

映像情報メディア学会誌

Vol.52 No.1 1998

1998 1

THE JOURNAL OF THE INSTITUTE OF IMAGE INFORMATION AND TELEVISION ENGINEERS



社団法人 映像情報メディア学会 (<http://www.ite.or.jp>)
THE INSTITUTE OF IMAGE INFORMATION AND TELEVISION ENGINEERS

5. インタラクティブポエム — 芸術と科学の「融合地点」 —

土佐尚子[†], 中津良平[†]

1. まえがき

人間は、会話を意味を求める。たとえ相手が人形でも、物でも、話しかけることがある。そうしたときに、「たまごっち」のようなスイッチのリアクションでは、暖かみと豊かさがない。そこで、言葉の意味を踏まえたコンピュータキャラクタとのインタラクションというものを考えてみた。しかし、日本の情緒ある「あうん」の会話をコンピュータとリアルタイムに行うのは無理である。そこで、土台となるような状況の枠を設計することを考えていたとき、我々の古来の文化である連歌を思いつく。そして連歌形式で、コンピュータ詩人と人間が即興詩を作る考えにたどり着いた。本来「詩」は、詩人が表現したい「メッセージ」を、「言葉の力」を用いて個人の世界を作りあげたものであり、それを読んだり聞いたりすることによりその世界を、我々は享受している。「インタラクティブポエム」は、享受するだけではなく、能動的に人間がコンピュータと創作することにより、思いもよらぬハプニング(偶然性)、誰がコンピュータと創作するかによる個性の違いが出てくると期待される。これらは「インタラクション」作用により開放的になる。つまり、「対話」の機能を取り入れることで感性によるコミュニケーションの媒体ができあがる。

2. インタラクション

大きなスクリーンにギリシャ神話の音楽の女神「MUSE」の顔が浮かび上がる。MUSEは、まるでいっしょに歌うように、人間と対話しながら詩を紡ぎ出す。MUSEはユーザに対して、短い詩的な言葉を感情を込めて語りかける。それを聞いてユーザはその詩の世界に入り、同時に、ユーザ自身も詩的な言葉を、MUSEに語り返す。この詩的な言葉をやりとりする「対話」プロセスを通じて、「インタラクティブポエム」は、インスピレーション、フィーリング、そして、感性に満ちた即興詩の世界を、ユーザとコンピュータがいっしょに創りあげができる(図1: カラー頁)。

3. ソフトウェア構成

インタラクティブポエムのシステムは、システム制御、音声認識、イメ

システム制御ユニット

出力用
フレーズ
データ集コンピュータ
グラフィックス
アニメーション
生成ミューズ顔表情
モデルリング
データベース実写背景用動画
データベース

音声認識システム

フレーズ認識
感情認識

図2 インタラクティブポエムのソフトウェア構成システム

ジ生成、および、音声出力の4種類のユニットから構成される(図2)。

システム制御ユニットは、データベースを使用して、システム全体の動作をコントロールする。このシステムにおいて、最も重要な課題は、インタラクティブな詩を作成することである。最初にインタラクティブポエムを記憶しているデータベースの内容を説明する。従来の詩は、詩的なフレーズが連続したものと考えられる。つまり、従来の詩の基本的構成は、単純な状態遷移ネットワークによって表現が可能である。このネットワークでは、各フレーズは状態に対応し、各々の状態には1つの状態が接続する(図3)。

インタラクティブポエムの基本形式は、この遷移ネットワークによって表現される。この違いは、コンピュータの発声するフレーズとユーザの発声するフレーズが交互にあらわれることである。これは、コンピュータとユーザ

[†] ATR知能映像通信研究所

"Interactive Poem—A New Type of Poem Where Art and Technology meet—" by Naoko Tosa and Ryohei Nakatsu (ATR Media Integration & Communication Research Laboratories, Kyoto)

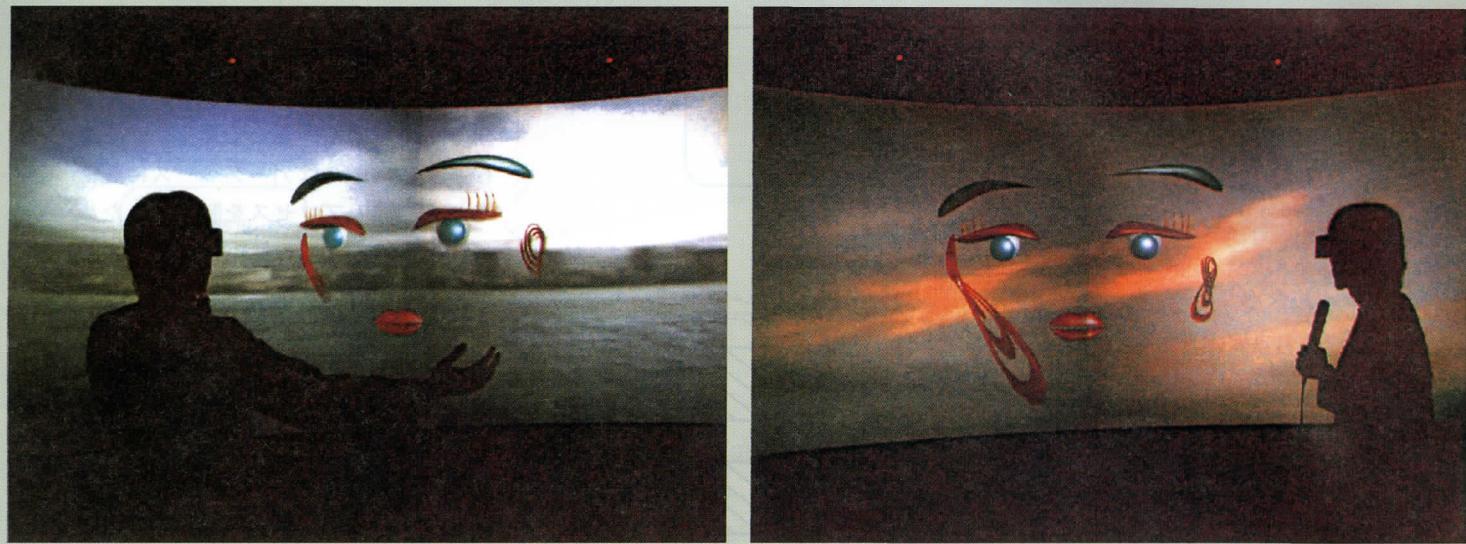


図 1 インタラクティブ詩エム

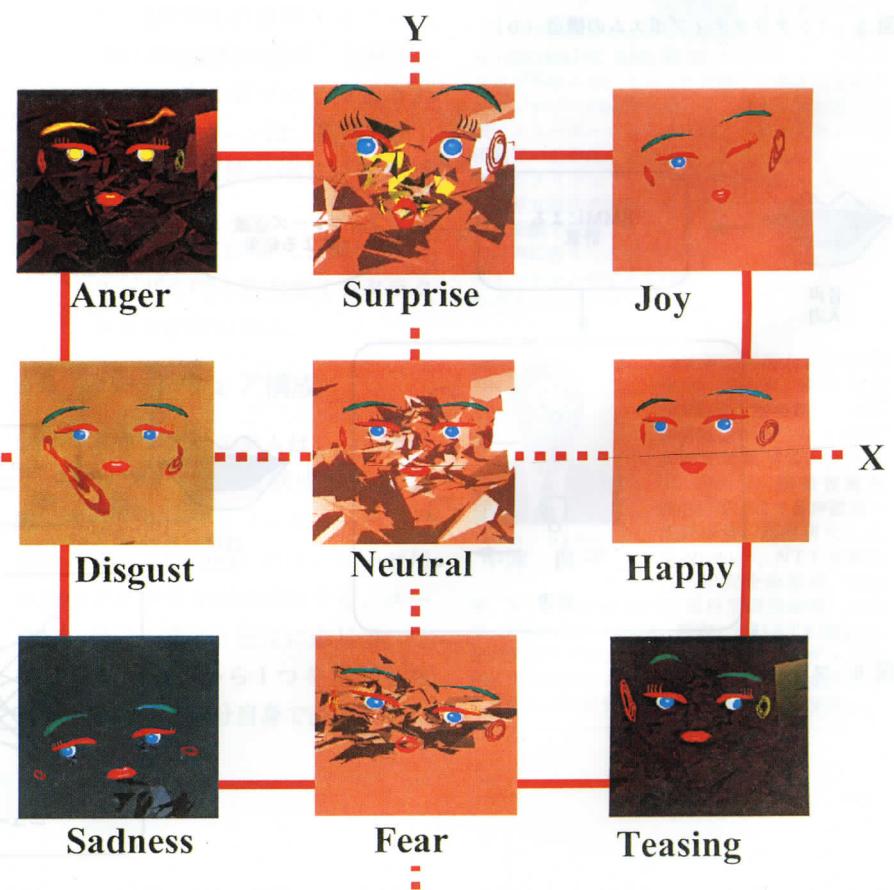


図 8 ミューズの
感情的顔表現

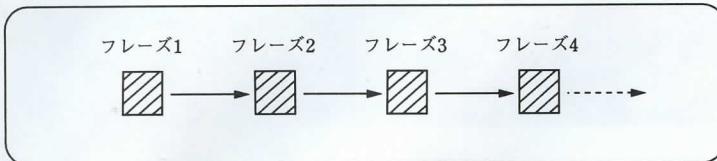


図 3 従来の詩の構造

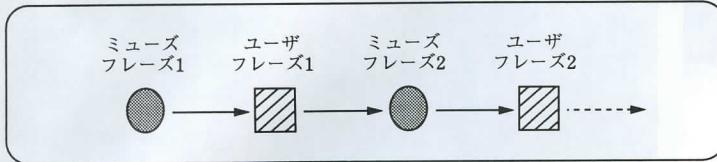


図 4 インタラクティブポエムの構造 (a)

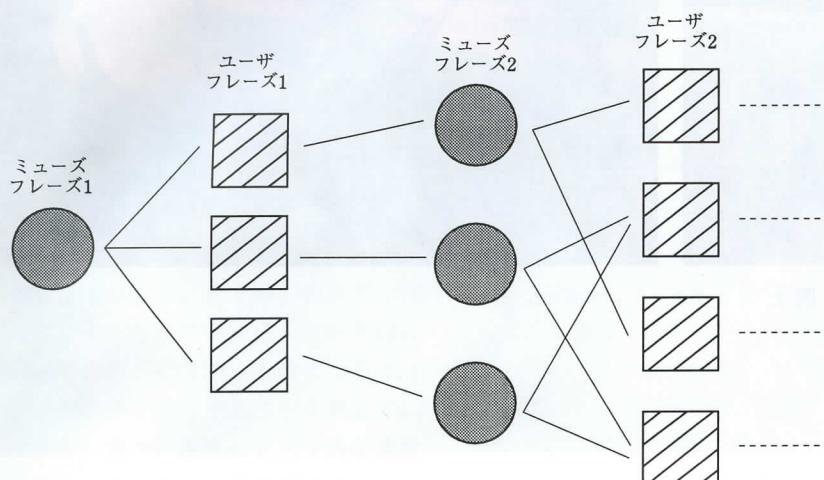


図 5 インタラクティブポエムの構造 (b)

が、あらかじめ決められた詩のフレーズを交互に読むという単純な対話に対応する(図4)。本インタラクティブポエムシステムでは、より高度な対話機能を実現するために、コンピュータの発声するフレーズに接続可能な複数のフレーズを用意してある。これらのフレーズは、コンピュータのフレーズとの音韻的なまた意味的なつながりを考慮して注意深く作成してある。この遷移ネットワークはデータベースに記憶され、全体のプロセスを制御するために使用される。この仕組により、ユーザーは自分の感情・感性に合ったフレーズを選択しコンピュータに語りかけることにより、オリジナルな詩の世界に自分自身の感情・感性を加えることができる(図5)。

音声認識ユニットは、意味認識と感情認識の2種類の認識機能を持つ。ユーザーが発声するフレーズの意味内容を認識するため、HMM (hidden Markov model)に基づいた不特定話者音声認識を用いる。発声される各フレーズは、音韻系列の形で表現され、認識用の辞書に記憶される(図6)。また、同時にユーザーの感情状態を認識するために感情認識を行う。感情認識のための基本アーキテクチャとしては、ニューラルネットワークを用いる。このニューラルネットワークは、多数の話者が8種類の感情で多数の単語を発声した音声データを用いて学習している。この学習により、不特定話者、内容独立型の感情認識が実現される(図7)。

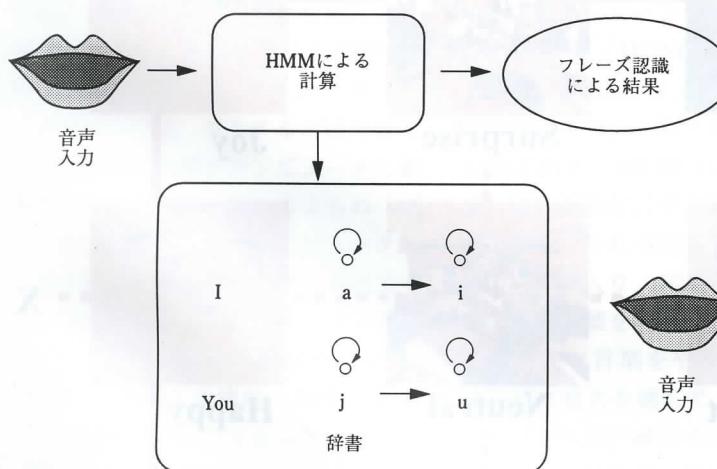


図 6 フレーズ認識

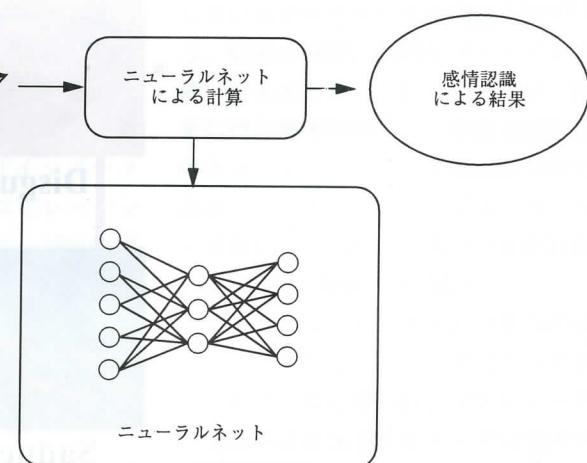


図 7 感情認識システム

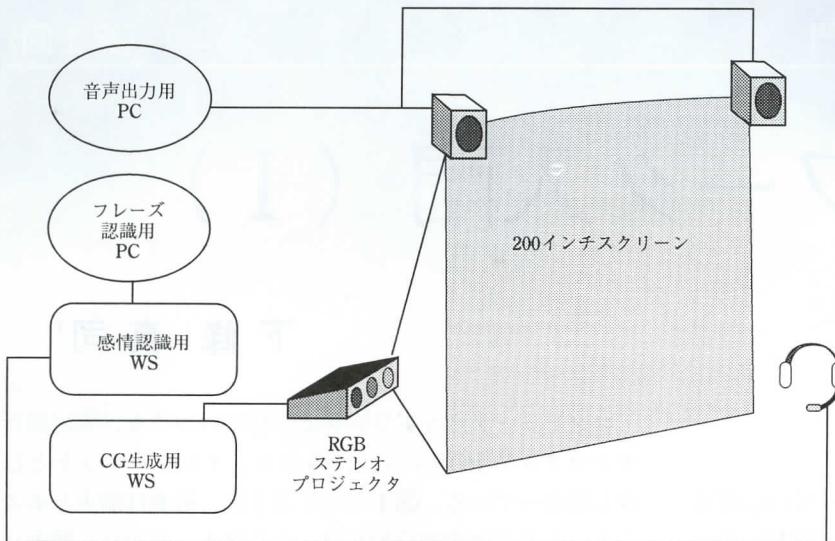


図9 インタラクティブポエムのハードウェア構成

ユーザの発声するフレーズに対するコンピュータキャラクタMUSEの反応は、音声と映像によって表現される。音声出力ユニットでは、MUSEの発声するフレーズに対する音声データが記憶されており、必要に応じて再生成される。コンピュータグラフィックス生成ユニットは、映像表現をコントロールする。映像は、MUSEの表情と背景シーンの2種類から構成される。MUSEの表情は、ユーザの感情に対するMUSEの反応を表す。これらの表情は、8種類の感情に対応した3次元モーフィングアニメーションによって表される(図8;カラー頁)。また、詩の世界の雰囲気を表現するために、何種類かの背景シーンを記憶しておき、遷移ネットワークの状態に応じて適切な背景映像を表示する。

4. 処理プロセス

- MUSEがフレーズを発声すると同時に、認識プロセスが起動される。ユーザが発声すると、その言葉は、コンピュータの意味認識機能によって認識される。この時、遷移ネットワークの次の状態に対応した辞書サ

ブセットを用いて認識を行う。同時に、ユーザの言葉に含まれている感情が、コンピュータの感情認識機能によって認識される。

- 認識結果と遷移ネットワークに基づいて、システムの反応が決定される。MUSEの表情は感情認識の結果に応じて変化し、MUSEが発声するフレーズは意味認識の結果と遷移ネットワークに基づいて決定される。背景シーンは、遷移ネットワークの状態に応じて変化する。
- このようにして、MUSEとユーザの間の詩的なインタラクションが行われる。

5. ハードウェア構成

ハードウェアシステムは、ワークステーションとPCから構成される(図9)。また、ユーザとコンピュータの対話をスムーズに行うために、ユーザ用の詩のフレーズを画面表示する。ユーザは、自分の感情と感性に応じて、これらのフレーズから1つを選択したり、あるいは、自分自身で詩的なフレ

ーズを作成して発声することができる。ユーザが辞書がないフレーズを発声した場合は意味認識機能は、それに最も近い、既存のフレーズを選択する。これによりコンピュータとユーザの即興的な対話も可能になる。

6. むすび

ユーザとコンピュータが、短いフレーズや文章をやりとりすることで「対話」を行い、そして、この対話から、オリジナルな詩の世界にユーザ自身の感情・感性を取り入れた新しい詩の世界の創作をコンピュータが助けている。これは、つまり人間自身の情操を高め、広げていく楽しさを持ちながら、コンピュータと心地良い関係の状態である。この快さを育てることが大切である。

(1997年10月16日受付)



土佐 尚子

アーティスト、ATR知能映像通信研究所客員研究员、神戸大学客員助教授、武蔵野美術大学映像学科非常勤講師。感情、意識、無意識といった見えないものを視覚化することをテーマにテクノロジーを表現ツールとしてメディアアート、人工生命、インタラクティブシアタープロジェクトを展開している。その作品SIGGRAPH/ARS ELECTRONICA(アルスエレクトロニカ)といった世界で代表的な芸術とテクノロジーの国際会議にて、講演と共に作品発表。ニューヨーク近代美術館、メトロポリタン美術館等の企画展に招待展示、アメリカンフィルムアソシエイション、国立国際美術館、O美術館、富山県立近代美術館、名古屋県立美術館、高松市立美術館で収蔵されている。最近の仕事には、人間の声に含まれる感情を認識し、人間とのコミュニケーションができる自律キャラクター「ニューロペイビー」がある。



中津 良平

1969年、京都大学工学部卒業。1971年、京都大学大学院修士課程修了。日本電信電話公社(現NTT)武蔵野電気通信研究所入所。1980年、横須賀電気通信研究所にて音声認識の基礎研究、応用研究に従事。1990年、NTT基礎研究所企画部長。1991年、NTT基礎研究所情報科学研究部長。1994年、(株)ATRに出向。現在、(株)ATR知能映像通信研究所代表取締役社長。マルチメディア要素技術の研究およびマルチメディア技術を応用した通信方式の研究などを行っている。工学博士。