ニューラルネットの学習は、以下のデータを用いて行った。

演奏者：男女各10名

演奏回数：上記の各感情に対応して、演奏を各奏者が10回ずつシロホンをたたく

また、入力に対する音によるミューズのリアクションとして、感情マップに適したフレーズを作成した。これらのフレーズのリズムとメロディは、ジャズ専門の作曲家が、楽典に基づいて注意深く決めている。図13にミューズの感情表現を示す。

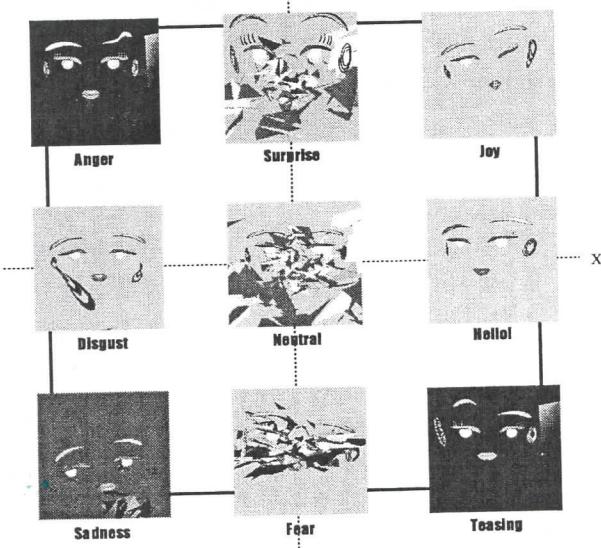


図13. ミューズのフィーリング表現例

Fig.13 MUSE's feeling expression

6. インタラクティブシアター [9]

6.1 コンセプト

ミック、ミューズによって、音声・動作で表現される感情、音楽等で人間と非言語のコミュニケーションのできるコンピュータキャラクターを実現することができた。しかしながら、ミック、ミューズと人間のインタラクションは、不自然な面を持つ。なぜならば、人間の感情は、普通の状態では、頻繁に変化するものではない。結局、人間が各々の感情に対応する音声・動作を意識的に作り出して反応を試していることで終始してしまう刹那的インタラクションになる。感情が自然に変化するには、その原因となる状況に没入すべき設定が必要である。その要因を設定するために、ミックやミューズのコミュニケーションには、ストーリーの導入が必要である。これを別の視点から考えてみると、ストーリーの導入された状況における感情、感性的インタラクションは、演劇、映画、小説、テレビゲーム、通信などを統合した観客参加、体験型の次世代映像メディアの可能性を秘めている。このメディアは、超現実的な映像やシーンからなるサイバースペースを作りだし、その中に我々自らが登場人物として入っていくことを可能にする。また、その中で他の人やCGキャラクターと感情を伴うコミュニケーションすることにより、情緒豊かに仮想世界を体験することが可能になる。他で行なわれているインタラクティブ・シアター研究[11][12]との最大の相違は、音声やジェスチャーによる自然なコミュニケーションおよび感情によるコミュニケーションが主観的没入感の達成の鍵であると考え、これらを実現している点にある。今回、このようなメディアの実現に向けた第1次のプロトタイプのシステム設計を行った。

6.2 システム構成

システムの構成を、図14に示す。

(1) スクリプトマネージャー：紙上で作られたインタラクティブな脚本をコンピュータ上に移植し、インタラクティブストーリーの展開全体をコントロールする。各シーン間の遷移の定義、インタラクション結果によって各シーンを制御する機能を持つ。スクリプトマネージャーは基本的には状態遷移図（各状態に各シーンが対応）で全体の動作を管理できる構造になっている。したがって、状態遷移図の作り方によってストーリー展開を自由に複雑化することが出来る。

(2) シーンマネージャー：シーンごとの登場人物の台詞と立体視CGアニメーション、効果音、BGM、ミキサー、インタラクション、実写映像とCGとの合成をコントロールする。

(3) インタラクションマネージャー：声とジェスチャーを用いて仮想世界とコミュニケーションするため、音声認識、感情認識、人間の動作を読み取る画像認識を採用した。インタラクションマネージャー音声認識では辞書の選択、画像認識では、人間のジェスチャーを認識して、キャラクターの意志決定をしている。

(4) 各種のハンドラー：シーンマネージャー、インタラクションマネージャーの配下にあり、各種の入力装置や出力装置を制御する機能を持つ。

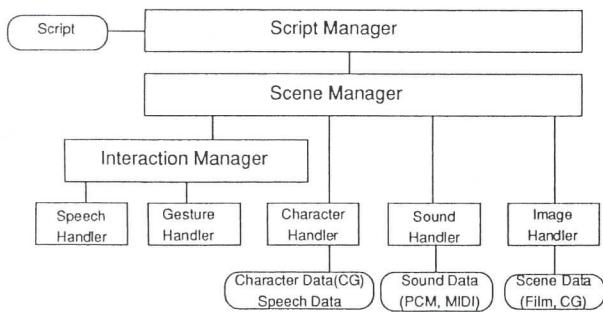


図 14 . システム構成

Fig.14 System configuration for Interactive Theater

6.3 インタラクティブシアター

「Wonderland」のコンセプト

我々日本人は、子供の頃、昔話を聞かされて育つ。語り継がれた昔話にほのぼのとした印象があるのは、大人の子供に対する愛だ。原書では、必ずしも暖かい話ではなく、話も単純ではない。隠された意味も様々な解釈があり奥深い。だから時代を超えて語り継がれ、いぶし銀のように魅惑的である。インタラクティブシアター用脚本は、日本古来の「浦島太郎」と、西洋の「不思議の国のアリス」から現代風に脚色した。採用した根拠は、「浦島太郎」は、現実の世界から竜宮城という仮想世界へ旅だって、再び戻ってくるという物語の構造を持ち、極めてVRを用いた物語の表現に適している。「不思議の国のアリス」も、また同じ様な構造を持ち、物語のビジュアル的表現が効果的である。このような理由から、この2つの物語を脚色し、「ワンダーランド」と名付けた。物語に登場する電子フェロモン城とは竜宮城の置き換えで、「テクノロジーの中毒性＝快楽の城」と意味付けている。

6.4 インタラクティブ脚本の構造

観客は仮想世界に入り、第3者ではなく実際の登場人物「浦島さん」を演じ、他の登場人物（CG俳優）とのやりとり（会話、ジェスチャーなど）により、様々な体験や問題を抱えて悩み、葛藤する感動を味わう。各シーンの中ではキャラクターと音声の感情認識によりインタラクションができる。

城の中では、インタラクティブポエム、インタラクティブミュージックなどのインタラクティブアートを導入している。インタラクティブポエムは、音声認識により観客とコンピューターキャラクターが、あらかじめフレーズごとに切ってある詩のフレーズを選んで、連歌のように掛け合いで即興詩を詠んでいくものである（図15右上）。インタラクティブミュージックは、観客がポップスの曲に合わせて、キーボードを即興演奏することで、そのひき方（高低、強弱、音数）によって、インタラクティブにダンシングシンガーのダンスが変わるものである。例えば、高音域で強くひくと、力強く踊り、単音でゆっくりした退屈な演奏をすると、座りこんだり、眠りだしたりする（図15左下）。

また、インタラクションに基づいたストーリーの展開を実現するため、随所に分岐点が設けてあり、例えば下記のようなものがある。

- a. いじめられている兎を助けるか、否か
(図15左上)
- b. 兎を助けなかった場合、あやまるか、否か
- c. 城で、どのアトラクションを選ぶか
- d. 城にとどまるか帰るか
- e. 玉手箱を開くか、否か (図15右下)

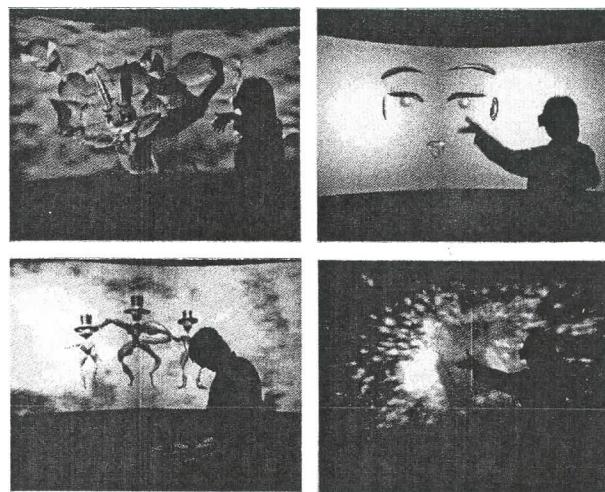


図 15.

インタラクティブシアター「ワンダーランド」より

Fig.15 Interactive Theater "Wonderland"

これらは、誘惑である。その誘惑にのった結果、期待通りの結果になるとは限らないが、誘惑とは魅惑的な物である。そこに観客の心理が働き、そこで起こる物事を主観的に受け止め、物語を体験する。具体的なストーリーの流れを図16に示す。

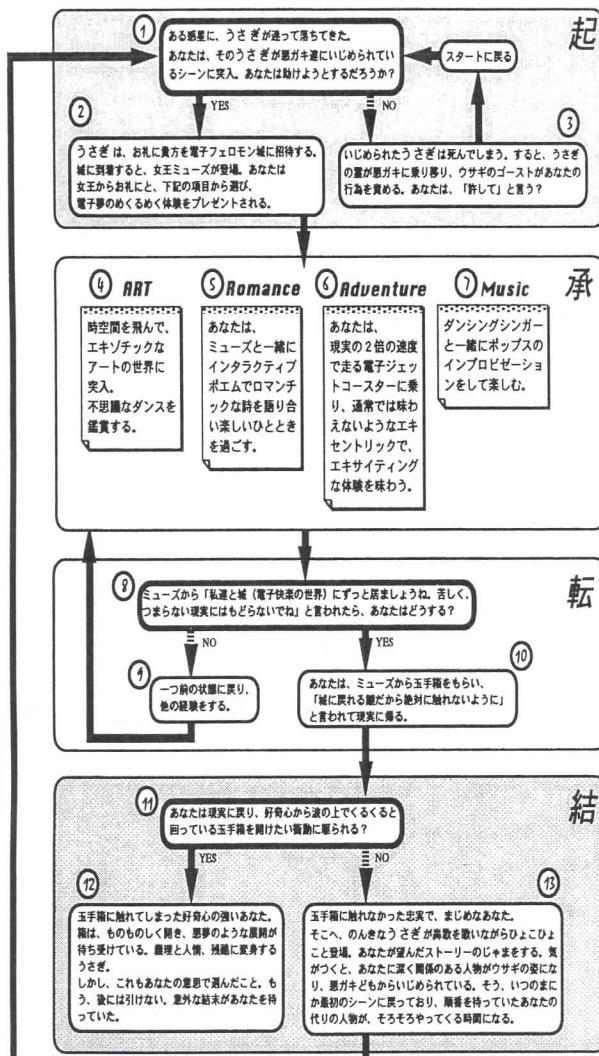


図16.ストーリーの流れ図
Fig.16 Story of Wnderland

7.まとめ

筆者らの作成した作品を通して、感情や音楽のフィーリングに反応する自律型インタラクティブアクターの生成から、その仮想世界の生成までを述べてきた。これらの研究においては、アーティストと技術者のコラボレーションが大きなキーワードになっている。ニューロペイビー、ミック&ミューズ、インタラクティブ・シアターというアプローチは次のように整理することが出来る。

(1) ニューロペイビー：人間のコミュニケーションの基本である感情（非言語情報の1つ）を用いて

人間とコミュニケーションできる自律エージェントの生成。

- (2) ミック&ミューズ：非言語コミュニケーション能力の向上・拡張（感情によるコミュニケーションの高度化、音楽によるコミュニケーション、ジェスチャーによるコミュニケーション機能の追加）。
- (3) インタラクティブ・シアター：上記(1)(2)で研究した自律エージェントと人間とのコミュニケーションにストーリーとシアター型アーチスクリーンを用いた立体視映像を導入。

アーティストが先端技術を用いて作品を制作するには、他の芸術と同じように、その技術を使いこなせるように理解しなければならない。理解した上で技術に対して技術者と違う自分なりの使い方を見い出して、初めてアート的な表現になる。そして、発見したアーティストは、既成概念を壊してオリジナリティを持って再構築していく。またそれに触発されて、技術者が既存の技術の最適な組み合わせにとどまらず、新たな技術の開拓し進める。このような協力関係により新しいメディアが作られていくのではないかだろうか。

参考文献

- [1] P. Maes, T. Darrell, B. Blumberg, and A. Pentland, "The ALIVE system: Full-body interaction with autonomous agents," Proc. of the Computer Animation'95 Conference (1995).
- [2] K. Perlin, "Real time responsive animation with personality," IEEE Transactions on Visualization and Computer Graphics, Vol.1, No.1 pp.5-15 (1995).
- [3] J. Bates, B. Loyall, S. Reilly, "An architecture for action, emotion, and social behavior," Proceedings of the Fourth European Workshop on Modeling Autonomous Agents in a Multi-Agent World (1992).
- [4] N. Tosa, et al., "Neuro-Character," AAAI '94 Workshop, AI and A-Life and Entertainment (1994).
- [5] N. Tosa, et al., "Network Neuro-Baby with robotics hand," Symbiosis of Human and Artifact, Elsevier Science B.V. (1995).
- [6] N. Tosa, R. Nakatsu "Life-like Communication Agent--Emotion Sensing Character 'MIC' and Feeling Session Character 'MUSE,'" Proceedings of the International Conference on Multimedia Computing and Systems, pp.12-19(1996).
- [7] Shirley Weitz, "Nonverbal Communication," New York: Oxford Univ. Press(1974).
- [8] S. Mozziconacci, "Pitch variations and emotions in speech," ICPHS 95 Vol. 1, pp. 178 (1995).
- [9] K. R. Scherer, "How emotion is expressed in speech and singing," ICPHS 95, Vol. 3, pp. 90 (1995).
- [10] 中津、土佐 「インタラクティブ映画構築に向けて」電子情報通信学会、画像工学研究会(1997).
- [11] P. Doyle, B. Hayes-Roth, "Computer-aided exploration of virtual environments," Proceedings of AAAI Workshop on AI and Entertainment, AAAI Press, Menlo Park, CA, 1996
- [12] F. Sparacino, et al., "HyperPlex:a World of 3D Interactive Digital Movies," Proceedings of AAAI Workshop on AI and Entertainment, pp.77-81(1995).

(1996年12月16日受付)