

# An Analog VLSI Chip with Asynchronous Interface for Auditory Feature Extraction

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

**Abstract**—We present an analog VLSI chip intended to serve as a front end of a speech recognition system. The chip architecture is inspired by biological auditory models common to humans and primate vertebrates. We include experimental results on a 1.2- $\mu\text{m}$  CMOS custom analog VLSI implementation and speech recognition results obtained from software simulations of the hardware on the TI-DIGITS database.

**Index Terms**—Analog VLSI, neural networks, speech recognition.

## I. INTRODUCTION

**H**UMAN performance in speech recognition tasks is superior to that of the state-of-the-art speech recognition systems. This is especially true under adverse conditions, such as noisy environments or when speech is transmitted through a telephone channel. It is hypothesized that the specific characteristics of the human auditory periphery may play an important role in the robustness of human speech perception. A significant amount of research has been performed to gain an understanding of basic signal processing steps in a mammalian cochlea [1]–[4].

It has also been demonstrated that feature extraction based on computational models of auditory processing leads to a signal representation that is more robust for speech recognition [5], [6]. However, because of their inherent complexity, application of auditory models to real systems poses a significant engineering challenge. For most applications, a system is constrained to be real time, low power, and low cost. However, as indicated by Jankowski [6], it takes 120 times the real time to compute auditory features on a general purpose workstation. Therefore, we should find other methods of computing auditory features.

Manuscript received October 21, 1997; revised March 13, 1998. This work was supported in part by the Center for Language and Speech Processing, Johns Hopkins University, by Lockheed Martin, by ONR/DARPA MURI N00014-95-1-0409, by NSF CAREER MIP-9702346, and by the National Science Foundation Neuromorphic Engineering Center at Caltech. This paper recommended by Guest Editors F. Maloberti and W. C. Sui.

N. Kumar was with the Center for Language and Speech Processing, The Johns Hopkins University, Baltimore, MD 21218 USA. He is now with Telocity Networks Inc., Germantown, MD 20874 USA (e-mail: nkumar@telogy.com).

W. Himmelbauer was with the Center for Language and Speech Processing, The Johns Hopkins University, Baltimore, MD 21218 USA. He is now with Micro Linear Corporation, San Jose, CA 95131 USA (e-mail: himmelba@engmail.ulinear.com).

G. Cauwenberghs and A. G. Andreou are with the Center for Language and Speech Processing, The Johns Hopkins University, Baltimore, MD 21218 USA (e-mail: gert@bach.ece.jhu.edu; andreou@olympus.ece.jhu.edu).

Publisher Item Identifier S 1057-7130(98)03961-5.

Since the seminal work of Lyon and Mead [7], several research groups have implemented analog subthreshold circuits and systems that emulate early auditory functions [8], [9], [11]–[13]. An important benefit of using analog circuit techniques for speech processing is very low power consumption and real-time operation.

We have adopted this approach to develop analog VLSI hardware for auditory-based feature extraction. The extracted features are the signal energies and zero-crossing time intervals obtained on the frequency-decomposed output channels in a cochlear filter bank. The system presented in this paper is intended as a demonstration vehicle toward a low-power real-time robust speech recognizer for portable applications. The paper is organized as follows. Section II briefly reviews the human auditory periphery and its signal processing function. Section III describes the implemented VLSI architecture and reports on preliminary chip testing results. In the final section, we describe the structure of a speech recognition system that uses auditory features, and we report recognition results for a digit recognition task.

## II. AUDITORY SIGNAL PROCESSING

Sound waves that reach the eardrum are mechanically transferred to the cochlea, which is a fluid-filled chamber partitioned by the basilar membrane (BM), illustrated in Fig. 1. The mechanical vibrations create standing waves in the cochlear chamber that cause the BM to vibrate at frequencies corresponding to the incident acoustic wave frequency. For each frequency value, there exists a location along the BM where the vibration is strongest. These locations roughly follow logarithmic ordering in frequency along the BM. Hence, in generic form, the BM can be modeled as a bank of frequency-selective filters, shown in Fig. 2, where the center frequency of each filter is equally spaced on a log scale, each representing a particular location equally spaced along the membrane [1], [17]. It has been shown that such filter-bank representation is equivalent to a wavelet analysis [10], [19].

The mechanical vibration along the BM is sensed by the inner hair cells (IHC's) that constitute the axon roots of the auditory nerve-fiber bundle. Each IHC is connected to about ten nerve fibers that differ in the motion level of the BM at which they fire. Beyond the auditory nerve, our understanding of biological processing is almost primitive. However, constituting the only signal input to the cortex, the nerve-firing patterns must contain all the information relevant for recognition. Therefore, an auditory-based feature extraction

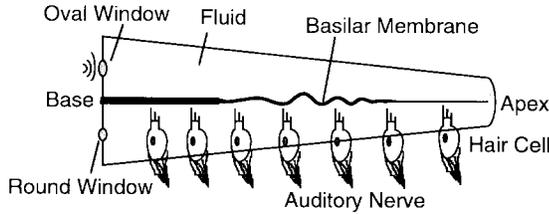


Fig. 1. A schematic diagram of the cochlea showing the main signal processing components.

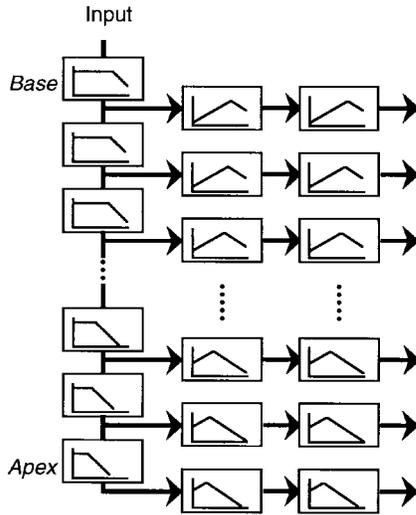


Fig. 2. BM can be modeled as a bank of filters, each with different center frequency. The center frequencies of the filterbanks are uniformly distributed on a logarithmic scale.

algorithm used for speech recognition must be capable of capturing important specifics of the firing patterns, which are hypothesized to be partly responsible for the robustness of human auditory.

The discrete-action potentials generated by the IHC and transmitted through nerve fibers to the cochlear nucleus in response to an auditory stimulus can be considered as zero-crossing events of the BM velocity [15]. This is especially true at medium sound-pressure levels, such as in a typical office environment. It has been shown analytically that the encoding of complex signals, such as natural speech, by the zero-crossing rates of its wavelet transform (in this case, performed by the BM filter bank) provides a robust representation. In particular, the formant phase locking—the property that the hair cells tend to fire in phase with the dominant frequencies of the input signal—is believed to introduce spectral enhancement and noise robustness [16]–[19]. Moreover, zero-crossing rates detected in spectral subbands are ideal for the fast detection of spectral changes [20].

Fig. 3 defines the term zero-crossing interval or the *instantaneous* zero-crossing rate. The depicted waveform is the output of a single BM filter that corresponds to some particular location along the BM. We define  $T_{ZC}$  as the time interval between two consecutive upward zero crossings in the ac component of the signal. In order to account for different fibers firing at different motion levels of the same output, we also compute an *energy* measure, which we define as the

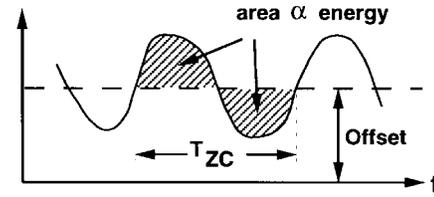


Fig. 3. Information coding by zero-crossing intervals and period energy.

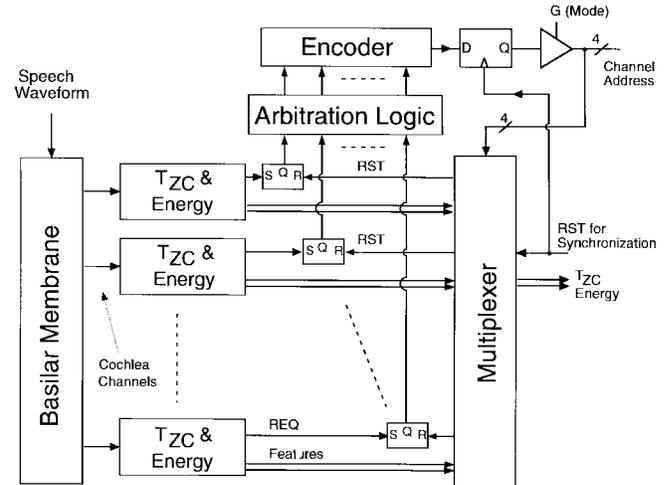


Fig. 4. Block diagram of the VLSI architecture for the electronic cochlea.

integral over the rectified ac component of the signal, within the period  $T_{ZC}$ .<sup>1</sup> The zero-crossing interval and the signal energy for the corresponding interval of the wavelet transform constitute a complete signal representation [21]. Also, since the accuracy in computing  $T_{ZC}$  depends critically on the accuracy of only a few components, it is easier to carefully design a reliable circuit cell for performing this computation [22]. Therefore, due to its physiological plausibility and powerful signal processing capabilities, we adopt a zero-crossing-based signal representation for abstracting the auditory nerve response.

### III. VLSI CHIP ARCHITECTURE AND CIRCUITS

The implemented hardware system emulating the auditory periphery includes both a model of frequency decomposition in the BM of the inner cochlea and a model of feature extraction in the inner hair cells of the cochlea.

Fig. 4 shows a block diagram of the auditory signal processing chip. Following the architecture proposed by Liu, we implement the BM as a filter-bank structure (as shown in Fig. 2), each segment of which consists of a linear first-order low-pass filter, followed by two linear bandpass filter sections [9]. The filter bank is tuned to frequencies spaced uniformly on a logarithmic scale, from 100 to 8000 Hz [9], [12], [13]. This range corresponds to the spectrum covered by speech sounds. The BM is implemented as a 15-segment filter-bank structure, each segment of which consists of multiple linear first-order sections followed by two linear bandpass filter

<sup>1</sup>The definition of an “energy” feature is short of rigorous physical meaning. It should be interpreted as a measure of signal strength or signal “power.”

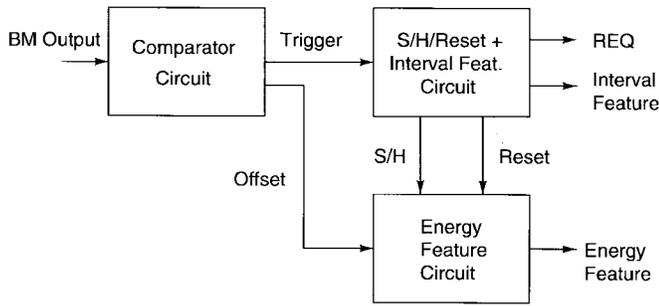


Fig. 5. Zero-crossing interval and energy feature computation block ( $T_{ZC}$  and  $Energy$ ).

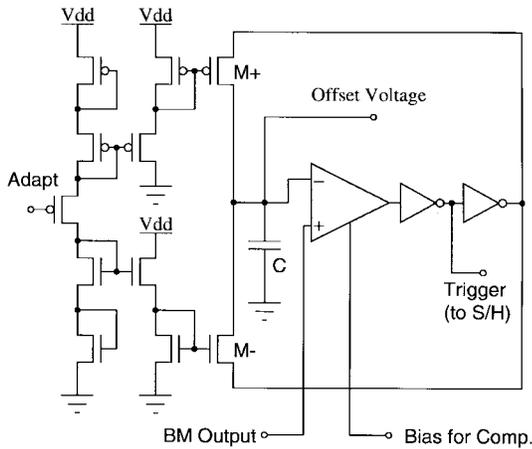


Fig. 6. Autoadaptive comparator for detecting zero crossing.

sections [9]. The filters are based on linearized transconductors developed by Furth [23]. For maximum power efficiency, their MOS devices are biased in below-threshold operation [24], [27]. The frequency-decomposed time signals from the BM are then processed locally to obtain a representation for the auditory-nerve firings. We employ a binary charge pump to establish an adaptive elimination of signal offsets. We use the same comparator, that provides the control signal to the binary charge pump, to detect the upward zero crossing and to provide control signals for circuit computing  $T_{ZC}$  [22]. The energy feature is obtained from integrating the full-wave rectified and threshold-adjusted signal on a capacitor.

The outputs of the BM are input to a feature computation block ( $T_{ZC}$  and  $Energy$ ), as also indicated in Fig. 5. The details of these circuits are described next.

#### A. Autoadaptive Comparator

The frequency-decomposed time signals from the BM are processed locally. We employ a binary charge pump [25] to eliminate signal offsets and cancel  $1/f$  noise from the comparator reference. The comparator, shown in Fig. 6, detects zero crossing and provides a control signal for circuitry computing the zero-crossing interval. The charge pump ( $M^+$ ,  $M^-$ ) controls the offset voltage stored on the capacitor through feedback. A change in offset in the BM signal will lead to a change in charge pump duty cycle and effectively charge or discharge the capacitor to follow the offset. The complemen-

tary mirror structure on the left controls the adaptation speed and provides robust bias voltages for transistors  $M^+$  and  $M^-$ .

#### B. S/H and Feature Computation

The approach followed for time-interval computation is similar to that of Kumar [26]. The circuit, shown in Fig. 7, performs a time-interval-to-voltage conversion at every zero-crossing event. Capacitor  $C1$  is charged with a constant current ( $IC$ ) and reset at the end of every period. Just prior to resetting  $C1$ , the follower on the very right is powered up and transmits the voltage on  $C1$  to  $C2$ .

The S/H and reset procedure described above requires two short, subsequent, nonoverlapping voltage pulses to be generated by the two-stage circuitry on the left. A stage consists of a NOR gate and an inverter. A falling signal edge from the comparator causes the NOR output to become high for a very short time, dependent on how fast the inverter toggles its state. This slew rate can be controlled externally by  $V_{bias1}$ . The first stage generates the sample pulse, the second stage resets capacitor  $C1$ .

This technique of S/H is different from the conventional scheme, where the follower is always powered up and a charge-compensated switch is used to S/H. The follower in this scheme is active only for the duration of the S/H pulse. Hence, a proper value for  $V_{Bias1}$ , that results in pulse durations of nanoseconds range, considerably decreases power consumption of the follower. Also, due to faster transition times, the short-circuit current in the digital sections is smaller, further reducing the power consumption by more than an order of magnitude. In this respect, the circuit is an improvement over the circuit described in [26]. To address the problem of charge injection, the follower is turned off slowly. The value of the capacitor  $C3$  and externally controlled voltage  $V_{Bias3}$  set the rate at which the follower is turned off. The fabricated circuit has been tested and found functional for signal frequencies up to 8 MHz.

Apart from the comparator, the computed offset, and S/H and  $Reset$  pulses, are also used for obtaining the signal energy by integrating the full-wave rectification of the threshold-adjusted BM output signal using a capacitor. We employ two transconductors to perform voltage rectification [27]. The energy feature is sampled and held the same way as the frequency feature.

The circuits just described are contained in the block indicated by  $T_{ZC}$  and  $Energy$  in Fig. 4 and interconnected as shown in Fig. 5.

#### C. Arbitration and Asynchronous Data Interface

The feature outputs from every channel are time-division multiplexed to the chip output, using an asynchronous protocol which is most efficient when dealing with communication problems involving a bandwidth-limited bus, and when bus requests are at arbitrary time and rate. The idea is similar to that used by Lazzaro *et al.* [28]. At every zero-crossing instant, the channel requests service by setting a set/reset (SR) latch (Fig. 4). The arbitration logic handles multiple requests at a time and favors the highest frequency channel. It initiates the

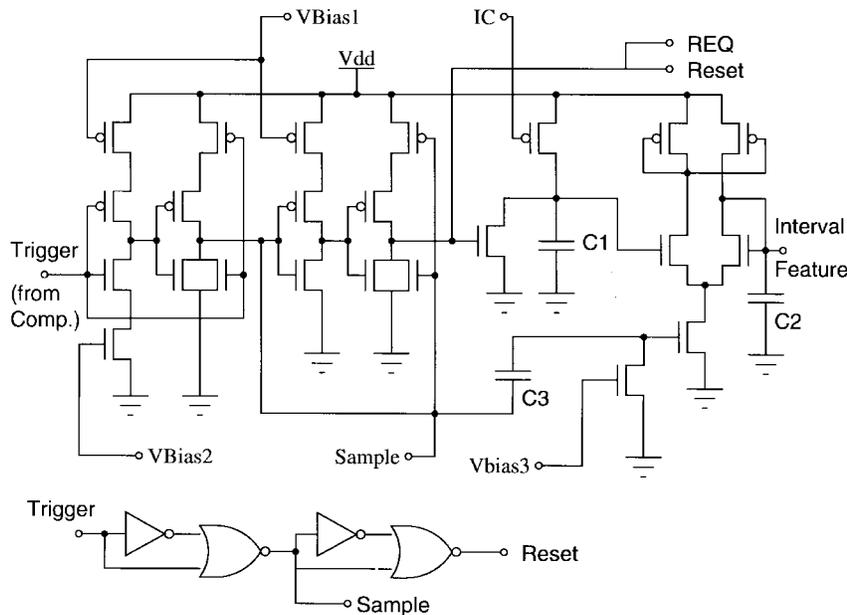


Fig. 7. Sample-and-hold (S/H) circuit and interval feature computation.

address encoder, which passes the winning channel address to a D-flip-flop (D-FF) that stores the channel address currently being serviced. The address is applied to the multiplexer, which steers the channel feature to the data output pins. Once the data has been acquired, an external reset pulse is expected. It is multiplexed back to reset the SR latch of the channel just being completed. The arbitration logic and the encoder generate a new address, which is held by the D-FF and applied to the multiplexer for the next acquisition. We verified this data acquisition scheme by performing software simulations for a speech signal. We found that a data acquisition bandwidth of 30 K samples per second is sufficient to collect all crossing events [29].

Apart from these features, the BM output can also be monitored externally through the multiplexing circuit and is used for tuning the basilar membrane filterbank model.

#### IV. CHIP TEST RESULTS

We have fabricated and tested a 15-channel (2 mm × 2 mm in 1.2- $\mu$ m BiCMOS technology) prototype chip, shown in Fig. 8. We report here some of the experimental results from the chip. Fig. 9 demonstrates the time-interval feature computation. The lower trace is the basilar membrane output signal of the highest frequency channel in response to a triangular FM-modulated input signal in the audio range. The bandpass properties of the basilar membrane channel are evident from the magnitude envelope of the output. As the input frequency decreases, the output amplitude first increases, and then decreases, in correspondence to the bandpass properties of the basilar membrane channel. The upper trace shows the resulting time-interval feature voltage. As the frequency decreases, the time interval between the zero crossings increases, and so does the output voltage. Also note that, since  $T_{ZC}$  is output every period, this feature is output less frequently at low frequencies, as is evident from the larger steps in time.

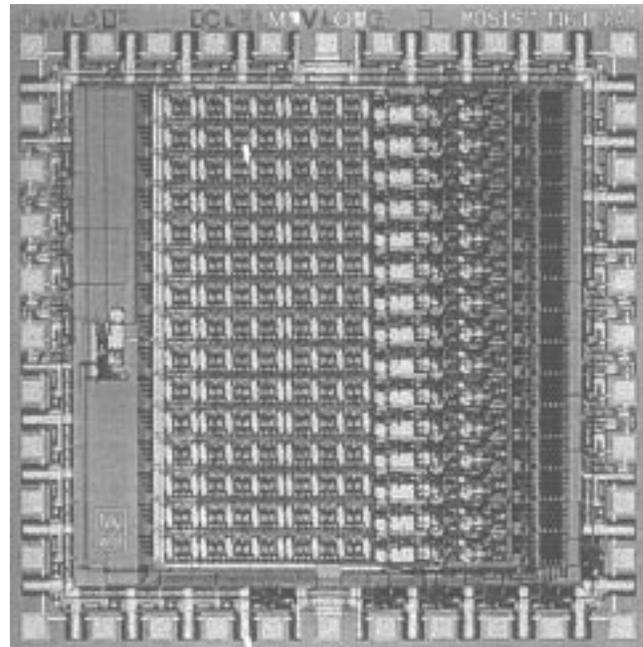


Fig. 8. Micrograph of the 2 mm × 2 mm feature extraction chip in 1.2- $\mu$ m BiCMOS technology fabricated through the MOSIS service.

The energy feature is also extracted every period of the signal. If the period is held constant, then the amplitude modulation of the signal will reflect in this feature. Fig. 10 illustrates this operation. The basilar membrane is supplied with an AM-modulated sinusoidal of constant frequency (uppermost trace). The trace below is the corresponding basilar membrane output. The lower two traces are the energy feature and the time-interval feature, respectively. We observe that signal energy changes, but, due to constant frequency, the time-interval feature remains constant.

Fig. 11 depicts the address-bus activity for four periods of a sinusoidal input signal. To obtain this trace, the reset

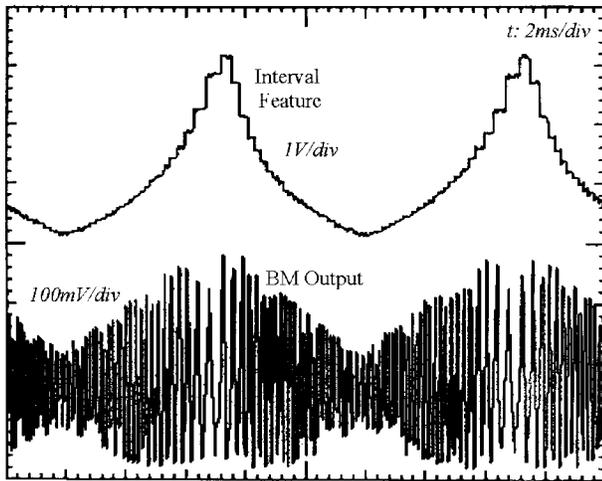


Fig. 9. Interval feature for FM-modulated input.

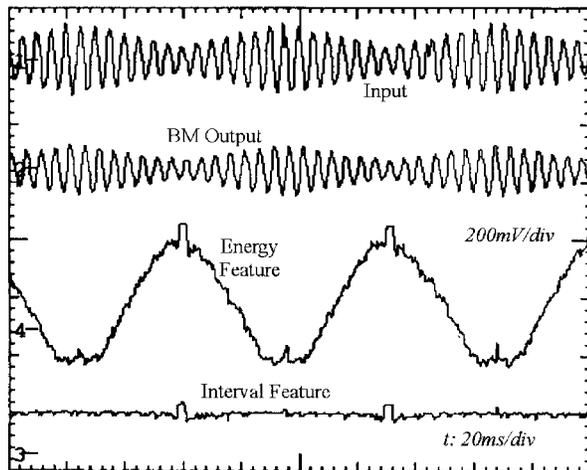


Fig. 10. Energy and interval feature for AM-modulated input.

pulse is externally applied at a clock rate of 50 kHz. Traces 1–4 correspond to interface address lines A0 (LSB)–A3 for 15 cochlea channels. Address zero encodes that none of the channels have requests to be serviced. Consistently, we observe that crossing events tend to be clustered around zero phase of the input signal, the zero crossing.

## V. SPEECH RECOGNITION ARCHITECTURE AND RESULTS

The analog VLSI chip outlined in the previous section emulates the *known* aspects of auditory signal processing. However, beyond the physiological level of neural firing patterns, the mechanisms in higher cortical processing stages are not well understood. Practical system implementations for speech recognition require a compact signal representation that is described by a small number of parameters and contains only (and all) the information relevant to speech recognition. We represent the signal properties by constructing an interval histogram (IH). We generate an IH by creating several bins corresponding to different ranges of values of  $T_{ZC}$ . For any zero-crossing event, we choose the bin that corresponds to the value of  $T_{ZC}$  for that event and fill it with nonlinearly compressed energy for the corresponding event. The IH is

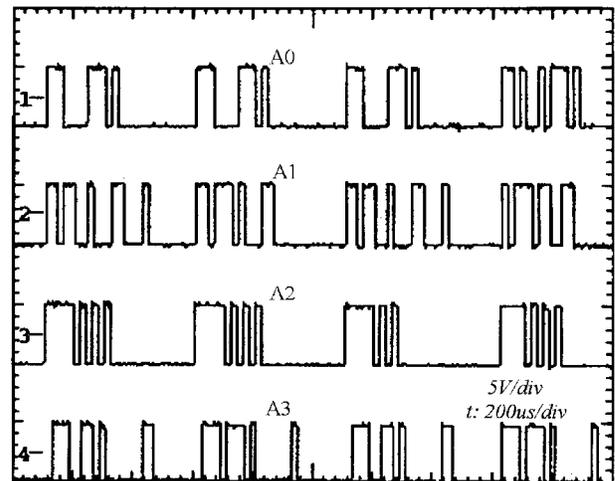


Fig. 11. Address-bus activity during four periods of sinusoidal input.

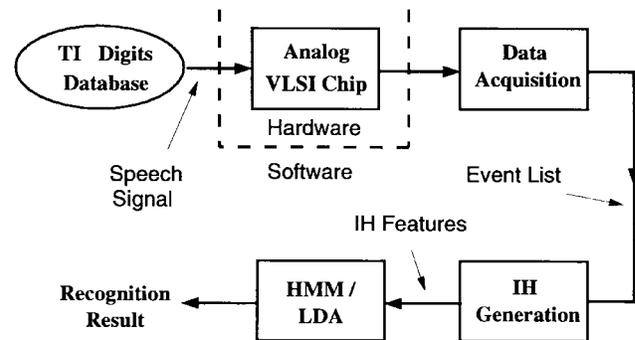


Fig. 12. System architecture for the use of silicon cochlea as a preprocessor to a speech recognition system.

computed from the last 20 crossing events in every channel and does not exceed a time span of at most 40-ms length and is produced at a rate of 100 Hz. As an alternative, we may also use pitch-triggered IH generation, that is, zero-crossing events in the lowest frequency channel trigger an IH to be produced.

When auditory signal representations are interfaced to recognition systems, the so-called *representation-recognizer gap* [11] becomes apparent. There is a significant difference between the conventional linear predictive coding (LPC) or Cepstrum features [14] and the auditory features. The conventional representations are generally uncorrelated and of low dimension, as opposed to auditory features that are highly correlated and of higher dimensions. To resolve this discrepancy, we use linear discriminant analysis (LDA) that also reduces feature dimensionality, thus enabling a more robust estimation of fewer parameters [30].

Fig. 12 depicts the recognition system architecture. The analog VLSI chip serves as the front end. The acquisition system collects zero-crossing intervals and the corresponding energy measure from all channels. Subsequently, a software module computes interval histograms, which are passed to the recognizer. We apply LDA to reduce the feature dimension and then use hidden Markov models (HMM) to perform digit recognition.

In our software simulations, we replaced the analog VLSI chip and the acquisition system by an equivalent software

module. We performed digit recognition experiments on the isolated digits part of the TI-DIGITS database. We modeled each digit by a seven-state single-mixture left-to-right HMM. We obtained a recognition accuracy of 99.47% on the TI-DIGITS database, which is obtained when a feature window size corresponding to the last 20 crossings is used. These results should be treated as preliminary when compared to the state-of-the-art systems [31]. We believe that the performance can be further improved by using better models and optimizing the IH-generation method. However, the recognizer performance certainly demonstrates the applicability of analog VLSI cochlea to auditory-based research. Other researchers have used features similar to the one reported here and found them robust in presence of noise degradation [32], and we expect to see similar robustness from the low-power real-time VLSI system.

Implementing a complete recognizer on a chip or chipset necessitates the implementation of a dimensionality reduction step (matrix vector multiplications) and a statistical decoder on a chip. The sophistication required by algorithms in state-of-the-art speech recognition decoders makes this not a trivial task and, certainly, a challenge.

The question of why the zero-crossing representation is more robust to noise is an interesting one [4]–[6], [18]. We have compared the standard fast Fourier transform and the IH histogram as applied to an input sine wave corrupted by white noise [29]. This comparison shows that the nonlinearity of the IH histogram leads to noise suppression around the signal frequency peak in the spectrum. The nonlinearity originates in the correlation of zero-crossing intervals across channels.

## VI. CONCLUSION

We have demonstrated an approach to real-time auditory-based signal analysis using analog VLSI as a front-end feature extractor. A chip that computes zero-crossing intervals and signal energies in frequency subbands was designed and successfully tested. Furthermore, we reported on recognition results on the TI-DIGITS spoken-digit database obtained from software simulation of the chip-based feature-extraction algorithm. In view of the fact that the recognition model is very simple, and the feature extraction process was not explicitly optimized with respect to various parameters, we consider these results very encouraging.

## REFERENCES

- [1] J. B. Allen, "Cochlear mechanics—A physical model of transduction," *J. Acoust. Soc. Amer.*, vol. 68, no. 6, pp. 1660–1670, Dec. 1980.
- [2] 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, Apr. 1979.
- [3] K. L. Payton, "Vowel processing by a model of the auditory periphery: A comparison to eighth-nerve responses," *J. Acoust. Soc. Amer.*, vol. 83, no. 1, pp. 145–162, 1988.
- [4] K. Wang and S. Shamma, "Zero-crossings and noise suppression in auditory wavelet transformations," Dep. Elect. Eng., University of Maryland, College Park, Tech. Rep., Aug. 1992.
- [5] C. V. Neti, "Neuromorphic speech processing for noisy environments," in *Proc. IEEE Int. Conf. Neural Networks*, 1994, vol. 7, pp. 4425–4429.
- [6] C. R. Jankowski Jr., "A comparison of auditory models for automatic speech recognition," M.S. thesis, Elect. Eng. Comput. Sci. Dep., MIT, Cambridge, May 1992.

- [7] R. F. Lyon and C. Mead, "An analog electronic co-chlea," *IEEE Trans. Acoust., Speech, Signal Processing.*, vol. 36, pp. 1119–1134, July 1988.
- [8] N. Bhadkamkar, "A variable resolution, nonlinear silicon cochlea," Elect. Eng. Dep., Stanford Univ., Stanford, CA, Tech. Rep. CSL-TR-93-558, Jan. 1993.
- [9] 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.
- [10] W. Liu, "An analog cochlear model: Signal representation and VLSI realization," Ph.D. dissertation, Elect. Comput. Eng. Dep., Johns Hopkins Univ., Baltimore, MD, 1992.
- [11] 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.
- [12] P. M. Furth and A. G. Andreou, "Cochlear models implemented with linearized transconductors," in *Proc. IEEE Int. Symp. Circuits and Systems*, 1996, vol. 3, pp. 491–494.
- [13] ———, "A design framework for low power analog filter banks," *IEEE Trans. Circuits Syst. I*, vol. 42, pp. 966–971, Nov. 1995.
- [14] L. R. Rabiner and B. H. Juang, "An introduction to hidden Markov models," *IEEE ASSP Mag.*, pp. 4–16, Jan. 1986.
- [15] E. D. Young and M. S. Sachs, "Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers," *J. Acoust. Soc. Amer.*, vol. 66, pp. 1381–1403, 1979.
- [16] H. E. Secker-Walker and C. L. Searle, "Time-domain analysis of auditory-nerve-fiber firing rates," *J. Acoust. Soc. Amer.*, vol. 88, no. 3, pp. 1427–1436, 1990.
- [17] K. L. Payton, "Vowel processing by a model of the auditory periphery," Ph.D. dissertation, Elect. Comput. Eng. Dep., Johns Hopkins Univ., Baltimore, MD, 1986.
- [18] K. Wang, S. Shamma, and W. J. Byrne, "Noise robustness in the auditory representation of speech signals," in *Proc. ICASSP*, 1993, vol. 2, pp. 335–338.
- [19] X. Yang, K. Wang, and S. Shamma, "Auditory representations of acoustic signals," *IEEE Trans. Inform. Theory*, vol. 38, pp. 824–839, Mar. 1992.
- [20] B. Kedem, "Spectral analysis and discrimination by zero-crossings," *Proc. IEEE*, vol. 74, pp. 1477–1493, Nov. 1986.
- [21] S. Mallat, "Zero-crossings of a wavelet transform," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1019–1033, July 1991.
- [22] N. Kumar, "Investigation of silicon auditory models and generalization of linear discriminant analysis for improved speech recognition," Ph.D. dissertation, Elect. Comput. Eng. Dep., Johns Hopkins Univ., Baltimore, MD, Feb. 1997.
- [23] P. M. Furth and A. G. Andreou, "Linearized transconductors in sub-threshold CMOS," *Electron. Lett.*, vol. 31, no. 7, pp. 545–547, Mar. 1995.
- [24] A. G. Andreou, "Low power analog VLSI systems for sensory information processing," in *IEEE ISCAS-95 Tutorial Book on Multimedia Communications*, 1995.
- [25] G. Cauwenberghs and A. Yariv, "Fault-tolerant dynamic multi-level storage in analog VLSI," *IEEE Trans. Circuits Syst. II*, vol. 41, pp. 827–829, 1994.
- [26] N. Kumar, G. Cauwenberghs, and A. G. Andreou, "A circuit model of hair-cell transduction for asynchronous analog auditory feature extraction," in *Proc. IEEE Int. Symp. Circuits and Systems*, 1996, vol. 3, pp. 301–304.
- [27] C. A. Mead, *Analog VLSI and Neural Systems*. Reading, MA: Addison-Wesley, 1989.
- [28] J. P. Lazzaro, J. Wawrzynek, and A. Kramer, "System technologies for silicon auditory models," *IEEE Micro*, vol. 14, pp. 7–15, June 1994.
- [29] W. Himmelbauer, "Investigation of a zero-crossing-based auditory model for digit recognition and its implementation in analog VLSI," M.S. thesis, Elect. Comput. Eng. Dep., Johns Hopkins Univ., Baltimore, MD, Oct. 1997.
- [30] N. Kumar, C. Neti, and A. G. Andreou, "Application of discriminant analysis to speech recognition with auditory features," in *Proc. 15th Annu. Speech Research Symp.*, Johns Hopkins Univ., Baltimore, MD, June 1995, pp. 153–160.
- [31] R. Haeb-Umbach, D. Geller, and H. Ney, "Improvements in connected digit recognition using linear discriminant analysis and mixture densities," in *Proc. ICASSP*, 1993, vol. 2, pp. 239–242.
- [32] D.-S. Kim, J.-H. Jeong, J.-W. Kim, and S.-Y. Lee, "Feature extraction based on zero-crossings with peak amplitudes for robust speech recognition in noisy environments," in *Proc. ICASSP*, 1996, vol. 1, pp. 61–64.



**Nagendra Kumar** (S'90–M'96) received the Bachelor of Technology degree in electrical engineering from the Indian Institute of Technology, Kanpur, India, in 1989 and the M.S. and Ph.D. degrees in electrical and computer engineering from Johns Hopkins University, Baltimore, MD, in 1991 and 1997, respectively.

He was a Research Engineer with the Indian Institute of Technology from 1991 to 1992. Since 1997, he has been with Telogy Networks Inc., Germantown, MD, as a Senior Member of the

Technical Staff. His research interests include signal processing, speech recognition and compression, machine learning, and VLSI for neuromorphic signal processing.



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

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

low-power coding and instrumentation.

Dr. Cauwenberghs received a Career Award from the National Science Foundation in 1997.



**Wolfgang Himmelbauer** received the M.S.E. degree in electrical and computer engineering from Johns Hopkins University, Baltimore, MD, in 1997.

From 1995 to 1997, he was a Research Assistant in the Sensory Communication Laboratory, Johns Hopkins University, where he was involved in analog circuit design for modeling auditory periphery. He is currently a Design Engineer with Micro Linear Corporation, San Jose, CA. His research interests include analog circuit design and speech signal processing.



**Andreas G. Andreou** (S'80–M'81) received the M.S.E. and Ph.D. degrees in electrical engineering and computer science from 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 University, where he became an Assistant Professor in 1989, an Associate Professor in 1993, and a Professor in 1997. During the academic year 1995–1996, he was a Visiting Associate Professor of Computation and Neural Systems at California

Institute of Technology, Pasadena. He is the co-founder of the Johns Hopkins University Center for Language and Speech Processing. His research interests are in the areas of sensory communication, integrated circuits, and neural computation.