

Life like Computer Character  
[MIC & MUSE]

土佐尚子 (ATR知能映像通信研究所)

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、からくり人形、ぬいぐるみ、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、感情移入しているではないか。アーティストも、人工生命的技法をアートの分野での表現技法

としてとりいれている。技法としてだけではなく、もはやこのメディア自体が作品のコンセプトになり、作品は作家の手から離れ自立し未来へ向かって成長していくことを目指すのだ。

私の表現テーマは、意識、無意識、深層心理、感情といった目に見えない一瞬で不確実で不条理な、しかし深く鋭く、恥ずかしくなるくらい人間的でもろい感性の視覚化である。もはや、従来の芸術がやってきたように人間の情念の断片を表現するだけでは、神経が

突出した現代人を描ききれない。むしろ現代社会におけるこの「人型」ラベルを用いて人工生命の美学と近未来的コードのコミュニケーションをシュミレーションしてみよう。近未来の個人のパーソナリティ像は、自己の多様化と増殖化である。我々は、エージェントと呼ばれる自己の分身をいくつか持つようになるであろう。今回の作品では、自己の分身であり自分に最も近い他者をキーワードに、音声認識研究者中津良平氏との共同研究で、話し掛ける人の声から音声認識により感情



27——土佐尚子《ミック》1995(平成7)年

抽出し感情をスキャンして、インタラクティブに答えるヒューマノイドエージェント「MIC」と「MUSE」との新しいコミュニケーションの形態を紹介する。

MIC(ミック)は3年前に富士通研究所の協力で開発制作したニューロベイビーを基に、さらに知識を習得して我々現代人とともに、現在進行形で成長していく人間の子供をモデルにしたパーソナリティーを持つ仮想の男の子のキャラクターである。彼は、下記のような8つの感情を\*に指示している声の抑揚から認識できる。

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、エキサイティングしてジャンプをし、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、おしゃまな性格である。

- a. Joy (happiness, satisfaction, enjoyment, comfort, smile)  
\* exciting, vigorous, voice rises at the end of a sentence
- b. Anger (rage, resentment, displeasure)  
\* voice falls at the end of a sentence
- c. Surprise (astonishment, shock, confusion, amazement, unexpected)  
\* screaming, excited voice
- d. Sadness (sadness, tearful, sorrow, loneliness, emptiness)  
\* weak, faint, empty voice
- e. Disgust  
\* sullen, aversive, repulsive voice
- f. Teasing  
\* light, insincere voice
- g. Fear  
\* frightened, sharp, shrill voice

- h. neutral  
\* normal voice

MUSE(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。ミックと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、ミューズとは感情表現する言葉に満たない、もっ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、人間は下記の感情を表すミュージカルグラマー(楽典)を用いて彼女と即興的なフィーリングセッションを楽しむ。

- a. Joy --- rising musical scale, elevated, allegro
- b. Anger --- vigoroso, 3 times same sound (repetitious)
- c. Surprise --- several times same sound (repetitious)
- d. Sadness --- falling musical scale, volante



- e. Disgust---dissonant sound, discord
- f. Teasing---scherzando
- g. Fear---pesante

マクルーハンの公式「メディアは、メッセージである」は、音やイメージの明示的内容ではなく、現実を等価記号の連鎖へと分解するというこれらのメディアの技術的本質に結びついたものであることを示唆している。基本的に技術水準の飛躍的向上とともにありのま

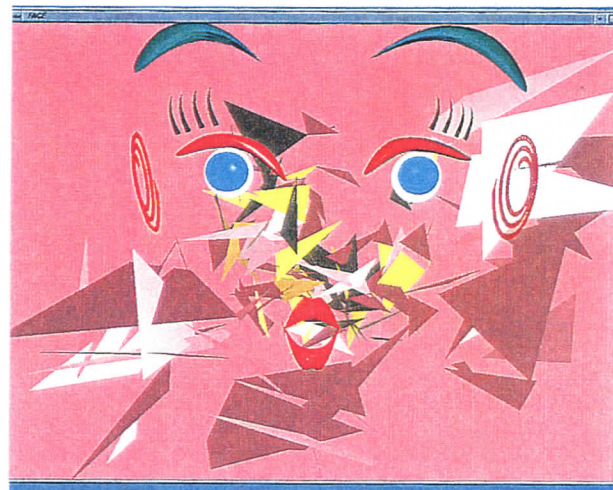
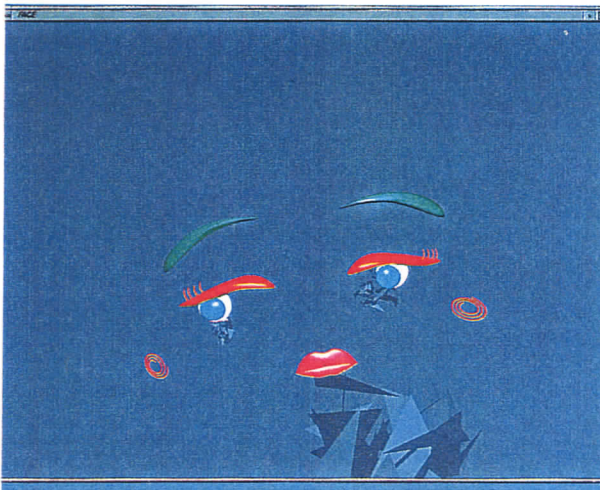
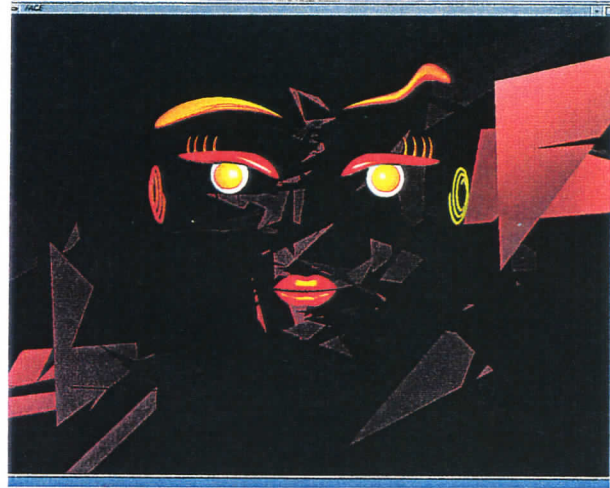
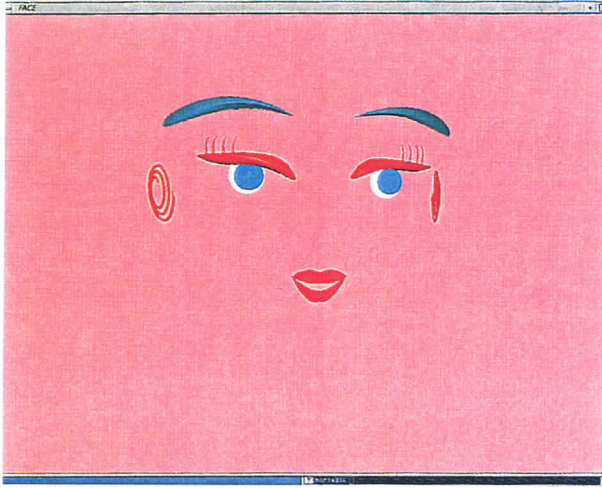
まの現実、我々から遠ざかっていく。  
今、まさに人工生命という仮想の生命を作り、それ自身とコミュニケーションをとる我々の未来は、どこへ行くかとしているのだろうか。

音声に含まれる感情の認識

中津良平 (ATR知能映像通信研究所)

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



28——土佐尚子《ミューズ》1995(平成7)年

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処理する場合、これまでは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。

2. 感情認識の方法

感情を認識するには、あらかじめ大量の音声データを用意しておき、これらを統計的に識別できるようにシステムを学習させておくという手法をとる。ここでは統計的手法の1つとして「ニューラルネットワーク」と呼ば

れる手法を用いている。これは、神経細胞に相当する多数のノードを相互に接続したいわば、脳の構造を模したネットワークであり、ノード間の接続の強さを調整することにより高い認識機能を実現できる特徴をもつ。多数の人が種々の単語をいろいろな感情を交えて発声した音声データを用意しておき、このデータを用いて事前にニューラルネットの学習を行わせてある。このネットワークに音声を入力することにより音声の中の感情の認識が行われる。