

ANALOG IMPLEMENTATIONS OF AUDITORY MODELS

Richard F. Lyon

Apple Computer, Inc.
Cupertino, CA 95014
and California Institute of Technology
Pasadena, CA 91125

ABSTRACT

The challenge of making cost-effective implementations of auditory models has led us to pursue an analog VLSI micro-power approach. Experiments with the first few generations of analog cochlea chips showed some of both the potential and the problems of this approach. The inherent exponential behavior of MOS transistors in the subthreshold or weak-inversion region leads to nonlinear filter circuits, in which the small-signal and large-signal behaviors can be quite different. Early problems with instability, poor dynamic range, and excessive noise are now understood in terms of the transition behavior between these regions, and this understanding has led us to design filter stages with appropriately compressive behavior, resulting in more robust cochlea performance. Several types of correlator circuits to follow the cochlea have also been developed into working demonstrations. Videotapes of circuit outputs and simulations illustrate the recent ideas and progress.

INTRODUCTION

The “Analog Electronic Cochlea” of Lyon and Mead [1] has been presented as a very efficient way to implement a cochlear model in silicon. It was shown to provide some of the basic filtering properties and adaptability potential needed for a comprehensive auditory model. However, we were not able to tune the filters to give a pseudoresonant gain peak greater than about 20 dB, due to poorly understood circuit misbehaviors.

In the hydrodynamic model on which our cascade filterbank model of the cochlea is based [2], the signals in the filter cascade correspond to either pressure (across the basilar membrane) or velocity potential at the membrane (the spatial derivative of velocity potential is the fluid velocity vector). For the case of a passive cochlea, the response of the filters should be monotonically decreasing with frequency; i.e., lowpass. But for a healthy active cochlea in a quiet environment, the active outer hair cells should provide enough gain to change the overall lowpass response to a pseudoresonant bandpass-like response with a broad gain peak of perhaps as much as 60 dB. Over a wide range of sound loudnesses, the filterbank gain should gradually transition between these extremes, acting as an automatic gain control to compress the dynamic range of signals at the output. We have previously discussed the evidence for this kind of function in fact occurring at the mechanical level [3]. We are now making progress on getting this behavior into our circuits.

COCHLEA IMPROVEMENTS

The original cochlea chip was nothing more than a cascade of two-pole filter stages (omitting the zeros of previous models that

had come from a long-wave analysis). Further analysis led us to expect three-pole stages to provide an even better fit to the hydrodynamic short-wave analysis. Most of the circuits and behaviors discussed here apply with minor changes to either type of stage.

Tests on early second-order filter stages, including a low-noise version based on the MOSIS low-noise analog BiCMOS process, revealed a problematic nonlinear effect related to the saturation of the *tanh* input nonlinearity of the transconductance amplifiers. The filter stages ended up with more gain for large signals than for small signals (i.e., they were “expansive”), and the result was that a given periodic input could lead to a pair of distinct periodic attractors. In a cascade of such filter stages, when the input became large enough to kick any stage into the large-signal mode, the final result was a chaotic output waveform resembling fractal mountains. This was neat, but not at all what we had in mind. We have since experimented with several ways to make inherently compressive filter stages, with good results. With recent circuits, we can tune the cochlea to have a peak gain close to 40 dB, rather than the 12 dB previously shown. This peak gain is closely related to both sharpness and dynamic range, as will be discussed.

Filter circuits

The first-order and second-order circuits of figures 1 and 2 are analyzed by Mead [4], including some of the large-signal behavior. The small signal, or linear, regime of operation of the modified second-order and third-order circuits of figures 3 and 4 are easily analyzed in the same way. The third-order circuit yields a complex pole pair plus a real pole; the same effect can also be achieved by cascading a first-order section and either of the second-order section designs, in either order. Several other circuit variations have also been considered, but here only the ones actually built and tested are discussed.

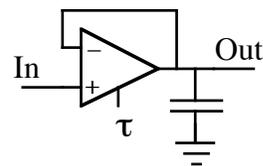


Figure 1: The circuit for a first-order (one-pole) section. The amplifier symbol represents a transconductance amplifier, which in subthreshold MOS has an inherent *tanh*(*v*) nonlinearity. Time constants in this design style are controlled by transconductances and capacitances, rather than by resistances and capacitances.

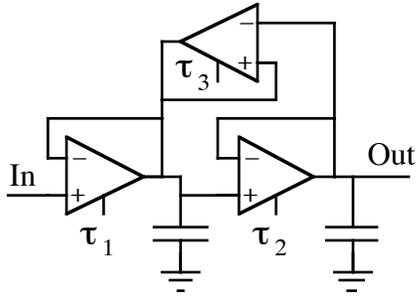


Figure 2: The circuit for Mead’s classic “ORD2” second-order section. All capacitors used are approximately identical, typically around 1 pF or less.

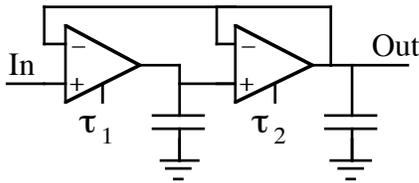


Figure 3: The circuit for the “DIF2” second-order section, which is a simpler circuit with slightly more complicated Q-control.

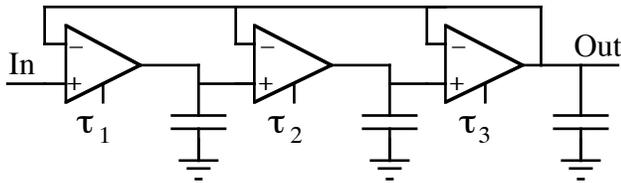


Figure 4: The circuit for the “DIF3” third-order section, which leads to a more accurate model of short-wave hydrodynamics.

In figures 1–4, the amplifier symbols represent abstract transconductance amplifiers. The circuits used to realize them inherently compress the input differential voltage through a hyperbolic tangent function whose “width” is fixed by the device physics and temperature, and largely independent of device geometry. The small-signal transconductance (gain) of each amplifier is a temperature-dependent constant times the bias current of that amplifier. The bias current is also equal to the output saturation current corresponding to $\tanh(\infty)=1$. The labels τ , etc., are used to indicate the time constants controlled by the bias currents of each amplifier ($\tau = C/g$, where g is the transconductance and C is the capacitance).

In testing an ORD2 stage, we noticed that for some moderate-level input signals of fixed frequency and amplitude, two different output amplitudes were possible. A frequency sweep showed a double-valued transfer function magnitude with hysteresis. Essentially, both the small-signal and large-signal regimes were stable, but with different gains—there were two distinct periodic attractors in this nonlinear dynamical system. This behavior was duplicated in Lazzaro’s “anaLOG” circuit simulator, one of the few simulators that implements a reasonable model of MOS transistors in subthreshold. Since the large-signal regime had more gain, a cochlea model built as a cascade of these stages could

be quite well-behaved while all the stages were in the small-signal region, and suddenly switch to a very nonlinear high gain mode if any stage got kicked up to its large-signal region (since all later stages would then be driven into their large-signal high-gain regimes). This result was very frustrating before we understood it—just as we would adjust the cochlea cascade to have a moderate gain peak, it would burst into uncontrolled fractal-noise-like behavior.

Mead showed that the second-order section can be stable for small signals and unstable for large signals, due to the \tanh nonlinearity, but he did not discuss the effect of the nonlinearity on stable behavior for moderate-size signals. The key to understanding the above behavior is a simple perturbed small-signal analysis. For any given frequency, the small-signal linear analysis will easily show which amplifier in the filter circuit has the largest differential input voltage (the answer may be frequency-dependent—consider frequencies near the pole frequency). That will be the first amplifier to start to feel the effect of the \tanh , which, to first order, will be to slightly reduce the transconductance of that amplifier. So, to predict how the poles will move and how the gain will change as input amplitude increases, it is only necessary to do a small-signal analysis with the original circuit parameters and then with a slightly modified parameter.

For the ORD2 circuit, the first forward amplifier sees the biggest signal, so its gain is perturbed downward, equivalent to lengthening the time constant τ_1 , reducing the pole frequency and increasing the Q and hence the gain. Having the gain increase with increasing input level is exactly the opposite of what we need—it is expansive rather than compressive.

The second-order DIF2 circuit of figure 3 is a simpler circuit in which the Q is varied by biasing the two amplifiers symmetrically about the point corresponding to the pole frequency (arithmetically symmetric bias voltage leads to geometrically symmetric bias currents and time constants, due to the exponential characteristic of the transistor in subthreshold). The Q is the square root of τ_2/τ_1 . For low enough Q settings, the first amplifier saturates first, increasing τ_1 as above. The pole frequency is also exactly proportional to the reciprocal square root of τ_1 , so the Q is reduced in proportion to the pole frequency. But with higher initial Q settings, the second amplifier saturates first, and the circuit is again expansive. Since a cascade of such stages will have a wide range of actual Q values, some will run into problems.

The third-order DIF3 circuit of figure 4 produces a transfer function that is flatter at low frequencies, as our hydrodynamic analysis suggests for the real cochlea; the resulting pseudoresonant gain peak is somewhat narrower than what we get with second-order stages. But the circuit still has the same stability problems.

To fix the problems, it is necessary to get some control over which amplifier reduces its gain first. In the ORD2, we want the feedback amplifier to compress while the forward amplifiers remain relatively linear. For the DIF3, we want the first stage to compress while the other two stages remain relatively linear. In the ORD2, the perturbation analysis then shows the poles staying exactly fixed in frequency and changing Q. For the DIF3, numerical studies in Mathematica™ showed that the poles move

in approximately the right directions, with Q being reduced and frequency changing only a little. How to get control of the linear range is then the key problem.

We have also built third-order filters by combining DIF2 with a first-order section. To try to stabilize this configuration, we considered always letting the first amplifier of the DIF2 saturate first. Such a circuit would have Q and complex pole frequency reduce proportionately as signal level increased. This sounds about right, and it does decrease the gain near the initial pole frequency, but it can actually increase the gain at lower frequencies, resulting in the same kind of two-region behavior over a limited range of parameters.

Amplifier circuits

The basic transconductance amplifier circuit is shown in figure 5. Compared to an operational amplifier, it is very simple and low-power, and has no critical transistor geometries if operated in subthreshold with audio-band signals. Typical device sizes used are 8-by-8 microns; minimum-size devices are avoided since flicker noise increases as the reciprocal of gate area, and there is little space advantage of making them smaller.

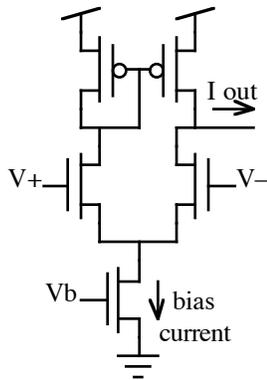


Figure 5: The transistor-level circuit for the simple transconductance amplifier.

Using the Mosis low-noise analog BiCMOS process, we have also experimented with amplifiers in which the upper mirror devices are replaced by quiet NPN bipolars, and the lower three devices are pMOS (which are quieter than nMOS due to the lower energy of the Fermi level relative to trap states). With this modification, the amplifiers proved to be relatively quiet, but we still got enormous fractal-like noise out of cochlea cascades, which is what put us on the right track.

Rather than focusing on low-noise amplifiers, we then turned our attention to expanding the linear input range. Linear micropower circuits are quite difficult to build. To achieve robust operation with the simple nonlinear circuits, it turns out that the ratio of transconductance to saturation output current needs to be modified on some amplifiers relative to the others. Reducing the transconductance relative to the output or bias current is equivalent to increasing the width of the linear input range.

The first method we used was capacitive voltage dividers at the inputs, as shown in figure 6, to reduce the signal seen at the

nonlinear differential pair. Since capacitors are essentially perfect and the differential-pair inputs are insulated gates, we figured it could operate essentially down to DC. So we used these for the forward amplifiers in an ORD2-based cochlea, keeping the simple amplifier in the feedback position.

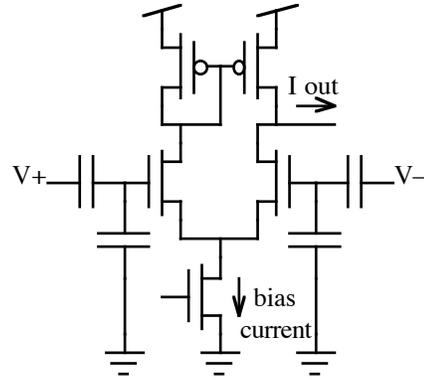


Figure 6: The circuit for the transconductance amplifier using capacitive dividers to extend the linear input range and reduce the transconductance relative to the bias current.

We were mistaken. It turns out that in this capacitively coupled circuit, a dominant effect that we ignored is the stray feedback capacitance (Miller capacitance) between the amplifier output and its inverting input. For the values we used, this reduced the open-loop amplifier voltage gain from about 500 to only about 30, resulting in a follower gain of only about 0.97, or a second-order section gain of only 0.94 at low frequencies. A cascade of these stages quickly attenuated the signal.

Some time later the usefulness of the capacitive coupling idea was revived when it was suggested, by Tobi Delbrück, that it could be used in the second amplifier of the DIF2 stage, since the loop is closed at DC by the first amplifier, which remains DC coupled. As discussed in the above subsection, the resulting DIF2 behavior was not quite right, but the idea extended nicely to the DIF3.

We built a DIF3-based cochlea cascade with this amplifier in positions 2 and 3, and did extensive characterizations on it. We were able to get gains as high as 40 dB after 36 stages, with an overall compressive characteristic, and we saw no signs of the previous tendency to burst into noise. But the signal level to which it tended to compress was only about 100 mV peak-to-peak, and at high gain settings the noise output with zero input was about 40 mV peak-to-peak. The circuit was usable but noisy over about a 40 dB input dynamic range.

At this point we revived another old idea that had been previously discarded due to its impact on the common-mode operating range of the amplifier. The idea was to interpose a pair of diode-connected transistors between the differential input transistors and the tail current transistor, as shown in figure 7. This modification results in a widening of input range and a corresponding gain reduction by a factor of about 2.5 to 3, depending on body effect. Since it is DC-coupled, this circuit works fine in the forward path of the ORD2.

This amplifier circuit can also be generalized to have any number of diodes in series, for even wider linear differential input range at the cost of a severe reduction in common-mode operating range. Lloyd Watts and Xavier Arreguit at Caltech have characterized many variations of this amplifier, as well as ORD2 circuits and cochleas based on it, which are quite well behaved.

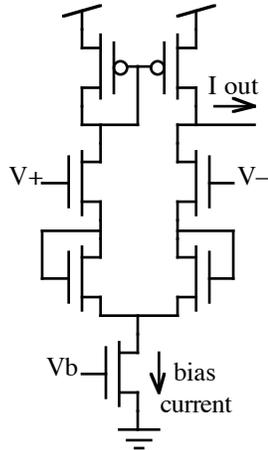


Figure 7: The circuit for the transconductance amplifier using diode-connected transistors to extend the linear input range and reduce the transconductance relative to the bias current.

At this point our best cochlea circuit uses DIF3 with two diode pairs in the second and third amplifiers, and one diode pair in the first amplifier. Due to the widened range at the first limiting amplifier, the output tends to compress toward about 250 mV peak-to-peak, and the noise is somewhat lowered due to the use of even larger transistors and capacitors. A cochlea of 168 stages fits nicely along one edge of our most recent correlogram chips, described in a later section.

AUTOMATIC GAIN CONTROL

None of our cochlea chips so far has had on-chip closed-loop gain control as hypothesized by the model. Gains have been manually adjusted, and with the new chips the instantaneous nonlinear compression does some of the job of adapting to signal levels as well. But to take advantage of what we learned in our early computer models, we really need a multi-channel coupled AGC with appropriate dynamics. Failures of early attempts are now understood in terms of the inherent poor stability properties of the old cochlea circuits. A new attempt to close the gain-control loop is underway.

In a coupled AGC, each cochlear place channel is rectified and fed back as a gain-reduction signal to not only that channel but also in lesser amounts to other nearby channels. The feedback path is smoothed through a time and space domain loop filter, in order to regulate the dynamics (attack and decay times) and the spatial spread (lateral inhibition or spread of suppression). The stability of the closed-loop system involves the phase shift of the loop filter, the added delay involved in feeding back to earlier channels (the effect propagates down the cochlea cascade at group velocity), and the loop gain. For loud attack transients, the nonlinear nature of such a loop tends to increase the loop gain,

quicken the overall time constant to the point where even a little extra delay in the loop has the potential to make it unstable [5].

We have recently tested a digital computer implementation of a new AGC loop filter structure that structurally resembles the multiple loops analyzed by Slaney [5]. But rather than cascading independent AGC stages as there, the four coupled-first-order time-space filters are run in parallel, each taking detected channel outputs, and adding their output to produce a gain-reduction signal. By choosing appropriate combining weights and time constants, it is possible to make the filter have roughly a constant 45-degree phase lag over a very wide range of frequencies while the gain drops off at about 3 dB per octave. This filter works well in computing good-looking cochleagrams and correlograms, and should be easy to get working in analog VLSI.

CORRELATOR ARRAYS

Correlograms, or moving images of arrays of correlation values indexed by cochlear place and a delay parameter, are an exciting sound representation with which we have been experimenting for a number of years [6]. While most of our experiments have been done on digital computers, we have built several arrays of analog correlation cells that each compute one pixel of a correlogram in real time. Several approaches to implementing the delays and multipliers needed in a correlator were discussed a few years ago [7]. Since that time, we have developed the surface-channel CCD analog delay line approach into a working correlogram demonstration, while Mead and his colleagues at Caltech have demonstrated yet another idea, due to Shamma—using a second (contralateral) cochlea to provide a relative delay in a binaural system [8, 9]. Lazzaro has continued to pursue designs based on pulse delay lines and simulated action potentials [10, 11, 12]. The architecture, motivation, and applications for these arrays are well discussed in the references; we focus here on implementation issues.

The key to a good correlator is a good delay mechanism. Neither the binaural “stereausis” technique used by Mead nor the “axon” pulse delay lines used by Lazzaro were judged to have adequate delay-bandwidth product or amplitude resolution, nor were they very area efficient. Charge-coupled devices provide a more ideal delay mechanism. The CCD delay line approach promises to provide the best performance for pitch detection and sound separation applications, since it has inherently good delay-bandwidth product, precisely matched delays, and continuous analog values.

We decided to base our correlator array on surface-channel CCD’s, rather than the higher-quality buried-channel ones, for several reasons. The use of surface channels results in non-ideal CCD’s, but for audio-band sample rates the transfer efficiencies are good enough. Their advantage is that unlike buried-channel devices, which can be made with the MOSIS low-noise analog process, the surface-channel devices can be built in an ordinary double-poly digital process, and can operate with on-chip 5-volt clock drivers. To debug the concept and tune up the cell design, we built a series of test chips using the cell of figure 8 and variations. For input voltages from near ground to about 1.5 V, the sense voltage follows the delayed input signal with a gain of

about one-third, which is adequate. Special on-chip clock drivers using current-starved inverters generate slow clock transitions to prevent charge dispersal. With a clock rate of 20 kHz, smearing due to charge transfer inefficiency across 70 stages is not noticeable.

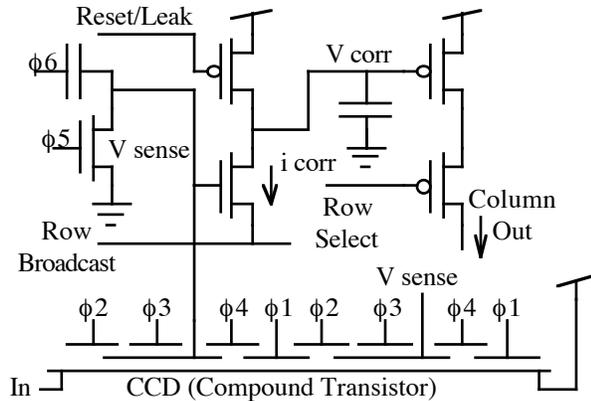


Figure 8: The circuit for two stages of four-phase CCD delay line with one correlation/integration/readout cell. The ϕ_3 clock couples capacitively to the poly1 gate that nondestructively senses the charge in the CCD channel. The correlation current being integrated on the 1 pF storage capacitor is small except during the sense interval when ϕ_6 and ϕ_3 are both high and the other clock phases are low (then it can be as high as 1 nA). If the broadcast and V sense voltages are appropriate logarithms of the input, the current i_{corr} can be an accurate one-quadrant product.

We have scaled up this design to an 84-channel by 70-lag correlator array, which fits in the MOSIS 4.6 by 6.8 mm die size using 2-micron rules. After several bugs in the video scanout timing circuits were fixed, we succeeded in demonstrating real-time moving correlograms on a video monitor. So far, however, the correlograms are redundant along one dimension, since the cochlea was not integrated onto this test chip, and all the rows were tied together.

Two new versions currently in fab combine the working correlator array with the working DIF3-diodes cochlea design and new circuits to tie them together. The circuits between the cochlea and the correlator were designed to allow flexible external control of average correlation current levels, degree of compression, degree of level-shift for the CCD charge input, etc. We expect these chips to produce high-quality correlogram displays of the sort that Licklider envisioned in 1951 [13]. What could possibly go wrong?

DISCUSSION

Lyon and Mead's 1988 cochlea chip worked, as documented in their publications, but was severely limited in its stable range of sharpness and gain tuning, and in its dynamic range of signal amplitudes. The problem addressed by that chip is still being explored: how to implement computationally intensive auditory processing in silicon, taking advantage of the power and area efficiencies of a subthreshold analog MOS approach, while working around its inherent limitations. Concurrently, we have

continued to explore digital implementations using custom silicon as well as standard programmable architectures and radically-parallel programmable architectures. Only the analog approach currently appears to have the possibility of (a) putting a cochlea and a correlator array on a single chip, and (b) operating them at a power low enough to consider for a battery-powered portable device.

The correlation arrays discussed above contain 5880 cells, each containing a one-transistor multiplier and two analog state-variable storage sites. Clocking at 40 kHz, the equivalent memory bandwidth is 470.4 million read-modify-writes per second. The equivalent floating-point operation rate is about the same. Digital versions on our Cray X/MP take advantage of block processing and fast transform algorithms to bring these numbers down to where they operate in near real time, but making a one-chip custom digital version is still too difficult. Using the analog approach, expanding the chip to 100 channels by 300 lags will be no problem using 1.2-micron CMOS rules and slightly larger chips.

Interpreting the correlograms is the next challenge. Distributing the next level of processing over the correlation array is a reasonable direction to keep real-time performance.

REFERENCES

1. Lyon, R.F., and Mead, C. "An Analog Electronic Cochlea," *IEEE Trans. ASSP* **36**: 1119–1134, 1988.
2. Lyon, R.F., and Mead, C. "Cochlear Hydrodynamics Demystified," Caltech Computer Science Technical Report Caltech-CS-TR-88-4, Pasadena, CA, 1988.
3. Lyon, R.F. "Automatic Gain Control in Cochlear Mechanics," *Mechanics and Biophysics of Hearing*. (June 1990 Madison workshop proceedings) In press.
4. Mead, C., *Analog VLSI and Neural Systems*. Addison-Wesley, Reading MA, 1989.
5. Slaney, M., "Lyon's Cochlear Model," Apple Technical Report #13, Apple Computer, Cupertino CA, 1988.
6. Lyon, R.F., "Computational Models of Neural Auditory Processing," *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 36.1.1–4, San Diego, March, 1984.
7. Lyon, R.F., "Analog VLSI Hearing Systems," in Brodersen and Moscovitz (eds.), *VLSI Signal Processing III* (workshop proceedings) pp. 244–251, IEEE Press, 1988.
8. Shamma, S.A., Shen, N., and Gopalaswamy, P., "Stereausic: Binaural processing without neural delays," *J. Acoust. Soc. Am.* **86**: 989–1006, 1989.
9. Mead, C., Arreguit, X., and Lazzaro, J., "Analog VLSI Model of Binaural Hearing," *IEEE Trans. Neural Networks*, in press.
10. Lazzaro, J. and Mead, C., "A silicon model of auditory localization," *Neural Computation* **1**: 41–70, 1989.
11. Lazzaro, J. and Mead, C., "Silicon models of auditory localization," in Zornetzer, Davis, and Lau (eds.), *An Introduction to Neural and Electronic Networks*. New York: Academic Press, pp. 158–174, 1990.
12. Lazzaro, J. and Mead, C., "Silicon modeling of pitch perception," *Proceedings National Academy of Sciences* **86**: 9597–9601, 1989.

13. Licklider, J.C.R., "A Duplex Theory of Pitch Perception," *Experientia*
7: 128–133, 1951.