NASA-CR-194712

(NASA-CR-194712) STUDIES IN ARCHITECTURE FOR AI APPLICATIONS Final Report, 1 Jul. 1988 - 30 Jun. 1989 (Stanford Univ.) 10 p

N94-71221

Unclas

Z9/63 0201692

Studies in Architecture for AI Applications

FINAL REPORT
For Period Covering 07/01/88 to 06/30/89

NASA Contract No. NCA2-388
Stanford Project No. 6336

January 31, 1990

Prepared by:

Michael J. Flynn
Principal Investigator
Department of Electrical Engineering
Stanford University
Stanford, California 94305-4055

Prepared for:

NASA-Ames Research Center
Information Sciences Division
Mail Stop 244-7
Moffett Field, California 94305
Attention: Dr. Henry Lum

Funds for the support of this study have been allocated by the NASA-Ames Research Center, Moffett Field, California, under Interchange No. NCA2-338

APPROVED:

Michael J. Flynn, Project Director
Professor of Electrical Engineering
Computer Systems Laboratory

NEURAL NETWORKS AND PARALLEL COMPUTATION

Recent work in neural networks has proceeded along two lines, distinguished by their goals. The first attempts to use neural networks to solve problems typically associated with computer science, e.g. optimization problems and classification problems for pattern recognition. For this effort, neural networks are viewed as collections of interconnected, "neuron" like elements arranged in one or more layers. The interconnections are mediated by weights that follow an adaptation law. Certain adaptation laws require the presence of only inputs to the network, while others require the paired presentation of inputs and outputs. In the former case, the network tends to discover the clusters of characteristics present in the input space. In the latter case, the network learns to associate classes of input with the particular outputs that were used during training.

The second stream of research attempts to design and build circuits that mimic, in limited ways, the signal processing behavior of biological systems. The emphasis here is on signal processing because most of what is known about the structure of biological organisms is concentrated at the periphery, that is, the sense organs, the brainstem, and the first projection to the sensory cortex areas of the brain. Models of the retina and cochlea are examples of such circuits that have been built using analog MOS technology, described in Mead(1).

Both streams of research utilize parallel computation as a fundamental paradigm. However, parallel computation alone does not qualify a system as "neural". For example, a matrix multiplication run on a parallel processing machine is not a neural computation. A true neural computation must posses one or more of the following characteristics :

(a). It must be collective or statistical in nature, and thus robust against individual component failures.

(b). It must be noise resistant, and therefore insensitive to small variations in the input, as commonly encountered with real world signals.

(c). It must be self adjusting, continuing to operate over the wide dynamic range that real world signal typically cover, without "crashing" due to register overflows, insufficient resolution, or other similar hardware problems.

"Learning" is not included as a criterion. Many computations performed by biological systems are hardwired into the system, not learned through experience. This is particularly true of the computations performed at the periphery. What is learned is the interpretation of the peripheral signals, performed at higher cognitive levels.

The remainder of this piece describes a signal processing system using parallel, neural computations. Its goal is to emulate certain aspects of our ability to preferentially attend to one sound source out of many in a room. This ability is colloquially referred to as the "cocktail party effect", and is greatly aided by our ability to listen binaurally, as described in Durlach and Colburn(2).

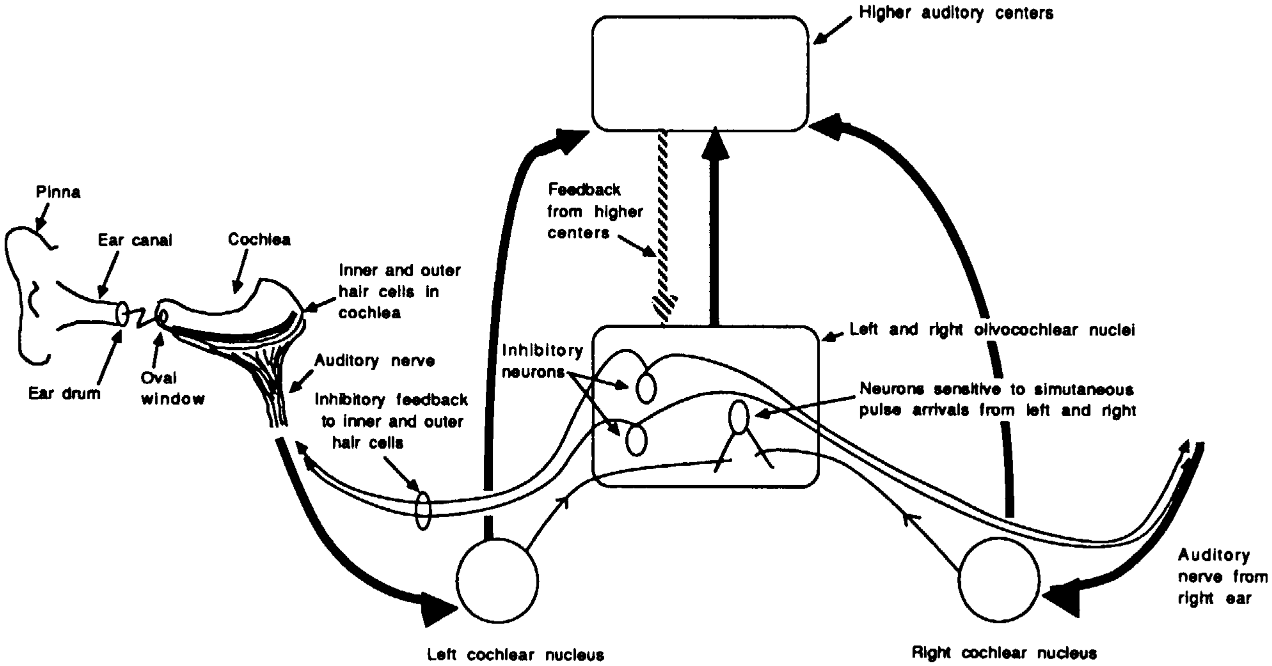
AUDITORY PHYSIOLOGY AND THE COCKTAIL PARTY EFFECT

The cocktail party effect can be described as a process of preferential amplification of one sound source relative to others, mediated by an attentive mechanism. Though the mechanism by which this occurs is not known, one can postulate a mechanism that is consistent with much of what is known about the auditory system . Figure1 summarizes some aspects of auditory neuroanatomy, covering the periphery, brainstem and higher auditory centers. Some key aspects of this organization are described below.

The external, visible part of the ear, called the pinna, acts to funnel sound into the ear canal. The shape of the pinna favors sounds based on their direction of arrival. In humans the pinna is relatively fixed, while in many animals, e.g. cats and dogs, the pinna is movable and allows them to better locate sound sources. The ear canal provides relative amplification to sounds in the 3Kz. to 5 Kz. range, thus tending to flatten somewhat the downward sloping shape of the spectrum of sounds made by the human vocal tract, which heavily concentrates energy in the range below 400 Hz. The middle ear consists of the eardrum and a system of bones that transform the vibrations of the eardrum into vibrations of the oval window connected to the cochlea.

The cochlea is a coiled structure filled with fluid. Vibrations of the oval window cause travelling pressure waves in the cochlea. This

FIGURE 1: SELECTED ASPECTS OF AUDITORY NEUROANATOMY



travelling pressure wave causes a displacement of the tectorial and basilar membranes in the cochlea. The cochlea contains two kinds of hair cells arranged in rows along its length. Outer hair cells have their tips connected to the tectorial membrane. They exert an active mechanical action on this membrane that tends to amplify the prevailing motion. The combined action of the travelling pressure wave and the active feedback from the outer hair cells makes the cochlea a highly resonant structure, wherein a particular input frequency causes a maximal amplitude of vibration at a particular point along the basilar membrane. High frequencies result in a resonant peak in the motion of the basilar membrane near the input end, while low frequencies map to resonances farther along the membrane. This system provides a mapping of frequencies into location along the basilar membrane. This mapping is maintained in the projections of the cochlea to the cochlear nucleus and the olivocochlear nucleus, described below.

Information from the cochlea is conveyed to the cochlear nucleus via the auditory nerve. This nerve carries information from both the inner and outer hair cells. The inner hair cells generate potentials based on the velocity of fluid movement in their immediate vicinity. The outer hair cells generate potentials also as their tips move with the tectorial membrane. Both the inner and outer hair cell potentials are converted into pulses that travel along the auditory nerve to the cochlear nucleus. The cochlear nucleus contains a variety of cells that transform and relay the information received from the auditory nerve. The cochlear nucleus projects, among other areas, to the olivocochlear nucleus, which is known to contain cells that respond only to simultaneous pulses received from the left and right cochlear nuclei. The olivocochlear nucleus projects to and receives feedback from higher auditory centers. The olivocochlear nucleus also projects back to the cochlea via cells that inhibit the action of the hair cells. One category of these cells inhibits the motor action of the outer hair cells, reducing the amplification they provide to the traveling pressure wave. In other words, these inhitory cells act as a form of automatic gain control to limit the amplification provided to high intensity inputs. Another category of inhibitory cells suppress the action of the inner hair cells. Further details on the organization and function of auditory subsystems can be found in Shepherd(3), and Kim(4).

The postulated scheme for explaining the cocktail party effect is as follows. A particular sound source generates pressure waves that arrive at the two ears with a fixed interaural time difference. For an average sized human, the maximaum interaural time difference is on the order of 500 microseconds. The left and right cochleas resolve this input into its frequency components, but maintain this interaural time difference in the pulse patterns sent along the auditory nerves. The timing difference between pulses emanating from the left and right cochleas causes a maximal excitation of cells in the olivocochlear nucleus at a particular lateral position along this nucleus, for each frequency band in which the input has energy. This excitation is integrated across frequency bands to provide an excitation peak whose lateral position in the olivocochlear nucleus is a function of the source location. When multiple sources are present, multiple excitation peaks with different lateral positions will result. This information is sent to higher auditory centers where an attentional mechanism acts to select a particular source to focus on. This information is sent back down to the olivocochlear nucleus, where it interacts with the excitation peaks in each frequency band. This interaction is used to suppress auditory nerve activity in frequency bands not corresponding to those of the selected source. This is done through the cells in the olivocochlear nucleus that project back to the cochlea and inhibit the response of the inner hair cells. Thus the information relayed to higher brain centers fron the cochlear nucleus is much richer than it would be otherwise in signal components corresponding to the selected source.

The fact that this improvement in signal-to-noise ratio (where noise refers to any signal other than that emanating from the chosen source) is achieved essentially by inhibiting the noise components is consistent with the following excerpt from Schubert(5) :

"It is readily apparent, though not easily quantifiable, that if one listens in the presence of two sounds, either of which might, under given circumstances, be the signal and the other the interfering noise, either of them can be made perceptually stronger than the other more or less at will. To a degree this must also be true of a signal in the interfering noise, and the process by which it is accomplished must in turn be some form of inhibition of the unwanted signal. The fact that with certain waveform differences between ears this can be accomplished binaurally to a greater degree than monaurally still leaves open the possibility that what is being accomplished is inhibition of the interference rather than some form of direct enhancement of a wanted signal."

SIMULATION OF THE COCKTAIL PARTY EFFECT

Figure 2 is a schematic of a large scale neural system that simulates the cocktail party effect. It consists of a number of subsystems, described below in more detail :

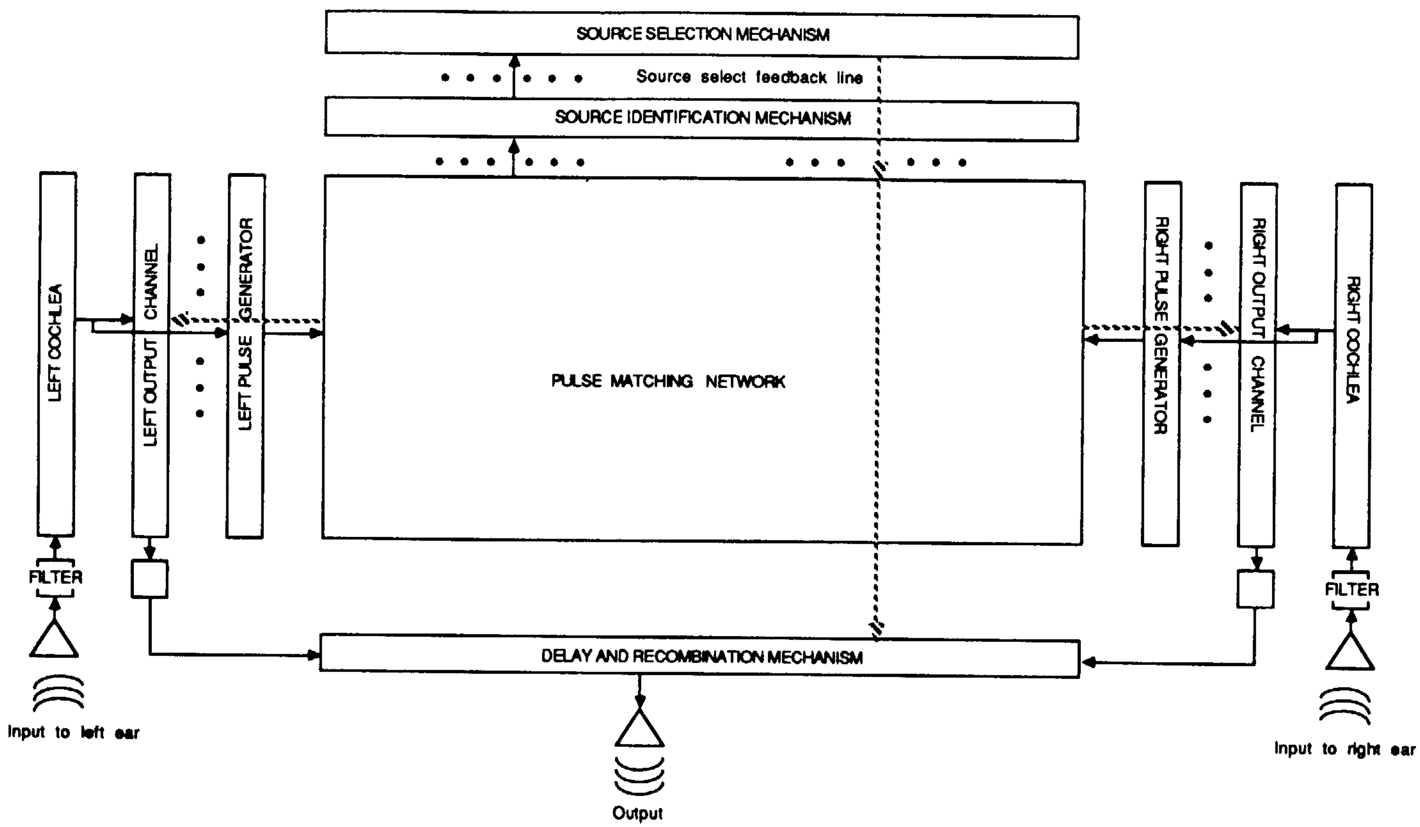
(a). Each input has a cascade of two high pass filters to mimic the kind of spectrum equalization provided by the pinna and ear canal system. Consequently, natural sounds reaching the input of the cochlear model have a flattened spectrum, which simplifies the output computations of the cochlea.

(b). The cochleas each consist of cascaded second order sections as described by Lyon and Mead(6). Automatic gain control is provided by a feedback mechanism from the output power of a cochlear tap to the Q-factor of sections closer to the input. An exponentially weighted temporal average of the power in the sum of left and right signals is used as the control signal for a common Q-control line to both left and right cochleas. This ensures that the automatic gain control mechanism does not distort the phase relationships between left and right cochlear outputs. The feedback mechanism operates by reducing the Q for a section as the activity at sections farther from the input increases.

(c). The sinusoidal outputs from the cochlear taps are fed into pulse generators that generate a pulse of fixed duration when a positive going zero crossing is detected in the cochlear output.

(d). The pulse matching network performs several tasks. Within each frequency band it delays and matches the pulses from the left and right cochlear taps, and creates a spatially and temporally decaying measure of matched pulse activity in a lateral zone dependent on the interaural time difference between the left and right signals. It integrates this measure of activity over frequency bands to create a global measure of matched pulse activity that corresponds to the presence of different sound sources present in the inputs reaching the two pickups. Finally, it 'and's the measure of activity within each frequency band with a 'source select' feedback signal (generated by the source selection mechanism, described below) to provide for a selective transfer of signals from the cochlear taps to the output channels on each side.

FIGURE 2: SCHEMATIC OF SYSTEM TO MIMIC COCKTAIL PARTY EFFECT



(e). The source identification mechanism identifies the centroids of local activity in the integrated outputs of matched pulse activity from the pulse matching network. Each centroid so identified corresponds to a particular source that can be focused on.

(f). The source selection mechanism selects a particular source to attend to and indicates this by activating a particular 'source select' fedback line into the pulse matching network. This 'source select' line also feeds into the delay and recombination system, described below.

(g). The output channels on each side combine the outputs from cochlear taps as mediated by the pulse matching network described above. Basically, what gets on to these output channels should be high in signal content corresponding to the selected source.

(h). The delay and recombination network provides a relative delay to the left and right output channels corresponding to the relative delay indicated by the 'source select' line. It then adds these outputs to create a single output in which the signals from the selected source add in phase. This is the output of the system.

There are obvious parallels between the postulated neurophysiological mechanism described earlier and the system described above. The pulse matching network seeks to mimic the postulated behavior of the olivocochlear nucleus. The gating of outputs from the cochlear taps to the output channels is similar to the postulated mechanism whereby inner hair cells picking up signals not emanating from the selected source have their outputs inhibited. The selection of which source to attend to occurs in a region analogous to the higher brain centers in the postulated mechanism.

Save for the logic based source selection mechanism (f) above, all the other processes in the system are neural in nature. The system is ideally suited for implementation in analog VLSI, but can also be implemented digitally. Efforts are currently underway to characterize the performance of this system.

REFERENCES

(1) Mead, Carver. 'Analog VLSI and Neural Systems'. Addison-Wesley Publishing Company, 1989.

(2) Durlach, Nathaniel I., and Colburn, H. Steven. 'Binaural Phenomena', in Carterette, Edward C., and Friedman, Morton P., (Eds), 'Handbook of Perception, Volume IV, Hearing', Academic Press, 1978, pp365-466.

(3) Shepherd, Gordon M. 'Neurobiology', Second Edition, Oxford University Press, 1988, pp304-325.

(4) Kim, D.O. 'Functional Roles of the Inner- and Outer-Hair-Cell Subsystems in the Cochlea and Brainstem', in Berlin, Charles (Ed) : 'Hearing Science', College-Hill Press, 1984, pp241-262.

(5) Schubert, Earl D., 'History of Research on Hearing', in Carterette, Edward C., and Friedman, Morton P., (Eds), 'Handbook of Perception, Volume IV, Hearing', Academic Press, 1978, pp72.

(6) Lyon, Richard F., and Mead, Carver, in Mead(1), pp279-302.