the Netherlands. Other researchers working on algorithmic composition include ***George Lewis, Christopher Dobrian, Carl Stone,*** and ***Clarence Barlow.*** Books such as *Music with AI: Perspectives on Music Cognition,* edited by Mira Balaban, Kernal Ebcioglu, and Otto Laske, summarize some of this research. ***Ron Pellegrino*** investigates "visual music," including systems to visualize underlying structures as a compositional tool.

Image and Speech Recognition

Over the years, AI researchers have toiled to create programs and computer systems that could successfully recognize speech or make sense of visual information. They have achieved some success. Commercial programs are available for speech recognition that can perform tasks once thought impossible, for example, speaker-independent, telephone-response systems. Image recognition has similarly flourished, for example, OCR in text scanning, missile targeting systems based on scene understanding, robot vision systems for patrolling assembly line problems, and fingerprint recognition systems for law enforcement. Each of these accomplishments require sophisticated knowledge representation and deduction systems that were only research topics a few decades ago.

As is typical of emerging research fields, few people refer to the exotic research origins once they become incorporated into everyday technical realities. For example, rarely does anyone hear mention of artificial intelligence in the product promotions for OCR. Many challenges still confront researchers in areas that could be called artificial intelligence, but they are no longer identified under that unified rubric. For example, researchers continue to work on face recognition, the interpretation of live video, and gesture recognition systems. Since speech recognition and visual interpretation are seen as uniquely human abilities, it seems appropriate that artists would be interested, although they, like the researchers, have moved into fields such as gesture recognition, motion tracking, and face recognition (see chapter 7.4).

Naoko Tosa

Japanese artist Naoko Tosa walks in both the art and research worlds. She works at the ATR Media Integration Research Center on pioneering new, more powerful interfaces between humans and machines. Her interests include speech recognition, gesture recog-

---

Ron Pellegrino: ⟨http://www.microweb.com/ronpell/VisualMusic.html⟩
Naoko Tosa: ⟨http://www.mic.atr.co.jp/~tosa/⟩

nition, artificial intelligence, and affective computing (see chapter 2.5). She creates installations and interactive movies in which the characters react to visitors in complex, sensitive ways. Her projects are full of poetry and scientific rigor, and win acclaim in both the art and research worlds. She exemplifies the new hybrid artist-researcher beginning to emerge as artists become pioneers in working with new technologies. A sampling of her presentations and papers illustrate the range of her work, for example, "The Esthetics of Artificial Life Characters," "Artistic Communication for A-Life and Robotics," and "Creating a Movie with Autonomous Actors That Respond to Emotions."[6]

Neuro Baby, one of her best-known installations, uses sophisticated neural-network programming to create a computer graphic entity that responds to the emotional tones of voices. The baby responds appropriately with crying or cooing depending on the way the viewer addresses it. The artist's statement from the Ars Electronica showing in 1993 explains the intention:

> 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 prerecorded 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.[7]

The artist sees *Neuro Baby* as an exploration of possible directions in future artificial life that will show signs of intelligence and sensitivities beyond normal speech recognition:

> *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.[8]

The extension of *Neuro Baby,* called *MIC* and *MUSE,* elaborated on the themes. *MIC* was a male baby with a distinct personality who responded more clearly to a variety of emotions, using voice patterns to make interpretations of emotions. For example, joy (happiness, satisfaction, enjoyment, comfort, smile) was extrapolated from an exciting, vigorous voice that rises at the end of a sentence. Anger (rage, resentment, displeasure)

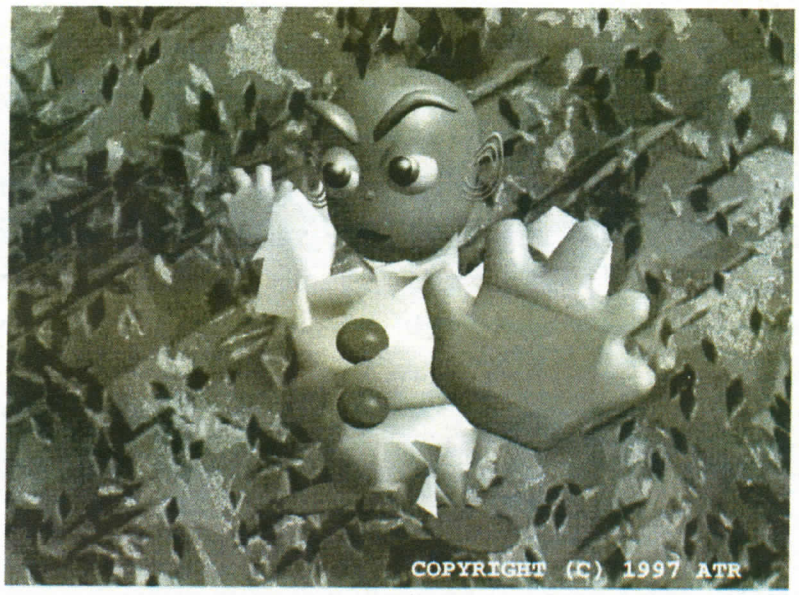

Fig. 7.6.4. Naoko Tosa, ATR—MIC Labs. Mic, an artificial computer character that detects emotional tone in speech and reacts with its own emotional response.

was derived from a voice that falls at the end of a sentence. Other emotions it attempted to read included surprise, sadness, disgust, teasing, and fear. The work also allowed viewers to make music to communicate particular emotions. An ambitious later system called *Play Cinema* (in collaboration with Ryohei Nakatsu and Takeshi Ochi) attempted a fully integrated, networked, multi-person, interact-anytime interactive cinema system in which participants could interact with fictional characters, such as in a demonstration version of *Romeo and Juliet.* The system included speech, gesture, and emotion recognition.

Other Artists and Projects

Nick Baginsky produced the installation *The Three Sirens* based on neural networks. These improvisational robots learn rules about improvisation and instrumental sound. They analyze the sound they hear and learn to control their motors and mechanical sound makers to create further improvised music. MIT Media Lab researcher ***Jonathan Klein*** has created *TurnStyles: A Speech Interface Agent to Support Affective Communication.* This system detects

---

Nick Baginsky: ⟨http://www.provi.de/~nab⟩
Jonathan Klein: ⟨http://affect.www.media.mit.edu/projects/affect/AC_research/projects/turnstyles.html⟩

the emotional content of speech using techniques such as conversational pacing, turn-taking patterns, and "expressive paralinguistics" in order to adjust its own voice output to be more in tune with the user's emotional state. Commercial researchers have pursued related lines of voice inquiry, although not in the same way, for example, voice analyzers that attempt to determine stress or lying. There is even one called *Sales Edge* that purports to let you know when a prospect is ripe for the closing. In these kinds of research, investigators are trying to create artificial intelligence that is at the edge of or beyond human capabilities. ***Ray Kurzweil*** developed the *Cybernetic Poet,* which can produce respectable poetry in a variety of styles.

Interactions with Artificial Characters

Researchers and artists are attempting to understand and model core elements of human personalities. Historically, researchers attempted to focus more on the information-processing capabilities of artificial intelligence, for example, could computers be programmed to excel at playing chess, solving mathematical problems, or helping a robot make simple visual determinations? Expert systems research attempted to model the decision-making processes of experts in narrowly defined realms such as medical triage, mineral prospecting, or making loans. A few researchers attempted to model more affective aspects of personality, such as Kenneth Colby's PARRY,[9] a modeling of a paranoid personality, but most of the field seemed to hold to the view that intelligence could be somehow disembodied and separated from affective elements.

Some artists question the wisdom of this approach. They attempt to model computer entities that demonstrate personality, intentionality, point of view, and emotions, and they create installations that investigate the responses of viewers to interactions that have these kind of digital characters. They suggest that the typical strategy of ignoring these elements of human functioning was doomed to create low-functioning and unacceptable digital characters. In my paper "AI Research as Art" I explained this view:

[M]any of the decisions to be made about the shape of AI programs are not purely technical. The simulation of human information processing outside of narrow realms, and the creation of machine partners which interest and satisfy humans, will depend on sophisticated artistic and psychological design choices as well. Some discourse about AI seems to imply that intelligence

---

Ray Kurzweil: ⟨http://www.kurzweilcyberart.com/⟩