

Reprinted from

Symbiosis of Human and Artifact

Symbiosis of Human and Artifact

Y. Anzai, K. Ogawa and H. Mori (Editors)

© 1995 Elsevier Science B.V. All rights reserved.

Network Neuro-Baby with robotics hand

(An automatic facial expression synthesizer that responds to expressions of feeling in the human voice and handshake)

Naoko Tosa* Hideki Hashimoto** Kaoru Sezaki** Yasuharu Kunii**
Toyotoshi Yamaguchi** Kotaro Sabe** Ryosuke Nishino** Hiroshi Harashima***
Fumio Harashima**

*ATR Media Integration & Communications Research Laboratories
2-2 Hikaridai Seika-cho Soraku-gun Kyoto 619-02 Japan

Department of Imaging Arts and Sciences, Musashino Art University
1-736 Ogawa-cho, Kodaira, Tokyo 187 Japan

**Institute of Industrial Sciences, University of Tokyo
7-22-1 Roppongi, Minato-ku, Tokyo 106 Japan

***Department of Electrical Engineering, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113 Japan



Network Neuro-Baby with robotics hand

(An automatic facial expression synthesizer that responds to expressions of feeling in the human voice and handshake)

Naoko Tosa* Hideki Hashimoto** Kaoru Sezaki** Yasuharu Kuni**
Toyotoshi Yamaguchi** Kotaro Sabe** Ryosuke Nishino** Hiroshi Harashima***
Fumio Harashima**

*ATR Media Integration & Communications Research Laboratories
2-2 Hikaridai Seika-cho Soraku-gun Kyoto 619-02 Japan

Department of Imaging Arts and Sciences, Musashino Art University
1-736 Ogawa-cho, Kodaira, Tokyo 187 Japan

**Institute of Industrial Sciences, University of Tokyo
7-22-1 Roppongi, Minato-ku, Tokyo 106 Japan

***Department of Electrical Engineering, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113 Japan

Abstract

Neuro-Baby (NB) is a totally new type of interactive performance system which responds to the human voice with a computer-generated baby face and sound effects. Emotion space model is employed to categorize the feelings of the speaker. To recognize the human voice we used a neural network which has been taught the relationship between a set of digitized wave patterns and the location of several emotion types in the emotion space. The facial expression is synthesized continuously according to the location which the neural network generates. The flexible design of NB is possible by changing the facial design, the layout in the emotion space, sensitivity to the transition of the feelings or the teaching pattern for the neural network.

By networking NB's, we can enjoy a non-verbal communication with each other. Such a Networked NB's will help the mutual understanding, absorption of cultural gap as well as international cultural exchange very much. The first result will be demonstrated in 1995, by connecting two NB's between Japan and USA. The networking issues concerning such a system is also addressed.

1. Introduction

A new creature has been born!! This creature 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. NB was born to offer such pleasures. The name "NB" implies the "birth" of a virtual creature, made possible by the recent development of neurally based computer architectures. NB "lives" within a computer and communicates with others through its responses to inflections in human voice patterns. NB is reborn every time the computer is switched on, and it departs when the computer is turned off. NB's logic patterns are modeled after those of humans beings, which make it possible to simulate a wide range of personality traits and reactions to life experiences. NB 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. NB is a use of contemporary media to express today's dream of a future being.[1]

2. System configuration for Network Neuro-Baby with robotics hand.

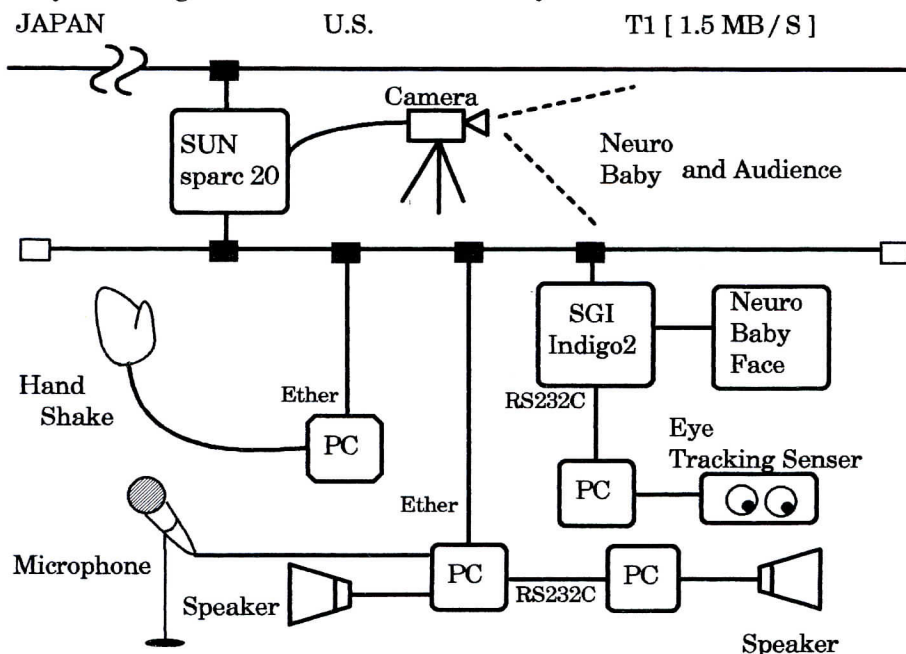


Figure 1.

3. Basic characteristic of Neuro-Baby and its interaction with the external world

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 sad 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.

Figure 2.

Figure 3 shows the general model of the NB from human input via recognition mapping R to a state in the emotion model, and then via the expression mapping E to the output

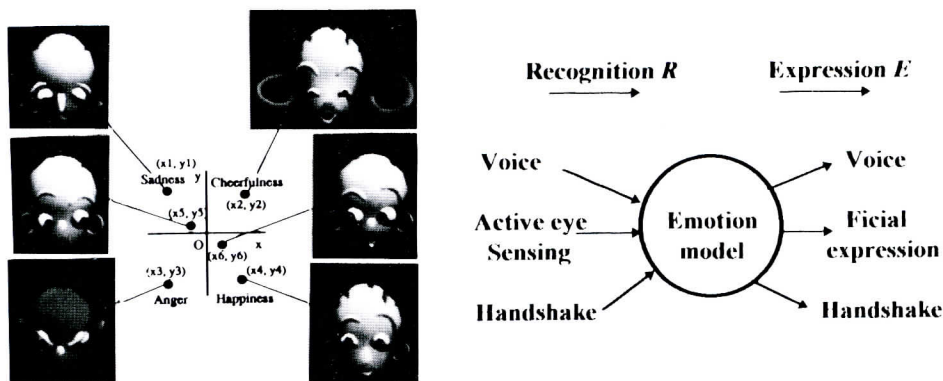


Figure 2. An assignment of the model faces. Figure 3. Processing model of the NB.

3-1. Customization and Learning of the relationship between voice and emotion

The principle function of NB is to make a map describing emotional responses evoked by voice input so that speakers can feel these emotional responses natural and comfortably. The emotional responses are expressed by using x-y axis coordinate shown in Figure 4. We call the coordinate as an emotional model. A point (x, y) corresponds to an action which NB performs to express his/her response. The coordinate of emotional model has been changed by a neural network with a set of sampled data such as sadness, cheerfulness, anger and happiness.[2]

NB has several types of emotional model and speakers can select one among them in accordance with speaker's characteristics. This selection is dependent on first input voice and reaction of handshaking machine. It is a kind of customization and realizes more delicate responses.

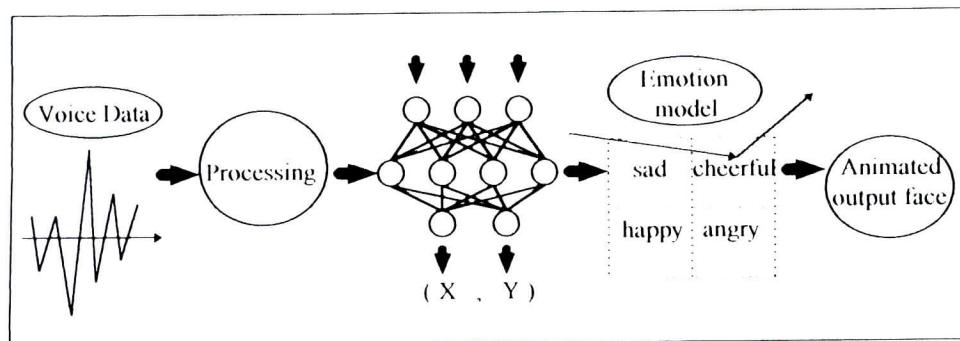


Figure 4.

4. Handshaking Device on Network Neuro-Baby

Handshaking Device (HSD) is an interface device which speakers can communicate with NB physically. The HSDs with NB are placed in Japan and USA so that people can communicate each other physically with HSDs through NB. The structure of HSD is shown in Figure 5. It can make force sensation to an operator and measure force pressure by pressure sensor. The HSD is assumed as a right hand and then the operator can feel existence of human through the force sensation generated by HSD. The other HSD is also grasped by other people. Those HSDs can send and receive force sensation through information network. The HSD is used as an input device to NB instead of Key-Board when the emotional model is customized.

Figure 6 shows system structure of HSD. The HSD is composed by 2 linear motors (AM20), and position sensor and force sensor are implemented to measure force from an operator. Those information are connected to a host computer via transputer mother board and i860 through 20Mbps link. The i860 is used to real time control. The host computer is connected to a host workstation (SS10) to communicate with the other HSD through information network (ATM and optical fiber).

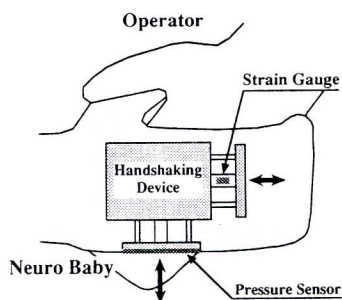


Figure 5. Handshaking Device.

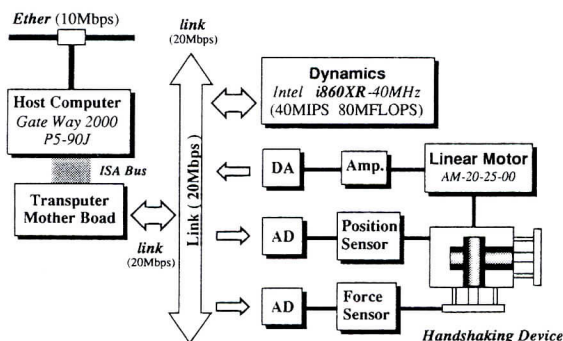


Figure 6. System of Handshaking Device.

5. Active Eye Sensing System

Active Eye Sensing System for NB is used to get information about speakers' face position so that NB can look at speakers. It means finding the most similar face among ones in the camera image by using template matching. In future it can recognize facial expression by understanding images.

Figure 7 shows the active eye sensing system. Each camera has two servo motors and then get two-degree-of-freedom such as yaw and pitch. The stereo camera system can identify the pose of moving object. The image is transformed into digital by Video Module and transferred to Tracking Module through VMEbus. In the tracking module three frames are stored and the motion between frames is estimated. These modules are controlled by VME master transputer which also calculate the pose of moving object. The pose information is used to control servo motors and transferred to NB through the host PC.

6. Ne

By ne
the co
many
applic
new in
techni
teleop
be on

7. App

A net
the cu
partner
custom
the NE
commu
user in

Active Eye Sensing System for Neuro Baby

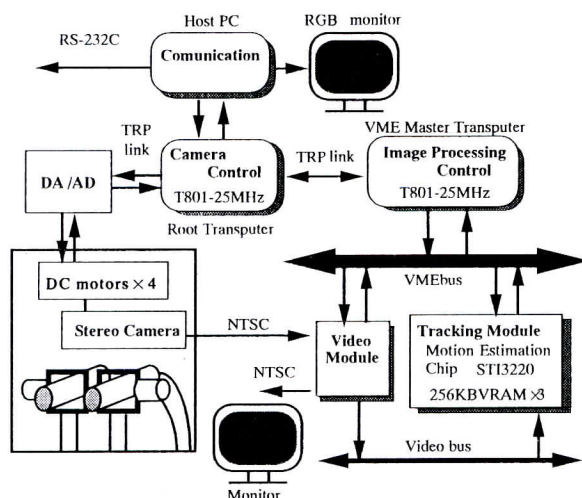


Figure 7.

6. Neuro-Baby as Network Sensitive Agent.

By networking, various new issues will happen. Since a network is subject to error and delay, the compensation of these effect is to be solved. For "conventional" media as image and voice, many techniques for it appeared in literature [3]. However, since networked NB is totally a new application, there exist no technique for delay or error compensation. Therefore, we developed new inter media and inter media synchronization technique suitable for the handshaking. This technique may also be used in general teleoperation systems, also a scaling technique is considered because the network might be one with long delay and severe packet loss.

7. Application: International Cultural Exchange

A networked NB can be used to help improving international cultural exchange and absorbing the cultural gap by customizing the NB at each site. Figure 8 shows two communication partners, one in Japan and one in the U.S., communicating via two NB's. The NB in Japan is customized for a Japanese user with appropriate recognition and expression mapping, whereas the NB in the U.S. is customized for its user with different typical mappings. The NBs communicate their emotional state over the network, which is then expressed to the individual user in both countries in an understandable form with customized expression mappings. A

communication setup like this can help reduce cultural differences between countries differing ways of communication and expression of feelings.

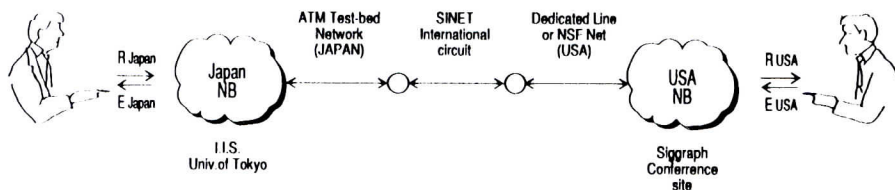


Figure 8. Network Based Neuro-Baby

8. Demonstration

As shown in Figure 8, we use ATM test-bed network from I.I.S. to the gateway of SINET international circuit which is located in Chiba, Japan. Then, SINET reaches to Stockton, CA. For the link between Stockton and Conference site, we will use either dedicated line or NSFnet.

The demonstration for SIGGRAPH'95 includes all above characteristics of the improved NB. Especially the demonstration of international cultural exchange will be exhibited. Communication will take place between the SIGGRAPH site (Los Angeles) and I.I.S. Univ. of Tokyo. Further it is to be expected that during the yearly I.I.S. Univ. of Tokyo open house event a large number of people will try to communicate with the NB.

Acknowledgment.

The authors would like to thank NACSIS staffs and Prof. S. Asano for arranging SINET international circuit. This work was partly supported by Grant-in-Aid for Creative Basic Research (Development of High-Performance Communication Network for Scientific Researchers). The first version of NB was developed in collaboration with N.Tosa and Fujitsu Lab. Special Thanks for ATR Media Integration & Communications Research Laboratories.

Reference

- [1] Tosa, N., Murakami, K., Kakimoto, M., Sato, S., "Neuro-Character" AAAI'94 Workshop: AI and A-Life and Entertainment
- [2] Philips, Sound Ideas, "Sound Effect Library LHH7910 Series," CD1001-CD1028.
- [3] See for example, Proc. 6th international workshop on Packet Video, Sep. 1994.