Approved for public release; distribution is unlimited.

A Miniature Acoustic Sensor that Mimics Mammalian Sound Processing

October 5, 2001

Allyn Hubbard[1], Tom Bifano[1], Jihad Boura[1], Fangyi Chen[1], Andrew Cherry[1], Howard Cohen[1], Socrates Deligeorges[1], Christopher Kapusta[3], Christian Karl[1], Piet Moeleker[3], Matt Nielsen[3], Errol Pinkney[1], Jim Simmons[2], Viktor Vajda[1], Stanley Weaver[3], Robert Wojnarowski[3], and David Mountain[1]

[1]Boston University
College of Engineering
Boston, MA

[2]Brown University
Department of Neuroscience
Providence, RI

[3]General Electric Corporate Research and Development
Schenectady, NY

ABSTRACT

We are developing a miniature acoustic sensor that contains mechanical and electrical counterparts to the auditory processing subsystems used by mammals. The overall plan is to mimic the mammalian processing scheme, eventually to the point of achieving mammal-like recognition of auditory objects.

The expected device is currently in development stages. The device is envisioned to consist of a mechanical frequency analysis stage, a subsystem for extracting both envelope and temporal information and converting analog information to events, and a digital subsystem that extracts features. Features currently slated for extraction are pitch and bearing, however the entire feature extraction subsystem is designed in a way that is amenable to modification for enhanced capability

The cochlea-like front end is being designed using a ribbed structure composed of polymer ground structures built up with metal, which comprise coupled resonators. The custom VLSI electronics are being designed to acquire a signal from the artificial cochlea obtained via electrically responsive materials or devices embedded into the structure, thus forming an integrated manufacturable apparatus. The VLSI

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
05OCT2001	N/A	-

Title and Subtitle A Miniature Acoustic Sensor that Mimics Mammalian Sound Processing	**Contract Number**
	Grant Number
	Program Element Number
Author(s)	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) Boston University College of Engineering Boston, MA	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es) Department of the Army, CECOM RDEC Night Vision & Electronic Sensors Directorate AMSEL-RD-NV-D 10221 Burbeck Road Ft. Belvoir, VA 22060-5806	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)
Distribution/Availability Statement Approved for public release, distribution unlimited	
Supplementary Notes Papers from 2001 Meeting of the MSS Specialty Group on Battlefield Acoustic and Seismic Sensing, Magnetic and Electric Field Sensors, Volume 1: Special Session held 23 Oct 2001. See also ADM001434 for whole conference on cd-rom., The original document contains color images.	
Abstract	
Subject Terms	
Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** UU
Number of Pages 12	

electronics, which comprise stages of special adaptive filtering and translation to digital events, are to send digitally encoded data to field programmable gate array (FPGA)-based processors.

1. Mammalian Hearing

Mammals are hearing specialists, although rudimentary forms of hearing are present in many life forms above the single cell level. In every case there exists some peripheral mechanism that transfers sound energy from the external environment to some sensory cell. Signaling from that cell may occur chemically, and the chemical event may, and will certainly in the case of higher organisms, be converted into an electrical event in the nervous system. The processing that takes place thereafter is most advanced in mammals. Thus, there is motivation to emulate a mammalian hearing apparatus using a nonbiological apparatus.

In the text to follow, physiology, anatomy, and neurophysiology of the ear will be introduced in a way that moves the scientist or engineer to a rather inevitable design for a small, biomimetic hearing device. Effort is made to note the engineering challenges as well as workarounds, such that the essence of the design might be implemented using technology.

1.1. The hearing apparatus

1.1.1. The external ear

The hearing mechanism in the mammal is well identified. The process begins with the head and the external ear, and in a sense they, along with the body are the true "external" ears. The body, especially the shoulders and the head modify the sound field that reaches the pinna. The pinna with its unique shape for various mammals and even individual shapes within species further conditions the sound that enters the external ear and reaches the eardrum. The eardrum itself is a complicated structure that acts as a simple piston only at lower audible frequencies, and moves in a complicated manner at higher frequencies.

1.1.2. The middle ear

The eardrum connects to the malleus, the first bone of the middle-ear chain that includes the incus and stapes. These bones serve to transfer mechanically the acoustic energy into the cochlea or inner ear. Once in the cochlea, the energy is acoustic. During the translation from acoustic to mechanical to acoustic energy, impedance matching is accomplished. This is not a trivial task, since the transfer is from air, which is a low-impedance medium. Acoustic energy in air generally reflects from water, but in the ear, most of the energy actually makes it into the cochlea itself. The exact mechanism of the middle ear is not well understood: Simple models consist of bandpass resonant mechanical lumped elements. More complicated models, for example ones carried out using Finite Element Methods, require detailed knowledge of the fine contours of each bone and their connections to each other as well as supporting ligaments [For general reading, see Rosowski, 1995.]

1.1.3. The cochlea

The cochlea is often drawn as (see Figure 1) a hairpin shaped tube, made from a straight tube bent back below the upper tube. The structure separating the tubes is not made of the same material as the hard walls of the tubes. The biological structure separating the channels includes sensory cells and other fine structure, but it is typically viewed as a single entity, called the basilar membrane (BM). From a modeler's perspective, the basilar membrane has characteristics of compliance, mass, and damping. The narrow region of the basilar membrane resonates at high frequencies, while the wide end resonates at lower frequencies.

Sound entering the stapes end of the tube travels at the speed of sound down the top tube, through the bottom tube, impinging on a compliant membrane at the distal end, which is called the round window. In simplest terms, it is a pressure relief. Thus, the stapes pushes in like a piston, and the round window bulges out.

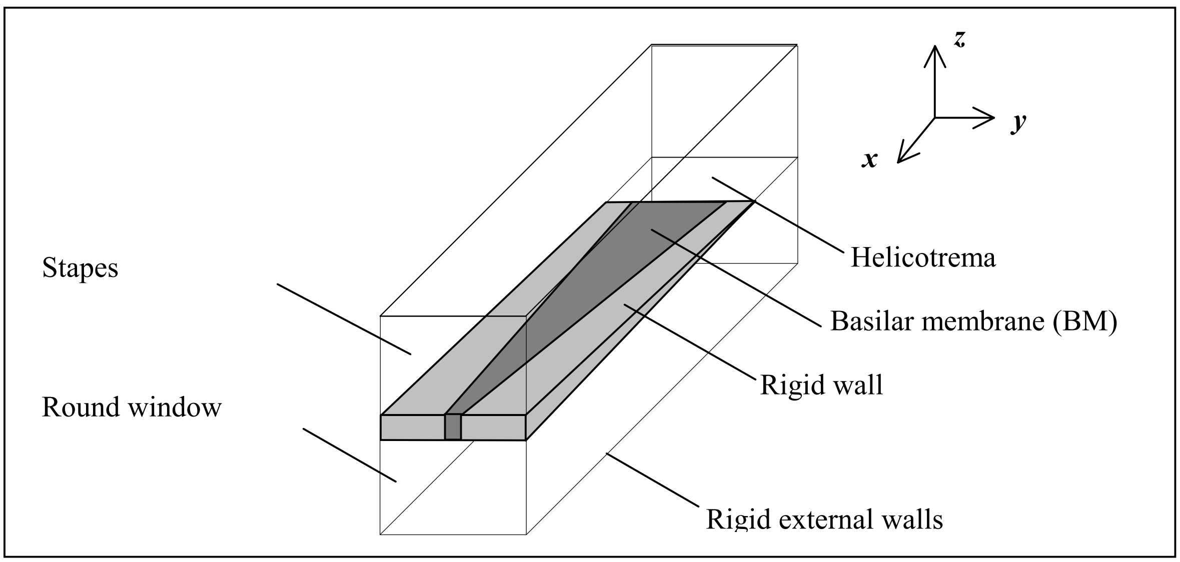

Figure 1 A schematic view of the mammalian cochlea.

What is most surprising is not the sound wave that travels at the speed of sound in water, but a second energy transfer modality, which is a slow wave. This wave can be thought of as a pressure-difference wave, because at any position along the length of the membrane that separates the two tubes is a pressure difference. This pressure difference is what moves the BM. This structure is of particular importance to ascertaining the function of the ear, as it is possible to measure *with great difficulty* the motion of the BM and so validate theories of how the cochlea might work. Suffice it to say that teleological, experimental, and model evidence supports the notion that the cochlea is a mechanical frequency analyzer. At every location along its length, frequencies are sorted out in a novel, sequential fashion. The structure is like a chain, as contrasted against a bank of filters. The energy enters the chain at the stapes end of the cochlea. The lower the frequency, the farther it propagates. The chain sequentially extracts high frequency energy. Thus, a single tone of high frequency produces a slow-traveling wave in the cochlea that proceeds only a short distance and its energy is, in a very real sense, deposited at a spot, a very small spot. A lower frequency tone would travel to a spot farther down the cochlea. Thus, in general, tone in air is impedance matched in a way that allows it to flow into the mammal's ear opening and all the energy (and more as we will see later) is concentrated at a spot, more accurately a region of about 10 microns by 100 microns.

In the mammalian cochlea exists another mechanism, the cochlear amplifier. It is not a focus of the present paper, whose scope is to encompass the cochlear apparatus previously elaborated (c.f. Figure 1). However this mechanism cannot be ignored in any reasonable tome on hearing, and may be included in a future generation of the present device, in a way that will be made obvious in the course of the paper. The cochlear amplifier has been modeled as a feedback loop [Mountain, Hubbard, and McMullen, 1983] that senses small amounts of acoustic energy (via the sensory cells, which are called outer hair cells (OHC)). These cells look somewhat like carrots, with the cell body being analogous to the carrot body and the hairs being analogous to the carrot leaves, although the hairs are straight, not bushy. These cells are actually little motors. When hairs on their top side are moved slightly, the whole cell changes length. Gain is involved, because a metabolic battery drives the process. Thus to summarize in a manner purposeful for this paper, motion in a region of the cochlea creates an electrical change that amounts to an electrical signal, which actuates a motion that amplifies the original sound signal in a frequency-selective manner. The analogy with piezoelectric devices cannot be more obvious [Mountain and Hubbard, 1994; See also for an engineering approach to modeling the cochlea Hubbard and Mountain, 1995].

1.1.4. Central processing

It is inner hair cells, which are not part of the feedback amplification loop, that sense motion and transduce that motion into electrical events including rectification and low pass filtering (akin to envelope detection), and finally the generation of neural impulses that are further processed by the brain [See for an engineering approach, Mountain and Hubbard, 1995]. A *simplified* depiction of auditory processing pathways in the brain is shown in Figure 2a. The brain processing is at each stage is not yet not completely understood. Investigators seek to understand the neural basis of the processing mechanisms by recording voltage signals from the brain, and through various analytical and sometimes imaginative ways deduce what processing takes place. Put together with data from psychoacoustics one may develop an abstraction of the hearing process as shown in Figure 2b. Working in concert, the mechanisms in Figure 2a perform the functionality of feature extraction and eventual regrouping for the purpose of identifying sources. Analogous, and more broadly-known processes are also known to exist in vision, e.g. edge detection, oriented spatial filtering, contrast gradient detection, etc. These mechanisms have, in turn, been carried out algorithmically or in hardware for purpose of effecting machine vision. Similarly, in the work to follow, we seek to emulate the essence of acoustic feature detection that mammals use to recognize acoustic objects.

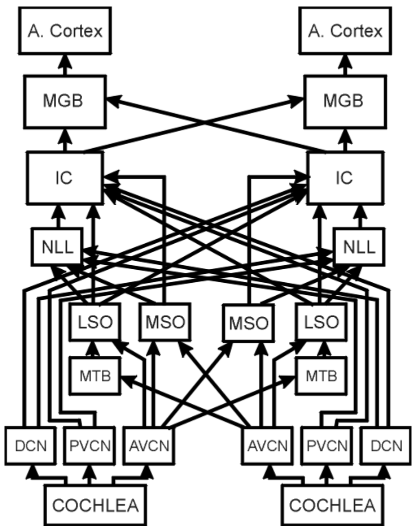

Figure 2a. A simplified schematic of mammalian auditory processing in the brain. The abbreviations in the boxes indicate standard anatomical nomenclature for various nuclei and brain regions.

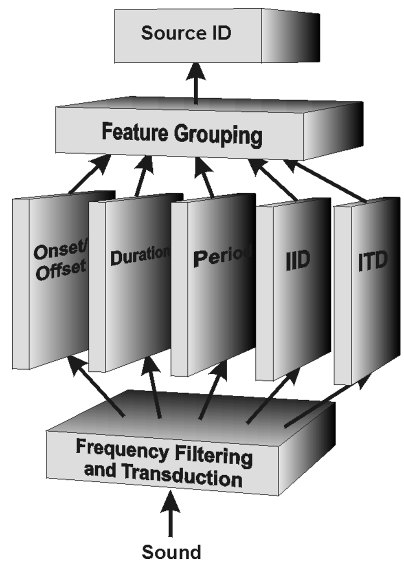

Figure 2b. A psychoacoustician's view of acoustic processing in the brain. The bottom block refers to the external, middle, and inner ear including transduction into the auditory nerve. From there, features are "extracted" by various neuronal groups. Finally, grouping of the features occurs, in order that classification can take place.

2. Cube hearing

Figure 3 depicts the proposed device as an embodiment of the external ear *plus* the cochlea shown in Figure 1 *plus* a front-end transduction and signal processing unit *plus* ITD and Pitch processing subsystems shown in Figure 2b. The assembly should fit inside a package on the order of an inch on a side.

2.1. Subsystems

Figure 3 further depicts a mechanism for delivering environmental acoustic energy to a MEMS cochlea, VLSI analog electronics, a digital event-based signaling mechanism, an intermediate processor stage that extracts certain features, and a means for connection to a computer that performs advanced analysis. The cube is well characterized in terms of its acoustic properties; and although some shaping may be done for purposes of modifying its directional characteristics; it might never resemble the head of a mammal. To get sound from the environment to the cochlea, we plan to use miniature microphones and a piezoelectric driver as the equivalent of the stapes (c.f. Figure 1). The MEMS cochlea needs to perform frequency analysis, and this is to be carried out using a cochlea-like structure constructed of polymers and metal with an appropriate transduction mechanism. Special modifications and variations will be made to this structure to achieve target mechanical performance as well as manufacturability. The front-end electronics are to be very-low power analog, current-mode circuitry that is

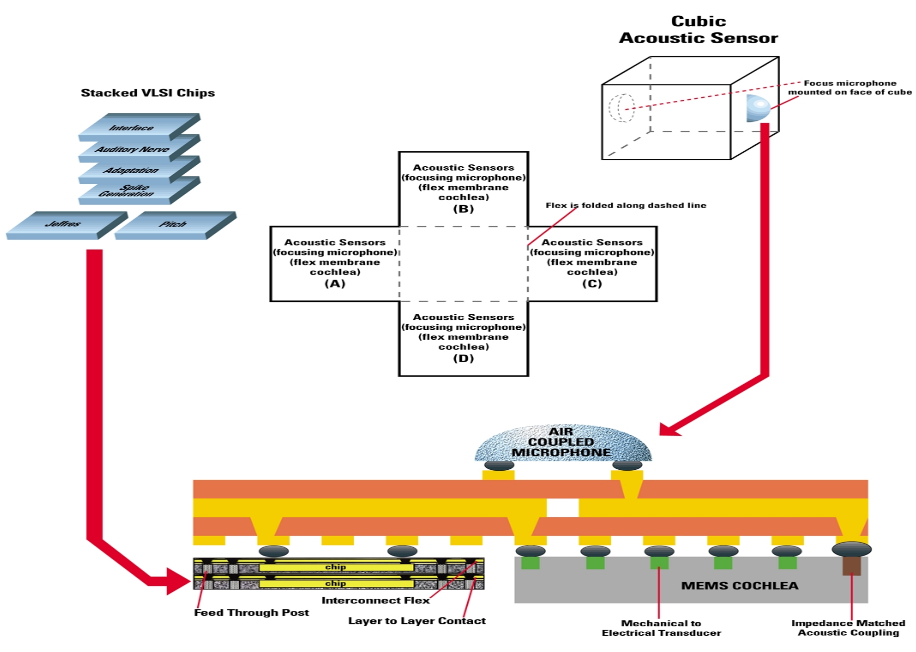

Figure 3 A schematic of the cube apparatus. The cube contains a mechanical cochlea and VLSI circuits that process signals from the cochlea.

temperature and supply-voltage compensated. The signal processing functionality of these units is similar to that used in the ear, but while the ear is limited by biological constraints that, for example, limit bandwidth, the electronics suffer no such constraint relative to audio frequencies. The same VLSI chip has asynchronous digital encoding and transmitting capability, such that "events" are codified and sent out to a subsequent processing stage, currently implemented using a XILINX FPGA. The FPGA is used because of its flexibility for recoding, as the auditory features currently slated for extraction are not the full menu of features used for source recognition by mammals. The final component is currently a PC, which allows total, if not belated, processing capability of the features extracted by the FPGA stage. This is prudent, because the ultimate way in which mammals handle sound is only partially known. Thus the subsystems make a transition from hard to soft, reflecting the current state of the knowledge about the auditory processing used by mammals. Thus, the back end of our device can be changed, while the front end is rather solidified, conceptually and cast in custom silicon.

2.1.2. Counterparts to the external and middle ear

Although originally conceived as a piezoelectric sensor comprising the entire face of the cube feeding a second piezoelectric driver juxtaposed into the fluid-filled artificial cochlea; the current embodiment uses miniature hearing-aid electret microphones. With sensitivity approaching that of ½ inch condenser microphones and a bandwidth out to 100 kHz, these microphones are excellent. Using this microphone also requires the utilization of a power source/amplifier, which will prove useful given that an artificial cochlea without a cochlear amplification mechanism will be less sensitive than its biological counterpart.

2.1.3. The artificial, mechanical cochleae

The mechanical cochlea, designed with a similar shape to that shown in Figure 1 is to be comprised of a sheet of polymer onto which are placed metal bars having various patterns, the simplest being that of a beam clamped on both ends. These beams, being of varying sizes, have varying resonant frequencies. The serial concatenation of the beams separated by polymer, forms the analog of the basilar membrane. The beam structure needs to be placed inside a fluid-filled cavity, because the coupling must be via the fluid mass and not the elastic membrane.

Figure 4 shows an example test structure comprising eight slots, each of which has a different design. In general, the slots have the taper of the mammalian basilar membrane (c.f. Figure 1). Designs have made using various polymer materials as the basement structure. Onto the polymer structure are imbedded beams made of metal, having a variety of widths, shapes, and edge-clamping.

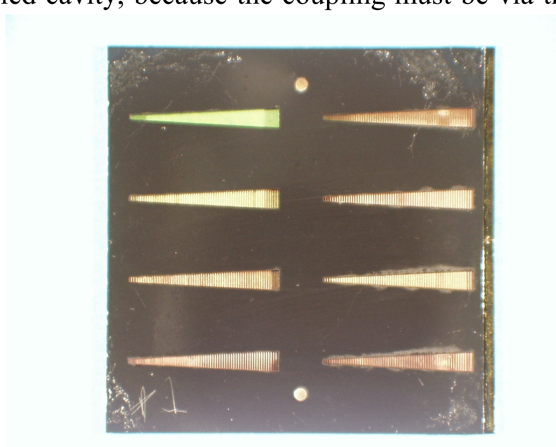

Figure 4 A sample containing eight different "basilar membrane" structures (c.f. Figure 1). The fine structure of differences between the sites (slots) is not apparent at this magnification.

We tested these structures using a variety of stimuli. One test strategy for low-frequency structures was to use a “tweeter”, which is a high-frequency sound source that works up to about 35 kHz but has poor low-frequency response. We have also used Polaroid ultrasonic transducers to reach up the 150 kHz range. We measure sound near the beam being measured using a B&K 1/8th inch microphone, whose response is essentially flat up to about 150 kHz. The velocity data measured on individual beams is scaled by the pressure.

Figure 5 shows the magnitude versus frequency response representing the velocity responses obtained from every one of 64 beams within a test slot. For this data obtained from a structure made with a coupling polymer having a relatively low bulk Young’s modulus, the individual beam resonances are estimated to be in the range of 3kHz to 12 kHz. The estimated resonances, as compared to exact calculations result from the fact that only a few of the beams are simple, clamped beams. Additionally, there is coupling between the beams by virtue of the polymer, whose effect is difficult to estimate and whose effect is one focus of the experimental measurements. The figure shows maximum responses change from about 15 kHz in the end of the cochlea with short beams (low-valued indices) to about 4 kHz (beam indices around 64).

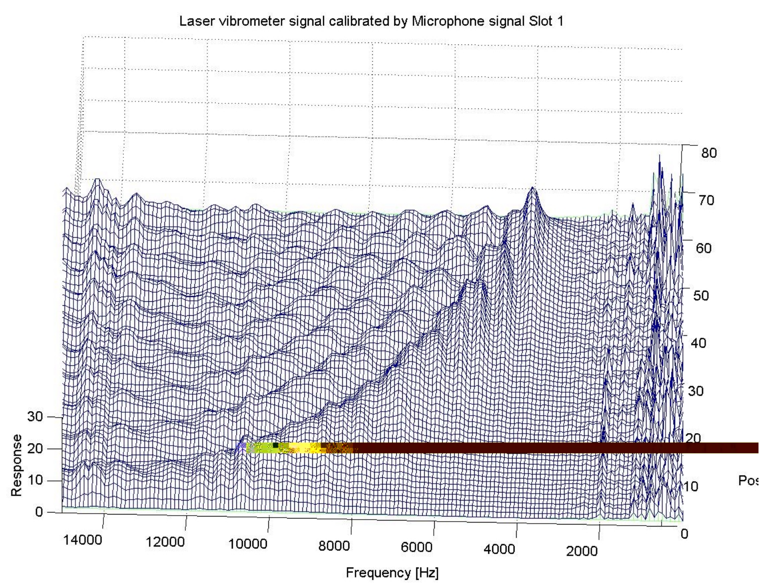

Figure 5 An example test result obtained using a sound stimulus facing the test structure in air. Notice that the frequency axis extends to the left, in order to more clearly display the contours. The narrow end of the structure supports higher frequencies than the wide end.

2.1.4. The VLSI electronics

The VLSI architecture depicted in Figure 6 is modular, in units of 64, since the original artificial cochlea is expected to have 64 mechanical resonators. The VLSI chip is to be capable of handling 16 such groupings of 64 frequency channels. This allows for the possibility of handling more cochlear channels without redesigning the VLSI chip. Such a VLSI processing chip with 64 channels requires about 32 square millimeters in a 0.25-micron technology. A separate chip handles the concatenation of up to 16, 64-channel processing chips. This was done so that as the artificial cochlea becomes more refined, hence more frequency channels, the VLSI chipset will not have to be refabricated to accommodate the increased channel density.

The VLSI electronics are currently designed to take a direct feed from the piezoelectric sensors that transduce motion in the artificial cochlea. The input MOSFET gates currently presume an ac signal ranging from 0.1 uV up to 100 uV with 0.0 dc value, coming from a piezoelectric source with a source capacitance that is determined by the size of the metalization on the cochlear beam.

The signal processing capability of envelope detection is carried out using current-mode circuitry. The dynamic range is shown by simulations to be approximately 80 dB. The output of the low-pass filtering feeds what is called an "adaptation circuit" which highlights changes in the signal's envelope for high-frequency signals, but produces highlights on a cycle-by-cycle basis for low-frequency auditory signals. The adaptation circuit feeds a thresholding circuit with a timeout, thus emulating the "firing" of a neuron attached to an IHC in the biological cochlea. Three different threshold levels are used to extend the dynamic range of the system, and these are carried out in triplicate circuitry, identical except that the "neural" thresholds differ. When a threshold crossing occurs in any of three channels having high, medium, and low threshold, a digital signal is emitted, which indicates cochlear location and threshold channel. Thus, the codification of auditory signals is place-specific (denoting in part, frequency content) and the time of codification also carries timing information. The timestamp occurs when the signal interrupts the FPGA.

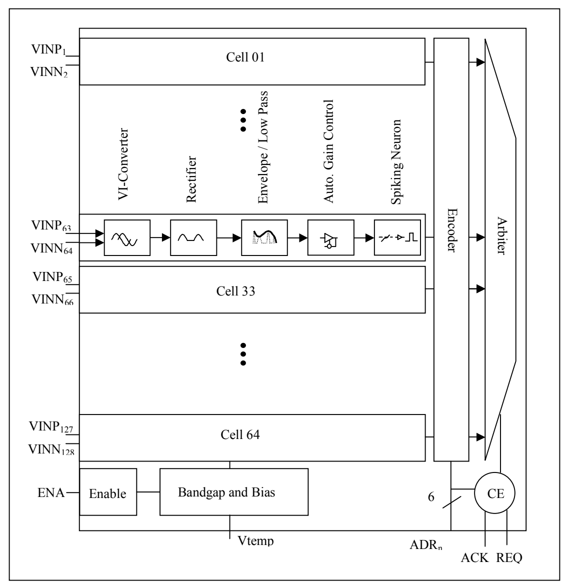

Figure 6 The custom VLSI electronics comprise 64 channels per chip, each of which is entails several signal processing operations.

2.1.5. FPGA-based intermediate processing

In the current design, pitch and bearing information is extracted by the FPGA. Both operations are carried out in a non-neural fashion. Fundamentally, operations are performed with timers. Pitch events are noted when two "spikes" from the same frequency channel (location in the cochlea) occur within predetermined times. These times are binned, and the bins are codified. Thus, the occurrence of a pitch event comprising its time and its bin number is sent to the computer. Similarly, for the same channel from the right and left ears, a bearing event is determined when right and left "spikes" occur within 60 microseconds. Again, the event and its time of occurrence, and the lead/lag are reported to the computer.

2.1.6. The entire assembly

Currently, the prototype boards are designed for the purpose of mounting and debugging left and right systems and the XILINX FPGAs. This is shown schematically in Figure 7 for a system comprising 128 channels from each ear, using two analog processing chips (APCs) handling 64 channels each. Portions of the 128 channel artificial cochlea connect to an APC, in groupings of 64. For each APC, 128 wires are needed, as the input to the APC is differential, not single-ended. Data from the APC are "events" coded into bit fields requiring eight bits. Two additional signals are needed for signaling purposes. The communications chip tacks on an additional four bits to encode which APC generated the data, in this case APC(0) or APC(1), thus allowing a fine-grained system to have up to 1024 channels before changing the electronics chipset. Each XILINX FPGA receives simultaneous data from the right and left ears via dedicated wires. More bits are needed to encode the features found by the XILINX FPGA, which are sent to a computer via parallel digital interfaces.

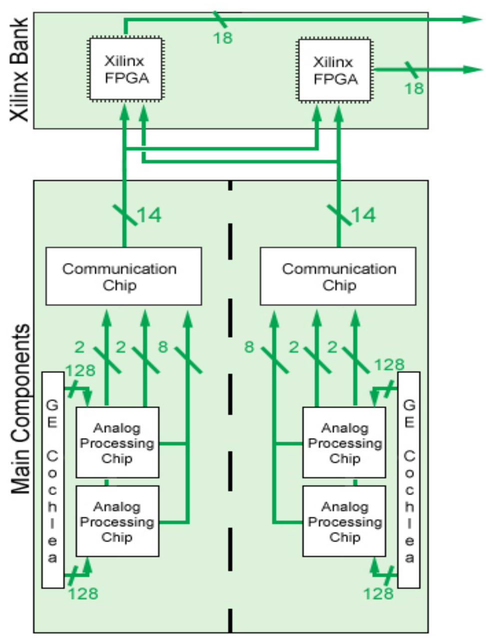

Figure 7 Schematic of assembly for testing the apparatus.

Ultimately, it is estimated that the artificial cochleae and the silicon, *not the chip carriers*, plus interconnect implemented via GE flex interconnect will fit inside a one-inch cube. The flex interconnect allows chips to be "stacked" up as depicted in Figure 3, and the scale of the interconnect may easily number in the hundreds. We have already made measurements using the cube with miniature microphones shown in Figure 8. Behavioral modeling has already been carried out using "real" signals heard through the "ears" on the cube.

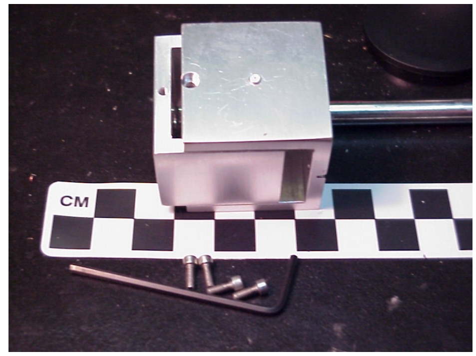

Figure 8 A one-inch cube with a miniature microphone mounted at the center of the top face.

2.1.7. The advanced processing stage

For test purposes, event-based data including a pass-through of every "spike" will be logged for off-line analysis. This at present is for the purpose of thorough testing. In the near future, the logged data will be used to quantify how much information is lost by only extracting two features from particular target signals. In addition, having all data at hand will afford an opportunity offline to understand which feature-extractors (of those shown in Figure 2, implemented in software) are most relevant to particular acoustic targets. Ultimately, the device should be stand-alone, with all processing implemented via hardware.

3. Future development

The work reported here is in various stages of development: This paper, therefore, should be considered a mix of what has been carried out and what is planned to be carried out over the short term. Eventually, we expect to include several other features inside the target one cubic inch frame. We plan to incorporate more feature-extractor mechanisms than the two now implemented. These include loudness queues, interaural loudness differences, and spatial-temporal filtering specific to a particular target's identification. As depicted in Figure 9, we also want to include electronics for telemetry that allow clustering of individual cubes forming dynamic networks, to allow conservation of energy resources and the formation of acoustic beams of very large spatial and numeric dimension.

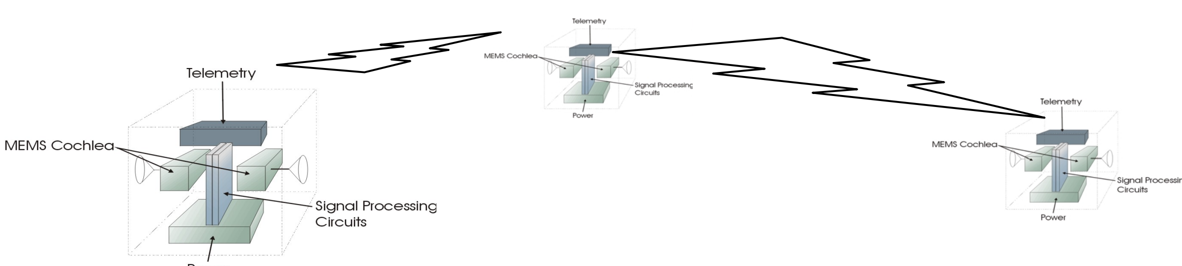

Figure 9 An interconnected array of devices in a field application

References

1. Hubbard, A.E., Mountain, D.C. "Computational analysis of the inner ear and auditory nerve processes," in *Auditory Computation*, Hawkins, H. L., McMullen, T. A., Popper, R. R., Fay, R. R.,(eds.) Springer Verlag, New York, chap. 3, 1995.
2. Mountain, D.C., Hubbard, A.E. "Computational analysis of the inner ear and auditory nerve processes," in *Auditory Computation*, Hawkins, H. L., McMullen, T. A., Popper, R. R., Fay, R. R.,(eds.) Springer Verlag, New York, chap. 4, 1995.
3. Mountain, D.C., Hubbard, A.E. and McMullen, T. Electromechanical Processes in the Cochlea, in Mechanics of Hearing. M.A. de Boer, E. Viergever eds., Delft University Press, 1983.
4. Rosowski, J. "Computational analysis of the inner ear and auditory nerve processes," in *Auditory Computation*, Hawkins, H. L., McMullen, T. A., Popper, R. R., Fay, R. R.,(eds.) Springer Verlag, New York, chap. 2, 1995.
5. Mountain, D., Hubbard, A.; "A piezoelectric model of outer hair cell function. J," *Acoustical Society of America*, V95, pp. 350-354, 1994.