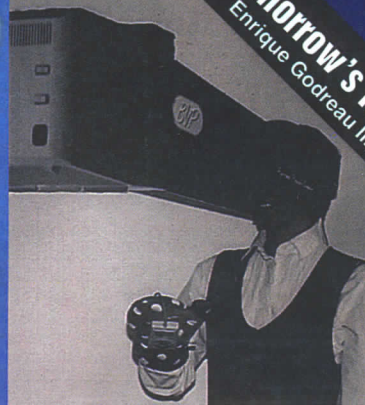# Visual Proceedings

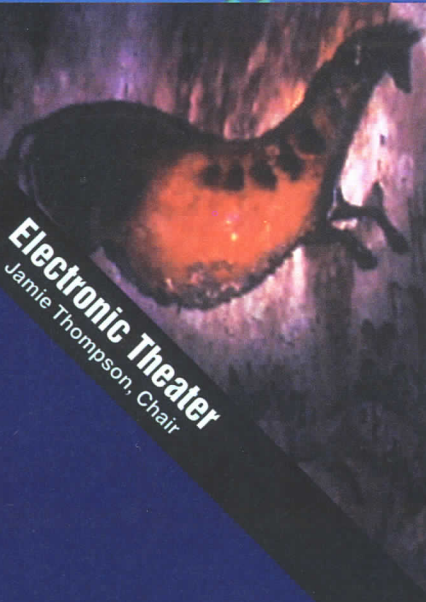## THE ART & INTERDISCIPLINARY PROGRAMS OF SIGGRAPH 93

Machine Culture
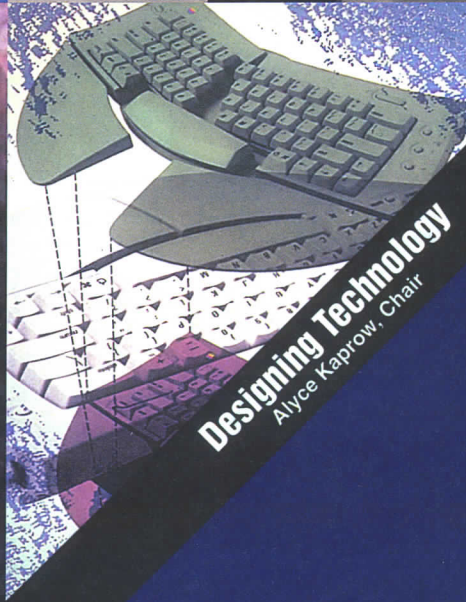Simon Penny, Chair

Tomorrow's Realities
Enrique Godreau III, Chair

Electronic Theater
Jamie Thompson, Chair

Designing Technology
Alyce Kaprow, Chair

Computer Graphics
Annual Conference Series, 1993

A Publication of ACM SIGGRAPH

Revised Second Printing

Thomas E. Linehan, Editor

Neuro Baby

- - - - -
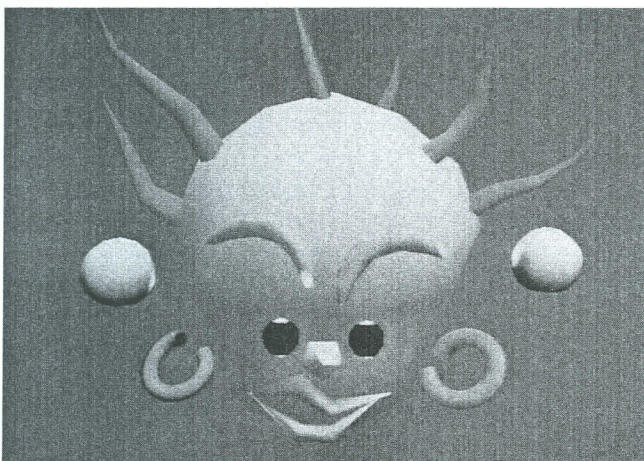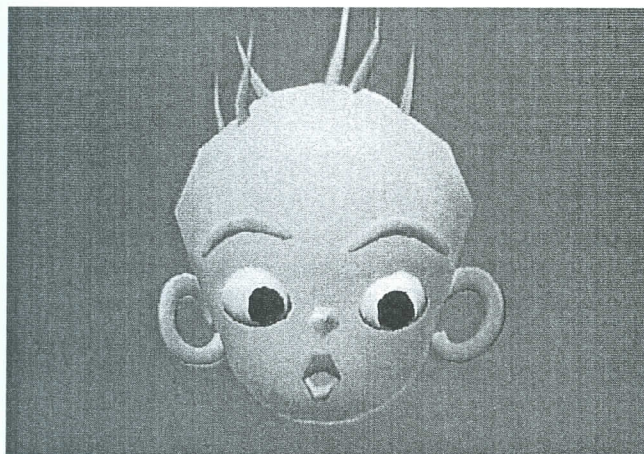
NAOKO TOSA

Musashina Art University

An automatic facial expression synthesizer that responds to expressions of feeling in the human voice.

I created a new creature or a piece of work that can live and meaningfully communicate with modern, urban people like ourselves, people who are overwhelmed, if not tortured, by the relentless flow of information, and whose peace of mind can only be found in momentary human pleasures. Neuro Baby was born to offer such pleasures.

The name "Neuro Baby" implies the "birth" of a virtual creature, made possible by the recent development of neurally based computer architectures. Neuro Baby "lives" within a computer and communicates with others through its responses to inflections in human voice patterns. Neuro Baby is reborn every time the computer is switched on, and it departs when the computer is turned off. Neuro Baby's logic patterns are modeled after those of human beings, which make it possible to simulate a wide range of personality traits and reactions to life experiences.

Neuro Baby can be a toy, or a lovely pet—or it may develop greater intelligence and stimulate one to challenge traditional meanings of the phrase "intelligent life." In ancient times, people expressed their dreams of the future in the media at hand, such as in novels, films, and drawings. Neuro Baby is a use of contemporary media to express today's dreams of a future being.

This work is the simulation of a baby, born into the "mind" of the computer. Neuro Baby is a totally new type of interactive performance system, which responds to human voice input with a computer-generated baby's face and sound effects. If the speaker's tone is gentle and soothing, the baby in the monitor smiles and responds with a pre-recorded laughing voice. If the speaker's voice is low or threatening, the baby responds with a sad or angry expression and voice. If you try to chastise it, with a loud cough or disapproving sound, it becomes needy and starts crying. The baby also sometimes responds to special events with a yawn, a hiccup, or a cry. If the baby is ignored, it passes time by whistling, and responds with a cheerful "Hi" once spoken to.

The baby's responses appear very realistic, and may become quite endearing once the speaker becomes skilled at evoking the baby's emotions. It is a truly lovable and playful imp and entertainer. In many ways, it is intended to remind speakers of the lifelike manner of the famous video-computer character Max Headroom.

Two major technologies were combined to create this system: voice analysis and the synthesis of facial expressions.

Voice analysis was performed by a neural network emulator that converted the voice input wave patterns into "emotional patterns" represented by two floating point values.

The neural network has been "taught" the relationship between inflections in human voices and emotional patterns contained within those inflections. During interaction with the baby the emotional patterns found in the observer's speech are continuously generated.

During the translation stage, the two values for emotional patterns are interpreted as an X-Y location on an emotional plane, onto which several types of emotional patterns are mapped. For example, "anger" may be located on the lower left of such a plane, while "pleasure" would be located on the upper right of the same plane. Each emotional pattern corresponds to a paired facial expression and a few seconds of voice output.

During the performance, the facial expression is determined by interpolating the shape, position, and angle of facial parts, such as eyes, eyebrows, and lips. These parts were pre-designed for each emotional reaction.

One FM TOWNS, Fujitsu's multimedia personal computer, is used for voice analysis, another FM TOWNS is used for voice generation, and a Silicon Graphics IRIS 4D is used for image synthesis.

Contact

Naoko Tosa
Musashino Art University
2-33-9-3A Ogikubo Suginami-Ku
Tokyo
Japan 167
81.427.44.9711
81.3.5397.3797 fax

Contributor

Koichi Nurakami