Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

Robotics and Autonomous Systems 46 (2004) 97–110

Robotics and Autonomous Systems

www.elsevier.com/locate/robot

Remote-neocortex control of robotic search and threat identification

Juan C. Macera [a,b,*], Philip H. Goodman [a,c], Frederick C. Harris Jr. [a,b], Rich Drewes [a,d], James B. Maciokas [a,e]

[a] *Brain Computation Laboratory, Mailstop 400, University of Nevada, Reno, NV 89557, USA* [1]
[b] *Department of Computer Science, Mailstop 171, University of Nevada, Reno, NV 89557, USA*
[c] *Department of Internal Medicine, Mailstop 355, University of Nevada, Reno, NV 89557, USA*
[d] *Program in Biomedical Engineering, Mailstop 256, University of Nevada, Reno, NV 89557, USA*
[e] *Department of Psychology, Mailstop 296, University of Nevada, Reno, NV 89557, USA*

Received 19 December 2002; received in revised form 20 September 2003

Abstract

Robots with remote processing capabilities would be useful in hazardous or complex environments presenting weight and cost constraints. We implemented a novel robotic system that incrementally triangulates and navigates towards a speaking target. This system comprises a distributed, biologically inspired, three-layer control system. High-level decision making is performed via Internet protocol by a pulse-coded neocortical simulator situated remotely in a secure location. The robot navigated towards and contacted an animated human mouth target in 75 of 80 trials ($\chi^2 = 20.3$, $P < 0.0001$). Current work involves classification and appropriate tactical response to acquired audio and visual speech information.

Keywords: Hierarchical robotic control system; Remote processing robot; Wireless robotic control; Sound localization; Neocortical simulator

1. Introduction

Robots typically have weight and power restrictions but voracious computing appetites. Wireless connections to powerful remote computing platforms combined with advances in software control make previously infeasible robotics problems realizable [14]. An interesting extension of this approach to robot control is a hierarchical command system that mimics biological brain function. Organic nervous systems have simple, local, high-speed motor and sensory circuits connected to more distant and slower (but more capable) brain structures for more complicated functions, and they call on the more powerful (but also slower and more remote) neocortex for the highest level of strategic decision making.

To test the hierarchical control concept, we set out to solve an example target threat identification problem, which we have broken down into two subtasks of different complexity.

1.1. Target localization

The system must be able to localize a target in the environment and navigate the robot towards it. The

* Corresponding author. Present address: Brain Computation Laboratory, Applied Research Facility 118, Mailstop 400, University of Nevada, Reno, NV 89557, USA. Fax: +1-775 784 1142.
E-mail address: macera@cs.unr.edu (J.C. Macera).
[1] http://brain.unr.edu.

0921-8890/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.robot.2003.11.005

target will emit continuous audio-speech signals, and the robot will localize it by binaural audio processing techniques.

1.2. Target threat assessment

Once it has reached the target, the robot must use further audio and visual information from the target, such as speech recognition with visual cues, to classify it as a threat or non-threat.

This experiment combines traditional robotic tasks, like movement and simple obstacle avoidance, with somewhat more computation-intensive tasks like target homing using audio cues along with the considerably more difficult (and more computation-intensive) challenges of speech and image analysis.

2. Method

A breakdown of the system's functions with their approximate neural correlates is depicted in Table 1. A sample experimental run is depicted in Fig. 1, and a short video clip demonstrating CARL (Cortex Accessed Remote Learning) in action is available online (http://www.cs.unr.edu/~macera/threatID.html).

Each of the three main components of the CARL system will be described in turn in the remainder of this section.

2.1. CARL robot

CARL is a remote-brained robot with three-level processing capability: on board, on a local PC/Laptop and on a remote computer cluster. The purpose of this

Table 1
CARL's three-level control system with biological correlates

	Body	Brainstem	Cortex
Functions performed at this control level	Movement, simple danger avoidance	Sound localization, preparing spike codes from audio and video data for input to neocortex simulation	Neocortical simulator software, speech recognition using audio and video (lip reading)
Location of this control function in CARL system	On board the mobile robot, called CARL	On nearby desktop class computer, called "Brainstem", connected to robot via wireless RF	On remote large-scale parallel computer, called 'Cortex', connected to Brainstem via the Internet

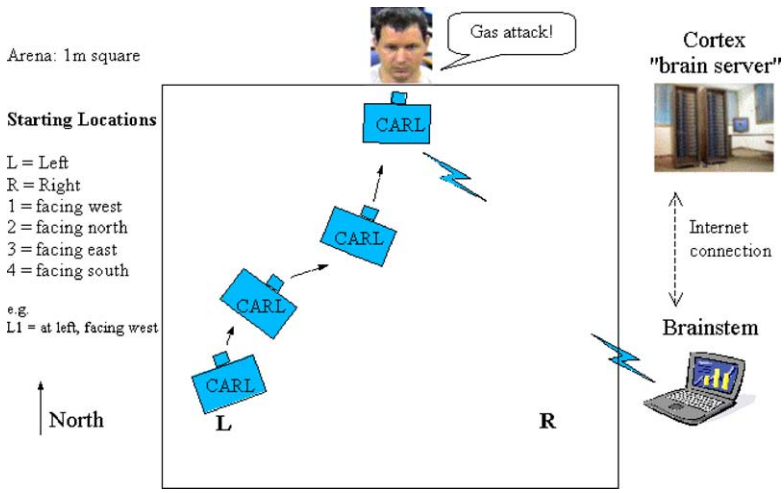

Fig. 1. CARL localizing a target.

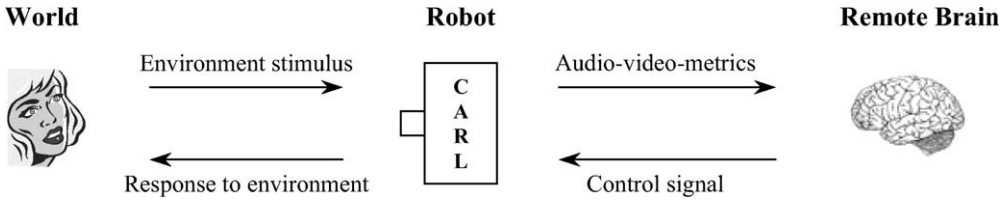

Fig. 2. Depiction of CARL's remote processing capability and its practical interaction with the environment.

processing architecture is to provide CARL with two main features: a limber and dynamic body that interacts its environment, and the potential of processing high-level AI techniques that usually requires massive computation. These features make CARL an excellent prototype for robotics and AI experimentation. Fig. 2 depicts this idea.

2.1.1. Onboard features

CARL by its own is a high-precision miniature programmable autonomous robot. It is a wheeled robot based on a dual processor architecture inspired by the biology of the nervous system. The secondary processor, a PIC16C71 RISC, is factory programmed to control speed and heading. This processor also communicates I/O and position information to the primary processor on board, a Parallax BS2-IC microcontroller, which allows sensor capture and navigation [3]. CARL's design offers high navigation accuracy, responsiveness and easy to command.

CARL features four true 8-bit A/D conversion ports, a speaker, four CdS light sensors and a thermistor to perceive temperature variations. Four ultra-bright LED's permit object proximity detection and status indication. Highly responsive bumper sensors detect objects around the entire front and sides of the robot. CARL has an efficient drive system that operates in near silence. Programmable PWM-controlled motors can deliver speeds from a crawl to more than 0.61 m/s. IR encoders monitor the motion of each wheel. The robot is approximately 0.18 m in diameter and 0.11 m tall overall.

Onboard, CARL contains a color video camera, stereophonic microphones to facilitate triangulation of sound, and a simple RF module (100 m span), that enables a wireless link to local computers. This hardware configuration is illustrated in Fig. 3.

2.1.2. Hierarchical processing and control capability

CARL has a unique remote processing architecture. Computing tasks and control commands have been distributed in a three-level fashion that correlates living creatures. *Reactive control*, which requires minimum computation, is performed on CARL's body. *Instinctive control*, which involves higher computation, is executed on a local PC. And, *decisive control*,

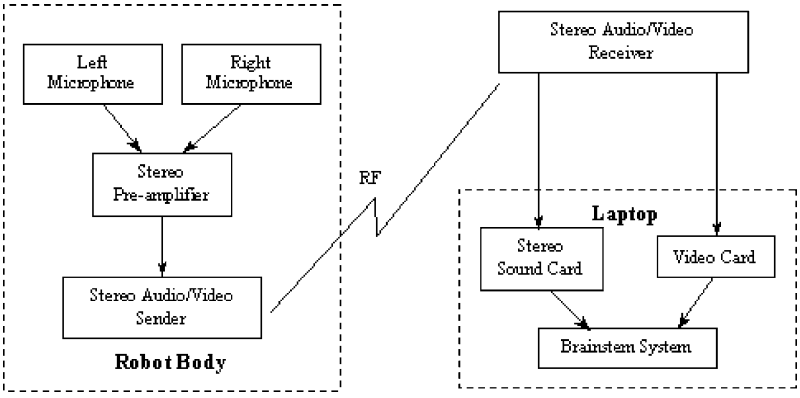

Fig. 3. Wireless audio–video hardware configuration.

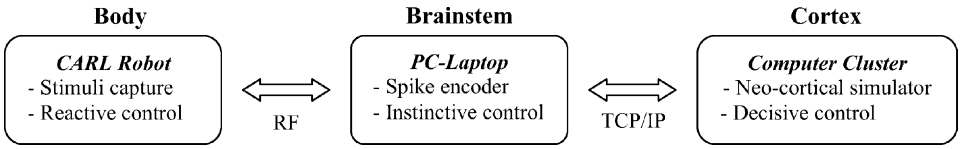

Fig. 4. CARL's three-layer control and processing architecture. Processing work is distributed according to its complexity. Control commands are defined remotely.

which demands massive computation, is performed on a remote computer cluster. This architecture is depicted in Fig. 4.

CARL connects to the local PC, dubbed 'Brainstem', using the RF link. Audio, video and metrics are gathered from CARL's sensors and passed on in real time to Brainstem for processing. Brainstem passes a preprocessed version of this data across the Internet to our remote located multiprocessor, named 'Cortex', which runs a program that simulate a biological neocortical network for perceptual analysis [30].

2.2. Brainstem functions

Brainstem resides on a nearby computer such as a laptop, connected to CARL via wireless RF link. The functionality of this system is three-fold: as a low-level decision and control system where computational tasks are not rigorous (for instance, sound localization); as a spike-encoded brain stimulus generator of auditory and visual signals for interpretation at the level of Cortex; and lastly as an interpreter of top-down signals from the remote Cortex, translating spike-encoded communications into instructions for CARL's navigation and tactical response systems.

2.2.1. Sound localization

Human beings localize a sound source primarily using cues derived from differences between the inputs received by the two ears. Two such cues play the dominant role in how humans estimate the horizontal position of a sound source: the interaural time difference (ITD) and the interaural intensity difference (IID) [8,15,16]. In the human brain, the IID function is performed by a brainstem structure called the lateral superior olive, and the ITD function is performed by a brainstem structure called the medial superior olive [23]. Correlation of these two structures are implemented in conventional software (i.e., in an algorithmic programming language) on our Brainstem computer.

In the localization sequence, sound from the target is received by CARL's two spatially separated microphones, and the data streams are passed on to Brainstem for processing. There the signals are digitized and the appropriate computations are performed on the Brainstem processor (an energy comparison [21] in the case of IID and a cross-correlation algorithm [2] in the case of ITD). The direction of the sound source is identified as either to the left side, right side, or approximately ahead, and the corresponding movement command is then radioed back to CARL. This sound localization technique is repeated between intervals of time for the next incremental move, and in this way CARL approaches the target step by step.

The methodology used for sound localization by ITD, which proved to be more resilient to noise and echo, is illustrated in Fig. 5. Stereo audio stream was captured at 16 000 Hz of frequency in a window of 0.02 s, every 0.5 s approximately for cross-correlation analysis and navigation decision.

The sequence ends when CARL's front bumper contacts the target. This initiates capture and RF transfer of audio-visual signals to Brainstem and Cortex, where threat assessment is performed. For this phase, Brainstem still serves as the intermediary for CARL's communication with the neocortical simulator. Brainstem is also responsible for converting the audio and video data received from CARL into spike-coded input suitably formatted for the neocortical simulator responsible for the deeper analysis.

2.2.2. Spike encoding of audio and visual data

Another function of Brainstem is to prepare, in real time, the spiking data for neural network simulation. This is accomplished as follows. CARL's audio data is processed using short time Fourier transforms. The result is a set of energies for 129 frequency bands. The energy in each frequency band is converted into

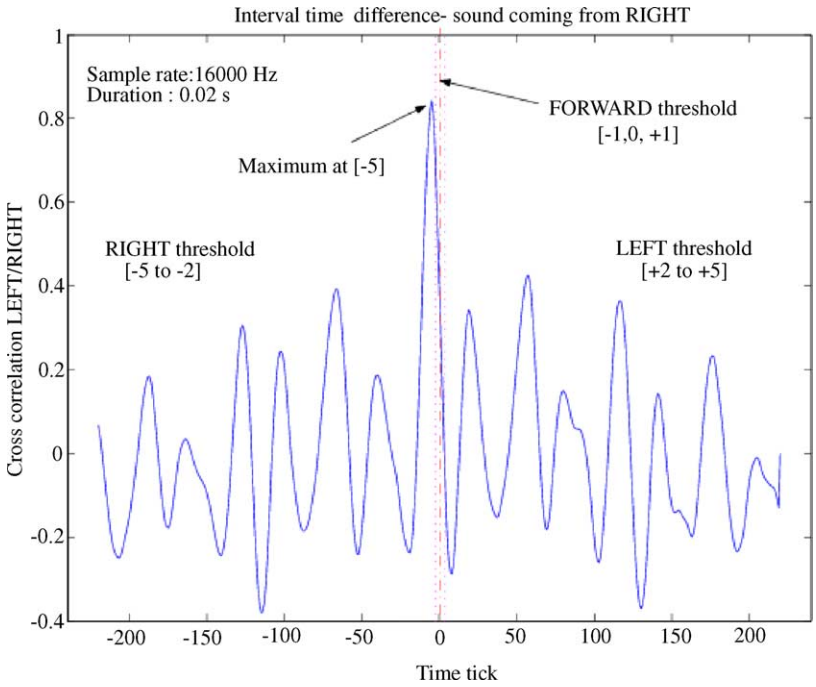

Fig. 5. Sound localization methodology by cross-correlation of binaural information.

a spike probability for a section of cells in the simulated auditory cortex. CARL's video data is processed by a Gabor filter on Brainstem, which results in spike probabilities for subregions of the image for different frequency ranges considered at different orientations [19]. These spike frequencies are sent to the neocortical simulator (NCS), on Cortex, where the data is input to groups of cells in the visual cortex of the simulated neocortex in retinotopic (spatially mapped) fashion analogous to human visual processing.

Fig. 6(a) is an example image of the target taken by CARL's own camera at the end of a successful localization run. Fig. 6(b) represents the output of a Gabor analysis of the image, which is subsequently converted into probability values for a region of the simulated visual cortex and then forwarded to Cortex

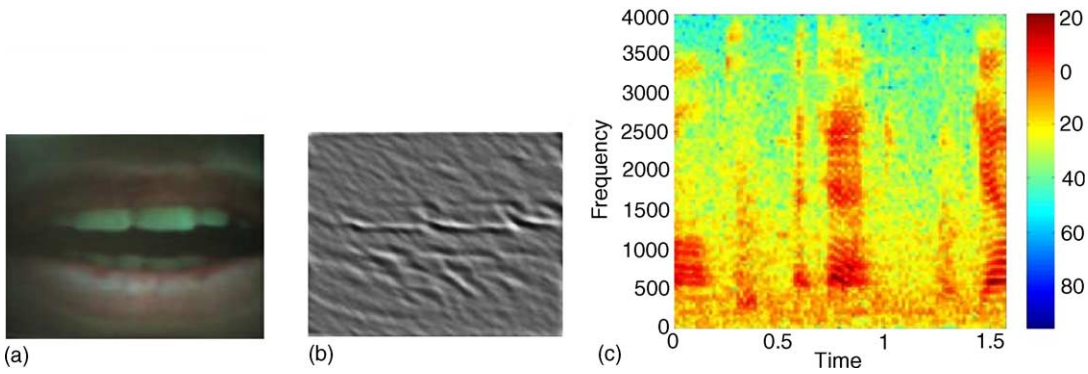

Fig. 6. Preprocessing of visual and audio data. (a) Video frame of the speaking mouth target; (b) Gabor analysis of the mouth frame; (c) spectrogram of the speech captured.

for processing. Fig. 6(c) is the STFT output, which is subsequently converted into probabilities by Brainstem for spike encoding input to the auditory cortex simulated on Cortex.

2.3. Cortex functions

Cortex is a powerful cluster of computers linked to the Brainstem system via TCP/IP protocol. Currently Cortex consist of 128 processors with a total of 256 GB of RAM and more than one Terabyte of disk storage, interconnected with a Myrinet 2000 high speed/low-latency interconnection network. The goal of Cortex is to deal with CARL's high-level decision and control tasks that require intensive computation, such as bimodal (audio-video) or multimodal (i.e., multisensory) processing. Data input or stimulus to Cortex is received from Brainstem, and the results are sent back to Brainstem for robot control.

2.3.1. Cortex: a generalized 'brain server'

At present, Cortex runs a biologically realistic spike-coded neural network, named NCS (neocortical simulator). In brief, the neocortical model employed here simulates several columns worth of a small visual and auditory cortex with thousands of cells and millions of synapse connections. NCS has the capability of simulating a chunk of spiking neocortex of thousands of cells with millions of synapses with a high degree of biological accuracy, including features such as ion channels, synaptic facilitation and depression, and Hebbian learning [12,29–31].

NCS was implemented in C language using MPI parallel programming library. A neocortical simulation session can be created de novo with each new set of data from CARL, according to a description in a configuration file, or, more commonly, a previously trained brain can be loaded from disk for that simulation run. For this work, NCS was trained in advance to recognize a number of phrases, some of which were classified as threatening and some as neutral. The design and training of NCS is detailed in Section 3.

The cluster performing the simulation can serve any number of simultaneous requests for simulations, either processing each request to completion in the order it was received or starting each request immediately as it arrives, even if other simulations are ongoing. In this way the cluster can function as a 'brain server' for multiple robots in the field.

2.3.2. Bimodal (sound and video) speech recognition

This section describes a particular high-level decision process performed on Cortex by NCS. Once CARL reaches the target, Brainstem begins to pass the audio and video data emitted by the target to the remote neocortex portion of the control system for threat analysis. The audio data captured by CARL's microphone consists of one of several spoken phrases emitted by the target for that test run, and the video captured by CARL's camera consists of short moving images, played on a small video screen embedded in the target, of human mouths articulating the phrase as it is spoken.

Bimodal recognition was chosen in order to support and link a related research project exploring the potential for increased phrase recognition accuracy when combining lip position information with sound information into a bimodal neocortical speech recognition strategy. Details of this learning algorithm are available in a technical report from our lab [19], and a summary is described in Section 3.

To reduce data transit demands, the audio and video information is converted into a neural spike-coded representation on Brainstem, before being passed to the neocortex simulator over a conventional computer network, usually the Internet. Since the neocortical processing takes from several seconds to minutes, depending on the complexity of the neocortical model installed, the network latencies introduced in communication to the remote neocortex are acceptable for our purposes.

3. Design and training of NCS for bimodal recognition

Our network was designed to processes bimodal, audio and visual, information. We decided on audio and visual information because these are the two sensory modalities that we as humans rely on the most and which we, the scientific community, have the best understanding. Next we designed our network to incorporate the current understanding of biological processing while keeping with reasonable simulation times. Both the transformation of audio/visual

information and the design of the network are discussed in more detail to follow.

3.1. Data acquisition and spike encoding

Audio–video-interleave (.avi) movies were recorded from 10 volunteers speaking the following three sentences: “Attack with gas bombs”, “He is a loyal citizen”, and “I’m not entirely sure”. Each .avi was recorded at 25 frames per-second with audio digitization of 11 kHz. Recordings were truncated to 1.6 s of audio and 40 frames of video to keep the sentences the same length. Auditory signals were processed using a short-time-Fourier-transform (STFT). STFT decomposes the auditory signal into 129 frequency bands and provides the power of each frequency as a function of time (Fig. 7).

By moving a narrow window (2.5 ms) independently for each frequency across time a probability of spiking is computed from the power within each window (normalized to the maximum power across all windows of all frequencies). In actuality the tonotopic representation of the cochlea is closer to a logarithmic scale, and the Fourier transform is a linear manipulation. In order to minimize the difference between cochlear processing and the STFT, a larger proportion of cells encoded lower frequencies than higher frequencies. Our auditory cortex included three columns. The first column received the first 20 frequency bands, the second column received the next 40 frequency bands, and the final column received the remaining 69 frequency bands.

Visual signals were first whitened and then processed using Gabor analysis. The receptive field properties of primary visual cortex (VI) simple cells resemble Gabor-like properties [28], minimizing the tradeoff between frequency information and spatial information. Fig. 8 shows two frames of an .avi movie before and after Gabor-filtering using horizontally oriented high and low band-pass filters. In order to preserve the retinotopic mapping, the filtered image was broken down into 5×5 subregions. The average intensity within a subregion was used as the probability of spiking for a group of cells encoding that position.

3.2. Network design

Our network was made up of 10 columns (6 visual, 3 auditory and 1 association). Each primary sensory column comprised two layers, an input (IV) and output layer (II/III). Layer IV included 300 excitatory cells. Layer II/III included 300 excitatory and 75 inhibitory cells. Layer IV excitatory cells connected to layer II/III excitatory cells with a 10% probability. Layer II/III excitatory cells connected with each other and to inhibitory cells with a 10% probability. Inhibitory cells connected to excitatory cells within layer II/III with a 5% probability. The association column was made up of one input layer (IV) similar to the output layers of the primary sensory columns. The excitatory cells of layer II/III for the six visual and three auditory columns each connected with layer IV of the association column using a 1% probability. Simulations

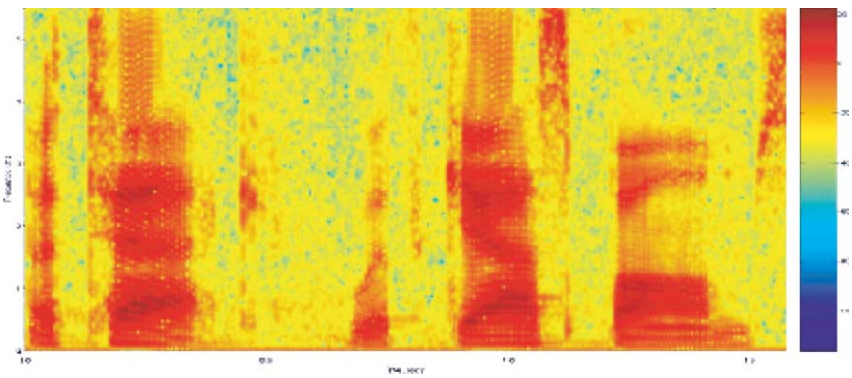

Fig. 7. Spectrogram of the spoken sentence “attack with gas bombs”. Vertical axis-auditory frequency (0–5.5 kHz in 129 bands). Horizontal axis: time in seconds (1.6 s). Pseudocolor legend: signal power in dB [−120, 25].

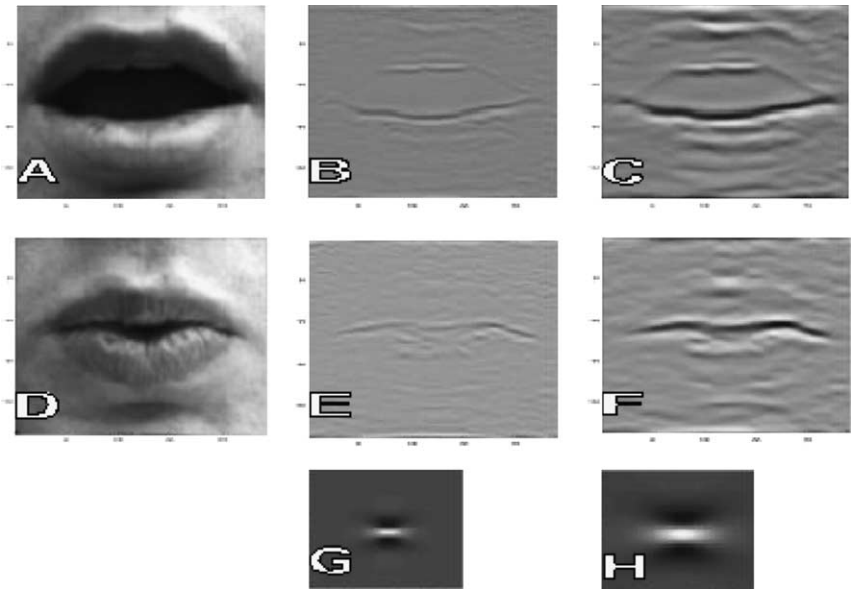

Fig. 8. Two frames (240 × 240) from an .avi movie before and after horizontal Gabor-filtering. (a and d) Before filtering; (b and e) after filtering with high band-pass filter; (c and e) after filtering with low band-pass filter; (g) high band-pass filter (30 × 30); (h) low band-pass filter (30 × 30).

typically took approximately 3–5 min to process a 3-s recording.

3.2.1. Cell design

Our cell models consisted of a single integrate-and-fire compartment. Resting potential and threshold were −65 and −40 mV, respectively. Once threshold was reached, a spike template consistent with literature-based measurements of chord-width, peak, rise and fall rates was applied for the following 2 ms. For pyramidal cells the most common firing pattern to sustained supra-threshold current is a repetitive firing rate with little to no decrease in frequency (RS1) [7]. In addition, a similar firing pattern occurs for several types of inhibitory cells [11]. For reasons of simplicity 80% of our pyramidal and inhibitory cells shared the same compartment design for regular spiking behavior. The other 20% of our pyramidal and inhibitory cells comprised adapting/accommodating firing patterns to sustained current. Again the inhibitory and pyramidal shared the same compartment design.

3.2.2. Channel design

Although our cell model used a spike template once threshold was reached, several sub-threshold responsive channels were used to achieve the desired repetitive firing pattern. Recent evidence has shown that AP repolarization is not Ca^{2+} dependent [22] and that voltage-gated K^+ channels, not Ca^{2+} activated K^+ channels, are largely responsible for AP repolarization [1,10,18]. In addition, A-type channels have been linked to neuronal repetitive firing without accommodation [6,25,27]. In light of this evidence we used only an A-type channel to model repetitive spiking behavior. Models of M-type channels [32] and SK channels [9] have been shown to contribute to spike frequency adaptation and therefore were included to achieve the desired firing behavior. Previous reports have shown that several after hyperpolarizations occur following an AP: fast (IC) calcium dependent, intermediate non-calcium dependent, and slow (IAHP) calcium dependent currents [27]. Therefore, we included a fast Ca^{2+} dependent K^+ channel, an M-type K^+ channel, and a slow Ca^{2+} dependent K^+ channel into our cell model. These channels were included in the ratio 1:1:2, respectively.

3.3. Network training

Learning and training were designed to take advantage of the synaptic properties observed in

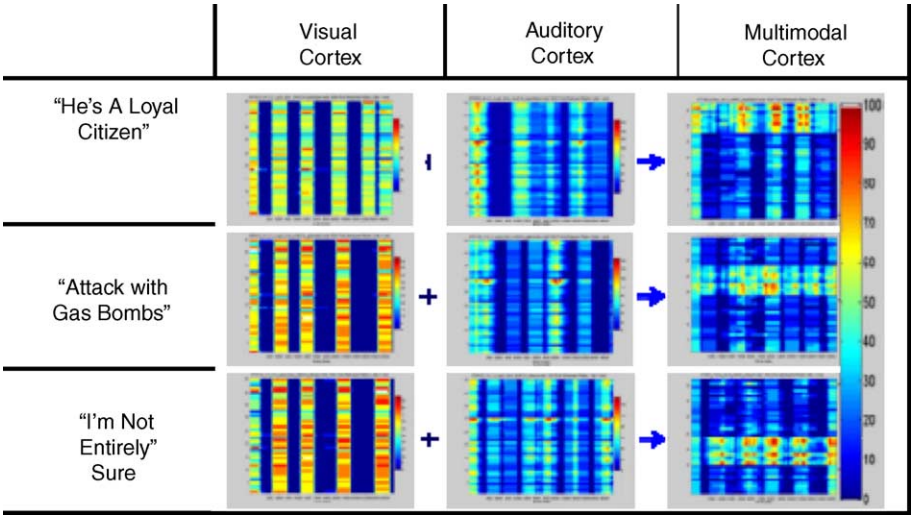

Fig. 9. Pseudocolor windowed spike rate plots in response to spoken sentences. Vertical axis: windowed spike frequency. Horizontal axis: time (each frame shown here corresponds to 1.6 s, the time for the complete sentence to be spoken once; sentences were repeated a total of seven times).

neocortical tissue. Both short-term transient and long-term Hebbian-like synaptic changes were modeled. In order to mimic the feedback projections of the frontal cortex training was accomplished by selectively injecting a unique subset of cells with current for each sentence presented to the network. Both the synaptic properties and training are discussed in more detail below.

3.3.1. Synaptic dynamics

Our synapse model included reversal potential, conductance, A (absolute strength, or product of quantal size and number of release sites), U (mean probability of release), D and F (the time constants to recover from depression and facilitation, respectively). Details of parametric equations are completely characterized in [20,26]. F1 synapses predominately facilitate (F:D, 9.04 ± 1.85), F2 synapses depress (D:F, 40.6 ± 4.6), and F3 synapses are mixed (2.82 ± 4.6); further details can be found in [11].

3.3.2. Learning

Spike-coded visual and auditory representations in primary sensory cortices demonstrated unique patterns for the three sentences (Fig. 9). When output layers of these primary cortices interacted in multimodal association cortex, there was again preservation of unique spiking patterns (Fig. 9, fourth column).

The first column in Fig. 9 shows three sentences modified from the TIMIT corpus. Columns two and three show the spiking response of neurons driven from the visual and auditory transformations. The fourth column is the response of associative multimodal cortex during reward depolarization of selected

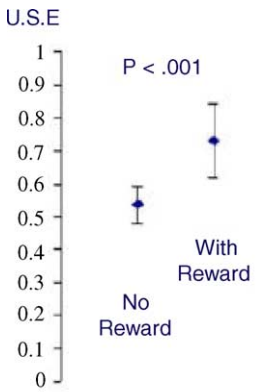

Fig. 10. Utilization of synaptic efficacy. Mean U.S.E. (±1S.D.) after seven presentations of spoken sentence "attack with gas bombs" among excitatory neurons in multimodal association cortex, unrewarded neurons (left) vs. rewarded neurons (right) during training for that sentence.

neurons for each sentence. Fig. 10 shows the change in synaptic strength (USE) after successive sentence presentations for the rewarded versus non-rewarded neurons. Rewarded neurons where given direct current injection during sentence presentation to bring their membrane potential closer to threshold.

4. Results

The integration of three dissimilar computing systems, with reliable linkage and synchronization, is the first achievement of this project. Our system responded effectively to different levels of computation complexity. On CARL the computation was kept simple. On Brainstem, the computation complexity was characterized by the cross-correlation, FFT and Gabor analysis. These are used for sound localization, auditory encoding and visual encoding, respectively. On Cortex, the spiking neural network algorithm defined the computation complexity. This is described in Section 3.2. The computation power of each system layer is described in Table 2.

With respect to data transmission characteristics, in theory, Brainstem is able to receive audio–video information from CARL at 12 Mbps, and exchange control commands and metrics at 9.6 kbps. Data stream between Brainstem and Cortex, over Ethernet link, rated 100 kbps. And, message passing between Cortex nodes at 2 Gbps.

Our remote-brained experiment performed as follows. First, sound localization via IID and ITD was assessed with CARL immobile and a moving sound source. These results are reported in Section 4.1. Next, CARL's ability to navigate toward a speaking target using one of these techniques (ITD) is reported in Section 4.2. We then briefly describe in Section 4.3 the communication that takes place between Brainstem and Cortex once CARL contacts the target and acquires the bimodal data for target threat assessment.

Table 3
Results of 80 localization experiments

Experiment no.	Left (CARL orientation)				Right (CARL orientation)			
	L1	L2	L3	L4	R1	R2	R3	R4
1	Ok	Ok	Ok	Ok	Ok	Ok	Ok	Ok
2	Ok	Ok	Ok	Ok	Ok	Ok	Fail	Ok
3	Ok	Fail	Ok	Ok	Ok	Ok	Ok	Ok
4	Ok	Ok	Ok	Ok	Ok	Ok	Ok	Ok
5	Ok	Ok	Ok	Ok	Ok	Ok	Ok	Ok
6	Ok	Ok	Fail	Ok	Ok	Ok	Ok	Ok
7	Ok	Ok	Ok	Ok	Ok	Fail	Ok	Ok
8	Ok	Ok	Ok	Ok	Ok	Ok	Ok	Ok
9	Ok	Ok	Ok	Fail	Ok	Ok	Ok	Ok
10	Ok	Ok	Ok	Ok	Ok	Ok	Ok	Ok

4.1. Sound localization using IID and ITD

Results of sound localization for IID are reported in Fig. 11. Comparable results for ITD are reported in Fig. 12. Although ITD and IID are generally complementary techniques for estimating a sound direction, we found ITD to be considerably more robust and less subject to calibration errors and errors due to noise or echoes. Therefore, the next experiment was conducted using ITD alone.

4.2. Navigation toward target

Ten trials of navigation toward target were performed from each starting location (left and right) in each of the four starting orientations (see Fig. 1), for a total of 80 trials (Table 3, Section 2.2.1). A trial consisted of repetitive cycles of sound localization and navigation. Each trial comprised multiple individual left/right/center ITD computations, resulting in an incremental rotation, or movement toward the target if the ITD orientation remained unchanged. A trial was considered successful if CARL's bumper made contact with the target and the middle 80% of the imaged

Table 2
Three-layer control system hardware comparison

	Processor	Processor speed	Memory
CARL	BS2-IC	20 MHz ~400 inst./s	2 K EEPROM 64B RAM
Brainstem	Pentium 4	2.2 GHz	512MB RAM
Cortex	Xeon 2.2 128 nodes	2.2 GHz per node	2GB RAM per node

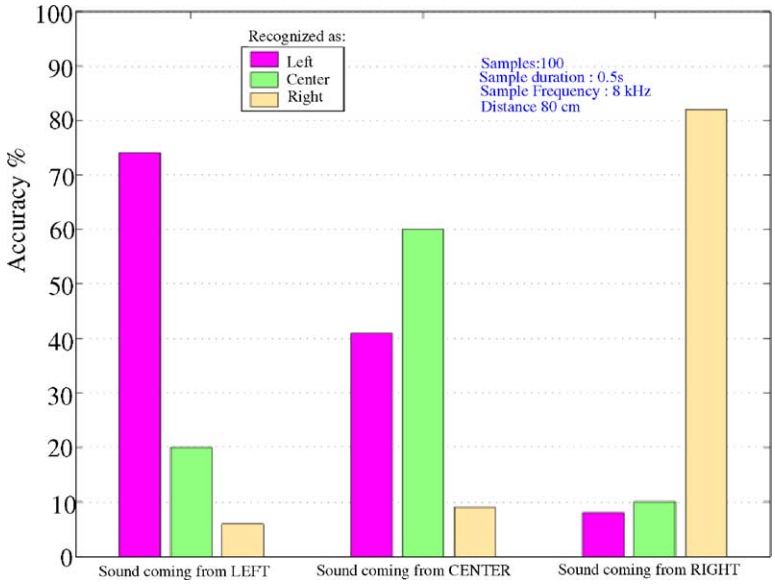

Fig. 11. Localization accuracy using IID technique.

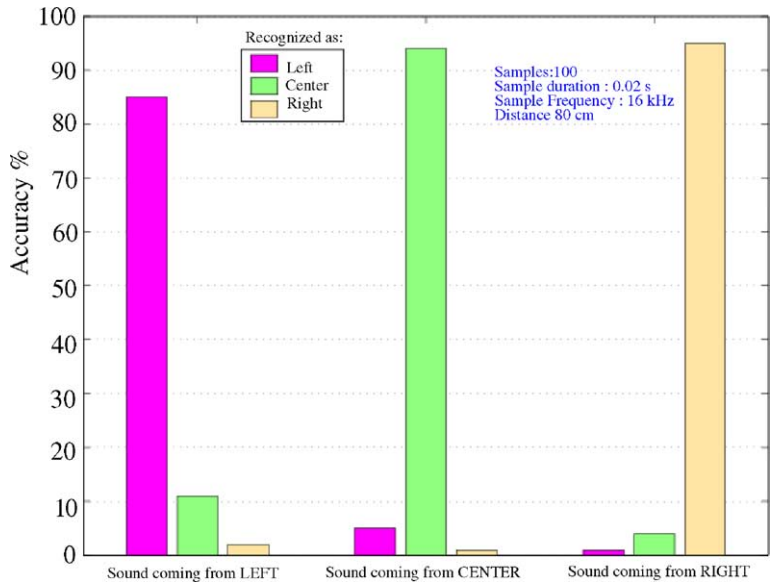

Fig. 12. Localization accuracy using ITD technique.

lip was visible from CARL's onboard camera. CARL successfully navigated toward and contacted the target mouth region in 75 of 80 trials ($\chi^2 = 20.3$, $P < 0.0001$, based on the number of possible endings along the edge of a meter square table surface). Each navigation experiment took between 25 to 30 s.

4.3. Threat assessment of target using bimodal neocortical speech recognition

After successful localization, 1.6 s of audio and 23 frames of video are captured and Brainstem codes the audio and video data into spike probabilities. This data

encoding took about 3.5 s on Brainstem during simulation. Next, a TCP connection is established between Brainstem and Cortex, and the spike-encoded data is transferred to Cortex. The NCS program is invoked on the data to perform the threat assessment. The initial brain state is typically loaded from a previously trained and saved brain. The interpreted result of the neocortical simulation is passed back to Brainstem, which then instructs CARL to take the appropriate action. The details of the bimodal speech recognition are described in [17].

5. Discussion

We implemented a wheeled robot with remote and hierarchical processing capability that can explore its environment from any initial orientation, detect the presence of a potential target, and reliably approach the target using sound localization techniques in order to perform bimodal (audio plus visual lip reading) speech recognition by invoking a biologically accurate remote neocortical simulation.

Although the concept of remote-brained robotics has been explored previously by Inaba et al. [14], in that work the brain and body were separated, both conceptually and physically. Our system is novel in that it incorporates three-level hierarchical processing intended to model the efficiency of human neurological perceptual processing and decision making. In this configuration, task selection and allocation are relevant and contribute to effective robot responsiveness.

Our system is also notable for its ability to map many-to-many robots and 'cortices' via a distributed communication network (here, the Internet). Each CARL could potentially communicate with many Cortex-like clusters globally distributed. In turn, each Cortex could simultaneously control (hence coordinate) many CARL robots. This would yield not only flexible distribution of computational powerful across a dynamic problem-solving environment, but also redundancy that could sustain the system in the event of focal destructive events.

Although the utilization of physical robots and real-world environments is important in order to constrain and supply boundary and noise conditions, our system can also function in a virtual reality mode, in which the robot and environment are emulated. This could be valuable for early stage projects such as extraterrestrial robotic landing or hostile terrain exploration.

From a cognitive science perspective, our remote-brained robotic system's massive parallel processing and its embodiment of perceptual decisions, make our system a valuable platform for investigating new types of artificial intelligence such as applied neurocomputing and evolutionary agents [24], where the active and strong relationship between the brain, body and environment is fundamental for model development [4].

Many algorithms for path planning (choosing a safe course through a set of obstacles) have been described in the computer science literature [5]. Few of these algorithms are biologically inspired, and even fewer effectively utilize spiking neural network techniques. We recently extended our work to include three responses contingent on cortical interpretation: uninteresting target (repositions and resumes searching), immediate threat to robot integrity (rapidly backs away to escape), and target identified for subsequent action or continued monitoring (alarms). Our work could be further extended to simulate the neocortical functions of human path planning. It should be noted that while CARL performs well in low noise and echoic environments tested to date, real-world implementation will require the ability to separate multiple sources of audio and video signals [13] (the problem of 'event formation').

Acknowledgements

This work was supported in part by US Office of Naval Research under grants N00014-00-1-0420 and N00014-01-1-0552. We would like to thank James Frye and Jason Baurick for their assistance with the hardware and software used in this project, and Dr. Dwight Egbert for his assistance with robotic design and control.

References

[1] J.L. Albert, J.M. Nerbonne, Calcium-independent depolarization-activated potassium currents in superior colliculus—projecting rat visual cortical neurons, Journal of Neurophysiology 73 (6) (1995) 2163–2178.

[2] P. Bourke, Cross-correlation, Swinburne University of Technology. http://astronomy.swin.edu.au/~pbourke/analysis/correlate, Australia, 1996.

[3] N.C. Braga, Robotics, Mechatronics and Artificial Intelligence, Newnes, Boston, MA, 2002.
[4] H.J. Chiel, R.D. Beer, The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment, Trends in Neuroscience, Elsevier, Amsterdam, 1997, pp. 553–557.
[5] H. Choset, M. La Civita, J. Park, Path planning between two points for the robot experiencing localization error, in: Proceedings of the International Conference on Field and Service Robotics, August 1999, pp. 98–103.
[6] J.A. Connor, C.F. Stevens, Prediction of repetitive firing behaviour from voltage clamp data on an isolated neurone soma, Journal of Physiology 213 (1) (1971) 31–53.
[7] B.W. Connors, M.J. Gutnick, D.A. Prince, Electrophysiological properties of neocortical neurons in vitro, Journal of Neurophysiology 48 (6) (1982) 1302–1320.
[8] R.O. Duda, Sound localization research, San Jose State University, USA, visited 23 July 2002. http://www-engr.sjsu.edu/~duda/Duda.Research.html.
[9] J. Engel, H.A. Schultens, D. Schild, Small conductance potassium channels cause an activity-dependent spike frequency adaptation and make the transfer function of neurons logarithmic, Biophysical Journal 76 (3) (1999) 1310–1319.
[10] R.C. Foehring, D.J. Surmeier, Voltage-gated potassium currents in acutely dissociated rat cortical neurons, Journal of Neurophysiology 70 (1) (1993) 51–63.
[11] A. Gupta, Y. Wang, H. Markram, Organizing principles for a diversity of GABAergic interneurons and synapses in the neocortex, Science 287 (5451) (2000) 273–278.
[12] F.C. Harris Jr., J. Baurick, J. Frye, J.G. King, M.C. Ballew, P.H. Goodman, R. Drewes, A novel parallel hardware and software solution for a large-scale biologically realistic cortical simulation, Brain Computation Laboratory, University of Nevada, Reno, NV, 2002.
[13] J.J. Hopfield, C.D. Brody, 2000. What is a moment? "Cortical" sensory integration over a brief interval, in: Proceedings of the PNAS'97, pp. 13919–13924.
[14] M. Inaba, et al., A platform for robotics research based on the remote-brained robot approach, International Journal of Robotics and Research 19 (10) (2000) 933–954.
[15] R.E. Irie, Robust sound localization: an application of an auditory perception system for a humanoid robot, MS Thesis, Massachusetts Institute of Technology, MA, 1995.
[16] D.J. Klein, et al., A neuromorphic approach to the analysis of monaural and binaural auditory signals, University of Maryland, USA, visited 7 August 2002. http://www.isr.umd.edu/CAAR/papers/ewns99.pdf.
[17] C. Koch, I. Segev, Methods of Neuronal Modeling, 2nd ed. MIT Press, Cambridge, MA, 1998.
[18] R.E. Locke, J.M. Nerbonne, Role of voltage-gated K^+ currents in mediating the regular-spiking henotype of callosal-projecting rat visual cortical neurons, Journal of Neurophysiology 78 (5) (1997) 2321–2335.
[19] J. Maciokas, P. Goodman, F. Harris, Large-scale spike-timing-dependent-plasticity model of bimodal (audio/visual) processing, Brain Computation Laboratory, University of Nevada, Reno, NV, 2002.
[20] H. Markram, et al., Potential for multiple mechanisms, phenomena and algorithms for synaptic plasticity at single synapses, Neuropharmacology 37 (4–5) (1998) 489–500.
[21] J.H. McClellan, R.W. Schafer, M.A. Yoder, Digital Signal Processing First: A Multimedia Approach, Prentice-Hall, NJ, 1998.
[22] J.C. Pineda, R.S. Waters, R.C. Foehring, Specificity in the interaction of HVA Ca^{2+} channel types with Ca^{2+}-dependent AHPs and firing behavior in neocortical pyramidal neurons, Journal of Neurophysiology 79 (5) (1998) 2522–2534.
[23] D. Purve, et al., Neuroscience, Sinauer Associates Inc., Sunderland, MA, 1997.
[24] E. Ruppin, Evolutionary autonomous agents: a neuroscience perspective, Nature Review of Neuroscience 3 (2) (2002) 132–141.
[25] P.C. Schwindt, et al., Multiple potassium conductances and their functions in neurons from cat sensorimotor cortex in vitro, Journal of Neurophysiology 59 (2) (1988) 424–449.
[26] W. Senn, H. Markram, M. Tsodyks, An algorithm for modifying neurotransmitter release probability based on pre- and postsynaptic spike timing, Neural Computation 13 (1) (2001) 35–67.
[27] J.F. Storm, Action potential repolarization and a fast after-hyperpolarization in rat hippocampal pyramidal cells, Journal of Physiology 385 (1987) 733–759.
[28] M.A. Webster, R.L. De Valois, Relationship between spatial-frequency and orientation tuning of striate-cortex cells, Journal of the Optical Society of America A 2 (7) (1985) 1124–1132.
[29] E.C. Wilson, Parallel implementation of a large scale biologically realistic neocortical neural network simulator, M.S. Thesis, University of Nevada, Reno, NV, August 2001.
[30] E.C. Wilson, P.H. Goodman, F.C. Harris Jr., Implementation of a biologically realistic parallel neocortical-neural network simulator, in: Micheal Heath, et al. (Eds.), Proceedings of the 10th SIAM Conference on Parallel Process. For Sci. Comput., Portsmouth, VA, March 2001.
[31] E.C. Wilson, F.C. Harris Jr., P.H. Goodman, A large scale biologically realistic cortical simulator, in: Charles Slocomb, et al. (Eds.), Proceedings of the SC 2001, Denver, CO, November 2001.
[32] W.M. Yamada, C. Koch, P.R. Adams, Multiple channels and calcium dynamics, in: C. Koch, I. Segev (Eds.), Methods of Neuronal Modeling: From Ions to Networks, MIT Press, Cambridge, London, 1998, pp. 137–170.

Juan C. Macera is a research associate at the Brain Computation Laboratory in the robotics field. He received the B.S. degree in Mechanical Engineering at UNI, Lima, and the M.S. degree in Computation at UMIST, Manchester. Currently, he is pursuing a doctoral degree in Computer Engineering at the University of Nevada, Reno. His research interests are in artificial intelligence, evolutionary robotics, cognitive neuro-robotics, and spatial auditory and visual processing.

Philip H. Goodman is a Professor in the Department of Internal Medicine and Program in Biomedical Engineering, and Adjunct Professor in the Department of Computer Science at the University of Nevada, Reno. He received his B.A. in Physics, B.S. in Biology, and M.D. from the University of California, Irvine, and his M.S. from the University of Michigan. As director of the Brain Computation Laboratory, he focuses on understanding the principles of mammalian neocortical processing using large-scale computer simulations, with applications in robotics, speech perception, and artificial intelligence.

Frederick C. Harris Jr. received the B.S. in Mathematics and M.S. in Educational Administration from Bob Jones University and the M.S. and Ph.D. in Computer Science from Clemson University. Since then he has been at the University of Nevada, Reno, where he is currently an Associate Professor in the Computer Science Department. His research interests are in Parallel and Distributed Computing as well as applications of Virtual Reality.

Rich Drewes received his B.S. in Electrical Engineering and Computer Science from the University of California, Berkeley, in 1990. He is currently enrolled in the biomedical engineering program at the University of Nevada, Reno. His research interests include techniques for computer simulation of organic neural nets and other approaches to artificial intelligence.

James B. Maciokas received his B.A. in psychology from Western Oregon University and his M.A. and Ph.D. in psychology from the University of Nevada. He is currently a postdoc at the Center for Neuroscience at the University of California, Davis. His research interests include investigating visual and attentional mechanisms by combining experimentation with computational modeling.