Auditory Feature Extraction Using Self-Timed, Continuous-Time Discrete-Signal Processing Circuits

Nagendra Kumar, *Student Member, IEEE,* Gert Cauwenberghs, *Member, IEEE,* and Andreas G. Andreou, *Member, IEEE*

***Abstract*—A compact integrated subsystem for accurate real-time measurement of level-crossing time-intervals, suitable for multiresolution feature extraction from an analog cochlear filter bank is presented. The subsystem is inspired by the function of the inner hair cells in the mammalian cochlea and is based on continuous-time discrete-signal processing circuits. Experimental results from a fabricated array of nine elements demonstrate instantaneous frequency-to-voltage conversion over a range covering the audio band. The power consumption is less than 20 μW per cell from a 5-V supply, when the system is biased to operate over the speech frequency range.**

***Index Terms*—Analog integrated circuits, neural network hardware, very-large-scale integration.**

I. Introduction

VERY LARGE SCALE INTEGRATION signal processing systems are often classified into analog, digital or mixed-mode. With an emphasis on low power real-time VLSI, there has been an intense discussion as to how much processing should be done in analog and how much in digital to achieve optimum performance [1]. Missing in most of these discussions is the fact that it is the design of the algorithm that gives an advantage to either analog or digital implementation. Certain algorithms map well to analog, while others map well to digital hardware.

There is another class of signal processing algorithms [2]–[6] that requires an entirely new hardware design paradigm. These algorithms necessitate an *event* based, asynchronous signal processing approach where the signal can take only discrete values, but is continuous in time. For example, signal could be quantized to two discrete levels, and change value at the event of a zero crossing. This paradigm is called *continuous-time discrete-signal* (CTDS) signal processing. It is a mixed-mode approach whereby algorithms exploit the robustness of discrete signal representations but preserve the continuity of events in the time domain. CTDS signal processing can only be approximated in a digital implementation by using a fast clock and thus *oversampled* in time. By using event based computation, unnecessary power dissipation can be eliminated by avoiding high-speed global clocks and associated unnecessary switching events.

The signal representation, and the system organization that follows from it, is similar to self-timed asynchronous digital design methodologies [7] and also with the address event representation (AER) [8] for interchip communication [9], [10]. A small CTDS system for centroid computation of visual stimuli was also presented in [11]. An analog VLSI event based system for speech processing and feature extraction has also been reported in [12]. However, in the latter system, a digital clock is used to provide time-stamp to each zero crossing and the spectral shape is extracted by relying on tuning characteristics of the cochlear filters.

There is evidence that biological systems employ an analogous representation [13] that is natural when computation must rely on individual components (neurons) that have limited intrinsic bandwidth, have a limited dynamic range and operate without global clocks. In particular, the location of zero-crossings of a signal in the time domain reveals much of its characteristics in the spectral domain [14]. An appropriately smoothed version of the time interval between consecutive zero crossings yields the inverse of a dominant frequency present in the signal [2]. Inner-hair cells attached to the basilar membrane [15], [16] are believed to encode the formant and tone information from audio signals through such zero-crossing intervals. Neural models of auditory processing in the inner-hair cells using level-crossing signal representations have been presented in the literature [5], [6]. In Ghitza's model [5], output from time interval measurements between level crossings are aggregated across cochlear channels to produce an ensemble interval histogram (EIH) spectral measure that has robust properties in the presence of noise.

This paper presents a compact subsystem for accurate and real time measurement of level-crossing time-intervals. The subsystem is suitable for multiresolution feature extraction from an analog cochlear filter bank. The CTDS approach deviates from the standard approach where audio signals are digitized by an analog to digital converter and the signal processing is performed by a specialized digital signal microprocessor [17]. The inputs of the subsystem are analog as it interfaces to a silicon cochlea [18], [19] and the outputs are also analog as they will be subsequently processed by

Manuscript received June 15, 1995; revised August 16, 1996. This work was supported by the National Science Foundation under Grant ECS-9313934 (P. Werbos, Program Monitor), by the Center for Language and Speech Processing at Johns Hopkins University, and by Lockheed-Martin Corporation. This paper was recommended by Associate Editor S. Kiaei.

The authors are with the Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, MD 21218 USA.

Publisher Item Identifier S 1057-7130(97)06581-6.

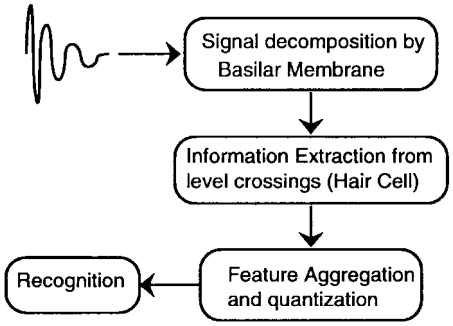


Fig. 1. Neuromorphic architecture for sensory preprocessing. Unlike the more traditional approach, the quantization of signal occurs at a much higher level and thus a single chip preprocessor replaces functions traditionally performed by the A/D converters and specialized digital signal microprocessors (DSP chips).

feature aggregation and dimensionality reduction functional blocks [20] followed by a vector quantizer.

An array of nine basic cells, each of which can be tuned for a given range of frequencies corresponding to one cochlear output channel, has been fabricated and tested. The area for the basic cell is (120 μm $\times$ 227 μm) and thus, allows for ultimately dense integration of a complete auditory-based signal processing subsystem. The output of the subsystem produces a voltage proportional to the time interval between consecutive upward crossings of the input signal with respect to a reference level, sampled at the end of every interval.

II. Circuit Description

The basic functional elements of the circuit cell are shown in Fig. 2. Essentially, the cell integrates a constant supplied current I_d onto capacitor C_1 over the time interval between consecutive upward level crossings. The capacitor voltage on C_1 is sampled and held onto C_2 at the end of the interval, while almost simultaneously resetting the capacitor voltage on C_1 for integration in the next cycle. Thus, at any time the output voltage (V_{out}) is a measure of the most recent level crossing interval. The comparator serves to indicate the location in time and the polarity of the level crossings. The most complicated part of the circuitry is the self-timed control logic which ensures that the capacitor voltage is reset *after* the output is sampled and held. The sampling and resetting should occur in a very short time interval at every upward level crossing. The time needed to sample and reset the capacitor voltage will ultimately limit the maximum frequency of time-to-voltage conversion. The different components are now described below.

A. The Comparator Circuit

The comparator circuit is shown in Fig. 3. It consists of a standard two-stage differential input CMOS amplifier, without frequency compensation. Although this circuit topology is typically used in above threshold MOS designs, transistors M_1 and M_7 are biased in the subthreshold region. Small-signal analysis of the comparator, using device sizing of Fig. 2 and typical values for the early voltage in a 2 μm process, yields a dc voltage gain of the order of 3×10^5. However, a large

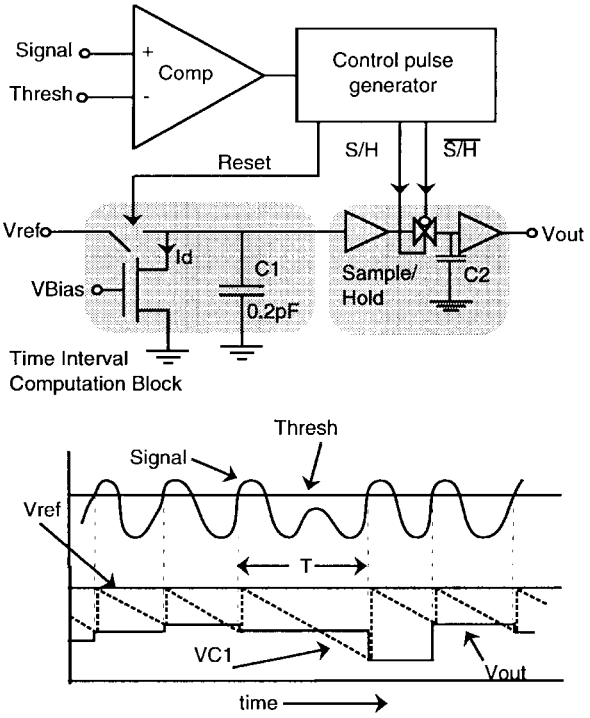


Fig. 2. Block diagram for a single cell: time intervals are computed by integrating a constant current on a capacitor.

signal analysis reveals a strong asymmetry in the transient behavior. A much larger output current results from M_6 (which is driven above threshold when the signal is high) compared to the fixed subthreshold current supplied by M_7. Therefore, the rise time of the comparator is several orders of magnitude smaller than its fall time, which is slew-rate limited. This asymmetry is used to create a refractory period in the response of the comparator, during which no second positive transition can be registered. It plays an equivalent role to refractory period in a neuron, avoiding spurious transitions due to noise. The duration of the refractory period can approximately be written as

$$T_{\text{refract}} = \frac{C_{\text{load}}(V_{DD} - V_{SS})}{2I_{M_7}} \tag{1}$$

where C_{load} is the load capacitance (mostly parasitics) at the output of the comparator, V_{DD} and V_{SS} are the supply voltages, and I_{M_7} is the drain current in the transistor M_7. The value for the refractory period cannot exceed the least time separation between consecutive positive transitions. For audio applications, the least time interval of interest is of the order of 25 μs. With an estimated parasitic load capacitance of 40 fF on the Cmp output node, it would require the drain current of M_7 to be 8 nA, which necessitates subthreshold operation for a square MOS device in 2-μm technology.

B. Self-Timed Control Pulse Generator

The circuit for generating the sample-and-hold (S/H) and reset pulses is shown in Fig. 4. A time delay element is realized using the transistors M_8 to M_{11}. Transistors M_8 and M_{11} are biased in subthreshold, and can set the rise and fall times of the output from a fraction of a microsecond to several

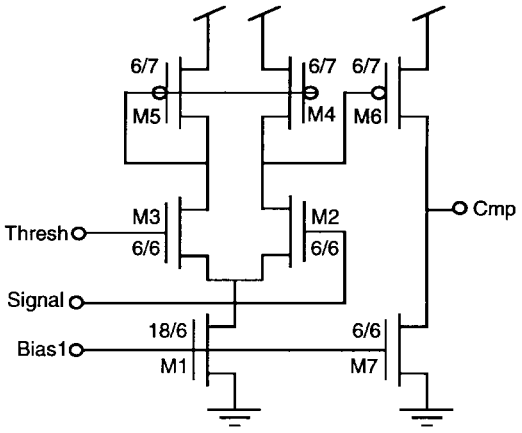


Fig. 3. Comparator circuit: A two-stage amplifier operated in an open loop configuration.

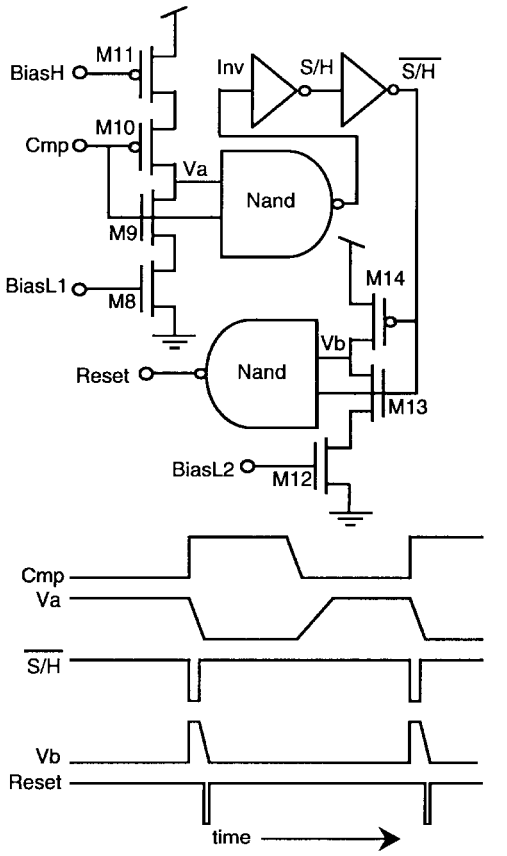


Fig. 4. Control pulse generator: The bias voltage Bias_H controls the rising slope and Bias_{L_1}, Bias_{L_2} control the falling slope, thus, creating a time delay element.

milliseconds as controlled by the bias voltages Bias_H and Bias_{L_1}. Because of this delay, the NAND gate is active low during the upward transition of the input Cmp. At that instant, a pulse is generated with a width approximately equal to the fall-time delay in the delay element.

As noted above, the comparator implementation shown in Fig. 3 has a very short rise-time but a fairly long fall time. The rise time of the delay element is chosen to be longer than the comparator fall time, to avoid a spurious pulse at the output of the NAND at the falling edge of the comparator output. The fall time, which is controlled by Bias_{L_1} at the gate of M_8 sets the width of the S/H pulse.

The circuit for generating the reset pulse is essentially identical to that just described for generating the S/H pulse, with the exception that the second delay element is configured for no delay in the upward transition of the input. The configuration ensures that a reset pulse is generated even for very short duration sample-and-hold pulses. The bias voltage Bias_{L_2} at the gate of the transistor M_{12} controls the width of the reset pulse.

C. Sample-and-Hold

The sample-and-hold (S/H) circuit in Fig. 2 is implemented in standard form, using an input voltage buffer, a switch, a hold capacitor C_2, and an output voltage buffer. The dummy-compensated complementary switch is driven by the S/H and $\overline{\text{S/H}}$ control signals. Presently, the buffers are implemented as differential transconductance amplifiers with unity-gain feedback, and are hence slew-rate limited. For a nonstationary input the output voltage not only depends on the voltage on C_1 when the S/H pulse arrives, but also on the slew rate of the first buffer in the presence of the load C_2. In case the interval between level-crossings changes significantly from period to period, slew rate may be a limiting factor in the performance of the circuit. In particular, the maximum change in output voltage between consecutive periods is given by

$$\Delta V_{\text{out}}^{\max} = T_{\text{S/H}} \frac{I_{\text{Bias}}}{C_2} \tag{2}$$

where $T_{\text{S/H}}$ is the duration of the sample-and-hold pulse, and I_{Bias} is the bias current in the S/H transconductance buffer.

For a quasi-periodic input, the steady state output voltage V_{out}, relative to the reset voltage V_{ref}, is given by the discharge current I_d integrated on capacitor C_1 over the level crossing time interval ΔT. Ignoring the output conductance of the current source I_d and the finite width of the reset pulse, V_{out} is approximately given by

$$V_{\text{out}} = V_{\text{ref}} - \Delta T \frac{I_d}{C_1} + V_{\text{off}} \tag{3}$$

where V_{off} is the offset voltage of the sample-and-hold output circuit. In the current design, I_d is the drain current of a MOS transistor in subthreshold. The value of the current I_d is controlled by the gate voltage V_{Bias} of that transistor. Therefore, I_d is an exponential function of V_{Bias}.

$$I_d \simeq I_0 e^{\kappa(V_{\text{Bias}}/V_T)}. \tag{4}$$

Thus, for an array of level crossing circuits, if V_{Bias} are linearly spaced from channel to channel (e.g., through a resistive polysilicon wire), then the operating center frequencies of such an array will be linearly spaced on a log-scale. This is desirable for multiresolution signal analysis [3].

Another useful characterization of the circuit is the maximum bandwidth F_B (in decades) for which the cell produces a useful output. The constraints result from power supply limitations and the resolution of the measurement equipment.

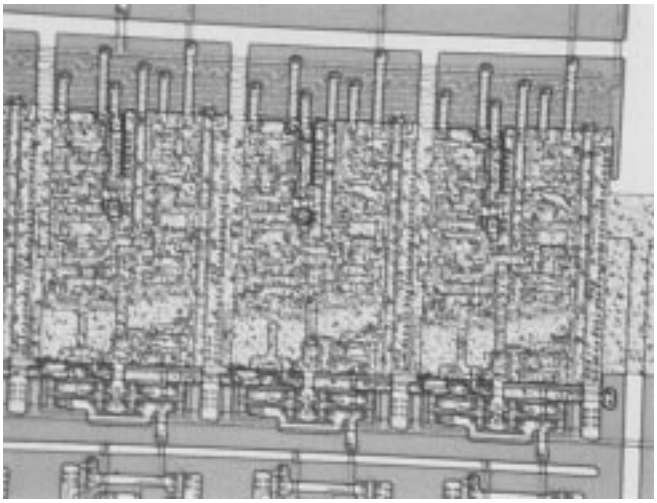


Fig. 5. Photomicrograph of three zero crossing cells on the chip. The cell dimension is 120 μm $\times$ 227 μm each, fabricated in N-well 2-μm CMOS technology.

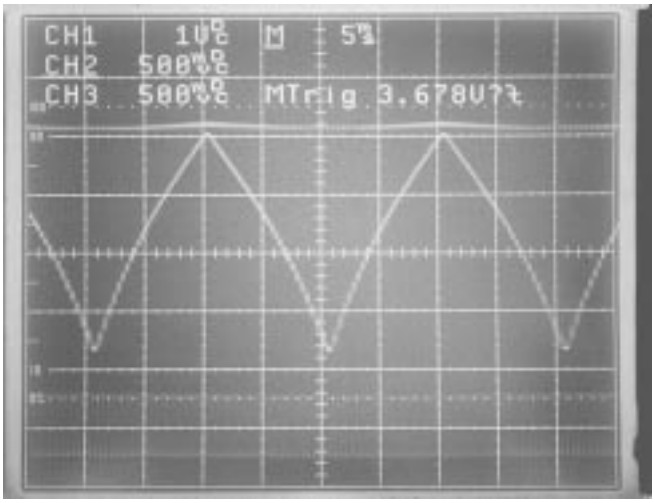


Fig. 6. Output waveform for a frequency modulated input—The lower waveform is the frequency modulated input, upper waveform (the straight line and thin vertical lines) show the resetting and discharging of the capacitor, and the triangular wave is the final output.

Ignoring V_{off} in (3), one can write

$$\Delta T = \frac{(V_{ref} - V_{\text{out}})C_1}{I_d}. \tag{5}$$

Taking frequency as reciprocal of the period, and dividing (5) for the maximum and minimum values of V_{out} gives

$$F_B = \log \frac{f_{\max}}{f_{\min}} = \log \frac{V_{ref} - V_{\text{out}}^{\min}}{\max[(V_{\text{ref}} - V_{\text{out}}^{\max}),\, V_\Delta]} \tag{6}$$

where the minimum and maximum output voltages $V_{\text{out}}^{\min}$ and $V_{\text{out}}^{\max}$ are determined by the power supply and the range of operation for the S/H, and V_Δ is the resolution limit due to the noise in the circuit and the measurement equipment.

III. Experimental Results

An array of nine circuit cells has been fabricated through MOSIS. A micrograph of the chip, is shown in Fig. 5. Common control signals are provided to all cells, except for the bias voltage V_{Bias}. Bias voltages are tapped from equally spaced points on a polysilicon wire that is used as a resistive voltage divider. Voltages at the two ends of the polysilicon wire may be controlled externally, thus locally controlling the discharge current I_d in the individual cells.

A. Temporal Response

Fig. 6 illustrates the response of a single cell when a frequency modulated signal is applied to the input. An input with minimum frequency of 1 kHz and a maximum frequency of about 4 kHz is used, with a triangular modulation at 50 Hz. It can be seen that the output voltage is nearly triangular, with a value proportional to the input frequency.

B. Tuning Characteristics

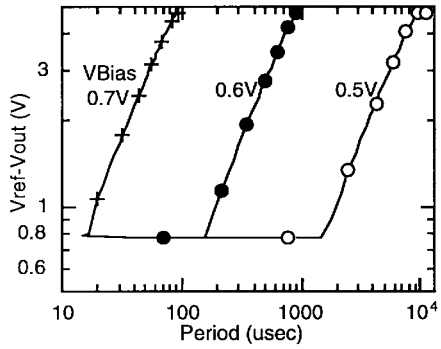
From (3), the relationship between ΔT and V_{out} is expected to be linear, and slope proportional to the discharge current I_d. Plots of the output voltage V_{out} versus the period ΔT of a periodic input signal are shown in Fig. 7(a), for three different values of V_{Bias} (see Fig. 2). As shown, the bias voltage setting allows to scale the linear frequency versus output voltage response over a wide range of frequencies. The output sensitivity S is defined as the change in output voltage V_{out} in response to a unit change in period ΔT. From (3), the quantity S is equal to I_d/C_1. The recorded output sensitivity as a function of bias voltage V_{Bias} is shown in Fig. 7(a). The exponential relationship between the S and V_{Bias} derives from the fact that in subthreshold region of operation, the saturation drain current I_d of a MOS transistor is exponential in the gate voltage V_{Bias}.

C. An Array of Zero Crossing Cells

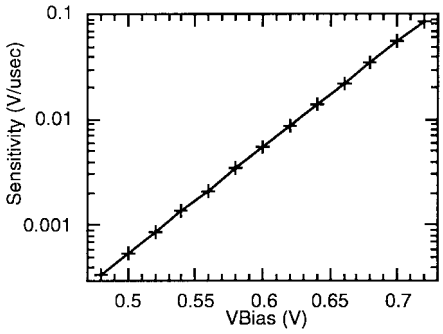
By applying a uniform linear voltage gradient across the polysilicon wire that provides V_{Bias}, an exponential distribution in bias current I_{Bias} of the cells is obtained. Using this arrangement, each cell can be tuned to a particular range of frequencies in the corresponding input signal, whereby the center frequencies of the cells are spaced uniformly in the log frequency domain. The motivation is to interface the array of level crossing cells directly with outputs from a cochlear filter bank with matched center frequencies. As a proof of concept, a linear gradient in V_{Bias} is constructed by tapping the bias voltages on equally spaced points along a resistive polysilicon line. The maximum and the minimum values of V_{Bias}, defining the corner frequencies, are applied at the ends of the resistive line. A frequency sweep in the audio range is input to the whole array. The resulting outputs are shown in Fig. 8. The useful frequency range of each cell is about a decade, suitable for use with cochlear filter banks which typically have bandwidths less than an octave per channel.

D. Power Consumption

The total power consumption can be broadly divided into two separate components. The first component is fixed. It



(a)



(b)

Fig. 7. (a) Output voltage versus input signal period, at three different discharge current bias settings and (b) output sensitivity versus bias voltage.

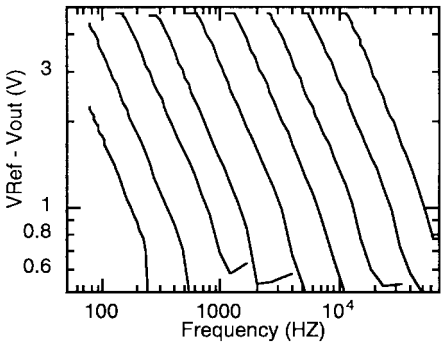


Fig. 8. Frequency-to-voltage characteristics of the transducer array over the audio frequency range.

is due to the constant bias currents in the comparator, the time-interval computation block, and the sample-and-hold. The value for these bias currents is determined by applying the criterion that the circuit should function in the worst case scenario. The second component of power consumption is due to the switching in the control-pulse generator. The main component of this dynamic part is the power consumption due to short circuit current generated from the slow transition of the time delay element. Since the amount of switching is proportional to the input frequency, this component is directly proportional to the average input frequency. Thus the total power consumption can be written as

$$P = \alpha + \beta f. \tag{7}$$

For power measurement, the comparator is biased such that it continues to function properly for input frequencies up to

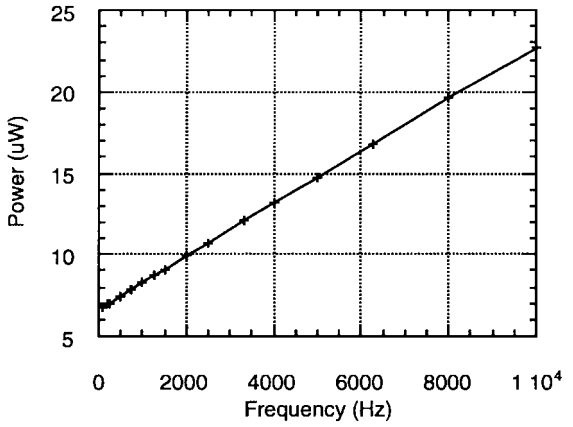


Fig. 9. Power consumption of the fabricated cell with a 5-V power supply.

40 kHz. The bias current in the S/H is set large enough to allow instantaneous changes in level-crossing-intervals to be recorded. The experimental data for power consumption is shown in Fig. 9. The estimated values of α and β are 7.4 μW and 1.64 nW/Hz, respectively. For the speech input frequency range (less than 8 kHz) the power consumption is less than 20 μW.

IV. Conclusion

An integrated circuit, small system design for an event driven paradigm in signal processing has been proposed and demonstrated experimentally. This is a compact and low-power quasi-analog VLSI system for real time level-crossing time interval measurement. Subthreshold exponential characteristics of the MOS transistor that controls tuning and real-time operation make arrays of these cells particularly suitable to multiresolution signal processing based on models of the mammalian cochlea. The time interval computation block of Fig. 1 could be replaced by an "average computation" to get the stabilized zero-crossing representation [3] or by the "spatial derivatives" [21] to model the lateral inhibition in neural networks.

Power consumption and robustness are key issues in the development of speech-processing devices for portable applications. A compact VLSI circuit has been presented that exploits the robustness [2] of zero-crossing-based signal processing at a very low power cost. The power consumption of the circuit is small because of two reasons. First, since the application involves audio-frequencies, is has been possible to use inherently low power subthreshold CMOS circuits. Second, due to the asynchronous self-timed design, unnecessary switching has been eliminated.

Acknowledgment

Chip fabrication was provided by MOSIS.

References

[1] E. Vittoz, "Analog VLSI signal processing: Why, where and how?" *J. Analog Integrated Circuits and Signal Processing*, pp. 27–44, June 1994.
[2] B. Kedem, "Spectral analysis and discrimination by zero-crossings," *Proc. IEEE*, vol. 74, pp. 1477–1493, Nov. 1986.

[3] S. Mallat, "Zero-crossings of a wavelet transform," *IEEE Trans. Inform. Theory*, vol. 37, July 1991.
[4] J. J. Benedetto and A. Teolis, "A wavelet auditory model and data compression," *Appl. Computat. Harmonic Anal.*, vol. 1, no. 1, pp. 3–28, 1993.
[5] O. Ghitza, "Auditory nerve representation as a front-end for speech recognition in a noisy enviornment," *Comput. Speech and Lang.*, vol. 1, pp. 109–130, 1986.
[6] K. Wang and S. Shamma, "Zero-crossings and noise supression in auditory wavelet transformations," Tech. Rep., Univ. Maryland, Dept. Elect. Eng., Aug. 1992.
[7] S. Hauck, "Asynchronous design methodologies: An overview," *Proc. IEEE*, vol. 83, pp. 69–93, Jan. 1995.
[8] J. Lazzaro, J. Wawrzynek, M. Mohwald, M. Sivilotti, and D. Gillespie, "Silicon auditory processors as computer peripherals," *IEEE Trans. Neural Networks*, vol. 4, pp. 523–528, May 1993.
[9] K. A. Boahen, "Retinomorphic vision systems," *MicroNeuro*, pp. 2–14, Feb. 1996.
[10] Z. Kalayjian, J. Waskiewwicz, D. Yochelson, and A. G. Andreou, "Asynchronous sampling of 2D arrays using winner-takes-all arbitration," in *Proc. ISCAS*, 1996, pp. 393–396.
[11] T. G. Edwards and A. G. Andreou, "VLSI phase-locking architectures for feature linking in multiple target tracking systems," in *Advances in Neural Information Processing Systems.* San Mateo, CA: Morgan Kaufmann, 1994, vol. 6, pp. 866–873.
[12] J. P. Lazzaro, J. Wawrzynek, and A. Kramer, "System technologies for silicon auditory models," *IEEE Micro*, vol. 14, pp. 7–15, June 1994.
[13] M. Abeles, *Corticonics: Circuits of the Cerebral Cortex.* Cambridge, U.K.: Cambridge Univ. Press, 1991.
[14] B. F. Logan, Jr., "Information in the zero crossings of bandpass signals," *Bell Syst. Tech. J.*, vol. 56, no. 4, pp. 487–510, Apr. 1977.
[15] J. B. Allen, "Cochlear mechanics—A physical model of transduction," *J. Acoust. Soc. Amer.*, vol. 68, no. 6, pp. 1660–1670, Dec. 1980.
[16] M. S. Sachs and E. D. Young, "Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate," *J. Acoust. Soc. Amer.*, vol. 66, no. 2, pp. 470–479, 1979.
[17] L. R. Rabiner, "Applications of voice processing to telecommunications," *Proc. IEEE*, vol. 82, pp. 197–228, Feb. 1994.
[18] W. Liu, A. G. Andreou, and M. G. Goldstein, "Voiced speech representation by an analog silicon model of the auditory periphery," *IEEE Trans. Neural Networks*, vol. 3, pp. 477–487, May 1992.
[19] P. M. Furth and A. G. Andreou, "Linearized differential transconductors in subthreshold CMOS," *Electron. Lett.*, vol. 31, no. 7, pp. 545–547, 1995.
[20] N. Kumar, C. Neti, and A. G. Andreou, "Application of discriminant analysis to speech recognition with auditory features," in *Proc. 15th Annu. Speech Res. Symp.*, Johns Hopkins Univ., Baltimore, MD, June 1995, pp. 153–160.
[21] K. A. Boahen and A. G. Andreou, "A contrast sensitive silicon retina with reciprocal synapses," in *Advances in Neural Information Processing Systems*, J. E. Moody, S. J. Hanson, and R. P. Lippmann, Eds. San Mateo, CA: Morgan Kaufmann, 1992, vol. 4, pp. 764–772.

Nagendra Kumar (S'91) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Kanpur, India, in 1989, and the M.S. degree in electrical engineering from the Johns Hopkins University, Baltimore, MD, in 1991, and is currently working towards the Ph.D. degree in the Department of Electrical and Computer Engineering at The Johns Hopkins University.

His research interests include VLSI circuits, signal processing systems and algorithms, machine learning, and speech recognition.

Gert Cauwenberghs (S'89–M'92) received the Engineer's degree in applied physics from the Vrije Universiteit Brussel, Belgium, in 1988, and the M.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology, Pasadena, in 1989 and 1994, respectively.

In 1994, he joined The Johns Hopkins University as an Assistant Professor in the Department of Electrical and Computer Engineering. His research covers VLSI circuits, systems and algorithms for parallel signal processing, and adaptive neural computation.

Andreas G. Andreou (S'80–M'81) received the M.S.E. and Ph.D. degrees in electrical engineering and computer science from The Johns Hopkins University, Baltimore, MD, in 1983 and 1986, respectively.

From 1987 to 1989, he was a Post-Doctoral Fellow and Associate Research Scientist at Johns Hopkins where he became Assistant Professor in 1989, Associate Professor in 1993, and Professor in 1997. During 1995–1996, he was on sabbatical leave as a Visiting Associate Professor in the Computation and Neural Systems Program, California Institute of Technology, Pasadena. His research interests are in the areas of integrated circuits, speech recognition, and neural computation.

Dr. Andreou is a member of Tau Beta Pi.