Fully integrated 500uW Speech Detection Wake-Up Circuit

Tobi Delbruck[1], *Senior Member, IEEE*, Thomas Koch[1],
Raphael Berner[1], *Student Member, IEEE*, Hynek Hermansky[2], *Fellow, IEEE*
[1]*Inst. of Neuroinformatics, University of Zurich and ETH Zürich, Switzerland,*
[2]*Center for Language and Speech Processing, Johns Hopkins University, Baltimore*

ABSTRACT

Speech analysis requires substantial computation. It is desirable to run this analysis only when needed and at other times to go to a low power state. Here we propose a self-biased low power speech detection wake up circuit which interfaces directly to standard electret microphones. The speech detector includes a microphone preamplifier, a power extraction squaring circuit, a bandpass filter passing power of the modulation spectrum in the speech band from 2-12 Hz, a half rectifier which extracts this phoneme band power, and a PFM silicon neuron which emits spikes indicating phoneme-rate modulation of the audio spectrum. The output of the speech detector circuit is an asynchronous stream of digital spikes at a rate of 1Hz to 20Hz whose temporal structure indicates the presence of speech. A subsequent conventional processor will go to sleep between spikes and only wake up for full power speech analysis when the temporal structure indicates speech. The circuit is built in 1.6um 2P-2M CMOS and consumes 500uW with a 3V supply when attached to a standard electret microphone.

1. INTRODUCTION

Some recent work in ultra low power sensor interface design has focused on sensor interface circuits which preprocess raw signals to extract relevant information [1-6]. In particular, [6] showed a sophisticated circuit capable of linear predictive coding of the input sound waveform and comparison of the code with stored templates; the prototype chip itself consumed only 450 uW but required the use of external timing FPGA and microphone preamplifier. References [1-4] describe the design and characterization of chips designed for uW-power digital impulse cross correlation, for the purpose of source localization in wireless sensor networks; these chip also did not include sensor interfaces. Julian et al. [5] reviewed the field in 2004.

Here we report on the design and characterization of a speech-detection wake-up circuit which was designed before and during [7]. Our chip implements a simple speech detector (**SD**) based on the amount of modulation spectrum power in the phoneme modulation band of 2-12 Hz, which has been shown to be useful in speech-non-speech classification [8]. In the intended scenarios, the device will be used in handheld, body worn, or at personal interaction distances under battery or scavenged power, where the absolute amount of modulation power is a useful metric for SD and power consumption is important.

This SD chip will serve as a wake-up detector for a more conventional programmable back-end microcontroller. The SD chip outputs digital spike events, between which the microcontroller will sleep, running only its real time clock (RTC). On each event, the microcontroller will wake up, measure the spike time from the RTC, perform some simple computations based on the current spike time and statistics computed from previous spikes, and whether it is worth fully waking for more power-hungry speech processing tasks. In this way the system will spend most of its time in a sleep state.

Fig. 1 shows the proposed overall architecture and the chain of analog signal processing. In this paper we report only on the implementation of the analog processing chain.
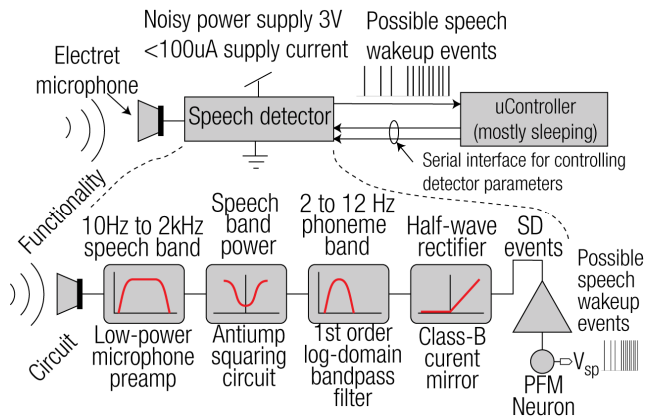

Fig. 1 System concept.

2. IMPLEMENTATION

The implementation of each part of the Fig. 1 SD analog chain is based on modifications of existing circuits. In the circuit schematic figures, transistor W/L geometry are shown in λ=0.8um values, e.g. 10/10 is equivalent to 8um/8um.

The microphone preamplifier (Fig. 2) is based on Baker et al. [9]. This circuit connects to an external JFET electret microphone. It "learns" the microphone DC current level $I_{mic,dc}$ on C_{ad} and linearly transduces the small-signal changes in current $I_{mic,AC}$ through the external feedback resistor R_f to produce the audio signal voltage V_{aud}. M_2 is sized to supply the DC microphone current. The 2-stage Miller-compensated opamp has estimated gain of 10^4.

The audio power in V_{aud} is extracted by the full wave squaring rectifier based on Delbruck's anti-bump circuit (Fig. 3) [10]. One input of the anti-bump circuit is the lowpass average audio signal V_{dc} and the other is V_{aud}. The output current I_{sq} is the sum of a constant current and a current I_{sq} that is proportional (for small signals) to the square of $\Delta V = V_{aud} - V_{dc}$:

$$I_{sq} = \frac{I_b}{1+16\operatorname{sech}^2\left(\kappa\Delta V/2U_T\right)} \quad (1)$$

Unit anti-bump circuit transistors were chosen to form an optimal squaring function for small signal input. Bulks were shorted to sources to eliminate the short and narrow channel effects; the DC output current is removed by the following bandpass filter. The antibump circuit transistors were drawn interdigitated to minimize variation.
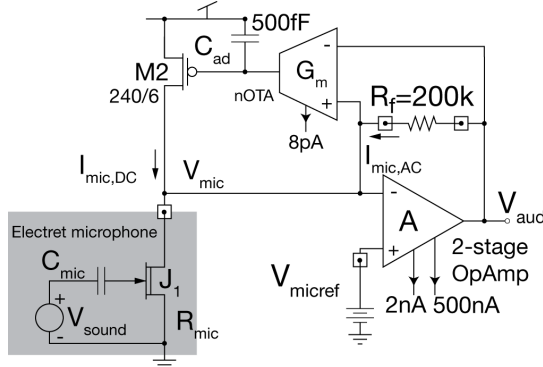

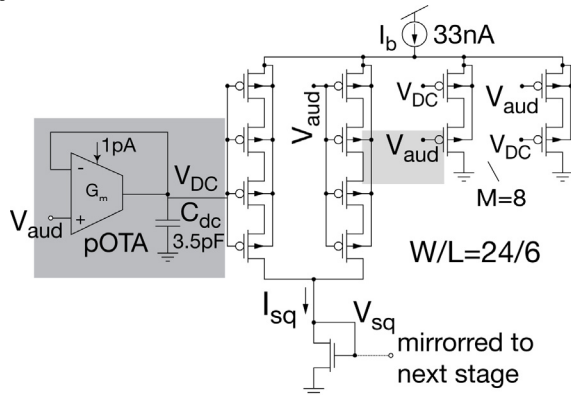Fig. 2 Microphone preamplifier circuit.

Fig. 3 Audio power extraction squaring circuit.

I_{sq} is input to Frey's log-domain bandpass circuit (Fig. 4), which is based on [11]. To form a bandpass filter, a lowpass filter is followed by a highpass filter. The first filter is set to cutoff at 12Hz and the second at 2Hz, forming the combined transfer function

$$\frac{I_{pp}}{I_{sq}} = \frac{A_{bp}\tau_{hp}s}{\left(\tau_{lp}s+1\right)\left(\tau_{hp}s+1\right)}$$

where the gain A_{bp} and time constants τ are determined by bias currents as shown in Fig. 4. Only V_{lp} is accessible externally. This log domain topology has the advantage that it does not require matching n and p type bias currents. Dummy transistors (not shown) were used to minimize mismatch which typically plagues these log domain circuits.

The bidirectional bandpass output current I_{bp} is half-rectified before driving the PFM neuron circuit. The active half-rectifier (Fig. 5) is based on [12]. The feedback holds V_{in} at $V_{g,ref}$ and thus actively mirrors the current, greatly increasing bandwidth near zero current compared with a simple current mirror. It also holds the input at a chosen virtual ground which improves the bandpass filter systematic drain conductance output offset current.

I_{rect} drives a low-power PFM neuron circuit [13] (Fig. 6) which has a small sink leak current set by V_{leak} that sets a threshold for generating events. The final output is the asynchronous digital V_s.
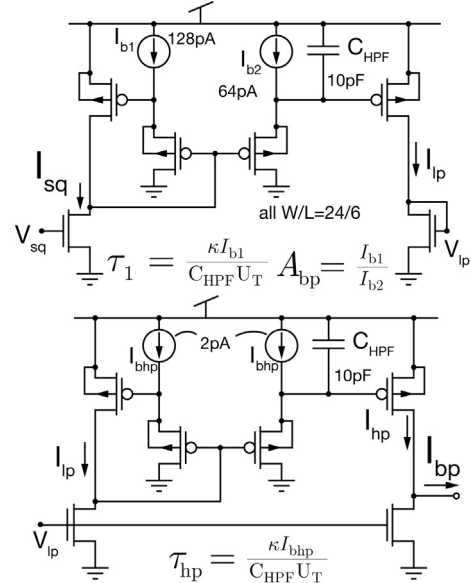

Fig. 4 Phoneme band bandpass circuit.

The SD chip integrates a fixed bias generator [14] which is configured to generate the 13 current biases which span the range of 0.5uA down to 1pA from a master bias current of 1uA. The values of biases were estimated by extensive SPICE simulation of the chip; however it turned out that 8 of 13 values needed to be overridden by external potentiometers. V_{micref} in Fig. 2 and $V_{g,ref}$ in Fig. 5 also were supplied externally. A future design will include configurable biasing, since simulation almost invariably cannot estimate proper values.

The test chip is built in MOSIS 1.6 um 2-poly 2-metal CMOS (Fig. 7). The speech detector including biasing circuits uses an area of about 1.45mm^2. Care was taken to use mismatch reduction layout techniques such as large transistors and common-centroid layout. Metal2 shielding over the circuit allows simultaneous operation of separate optical test structures on the same die. The standard supply voltage for this process is 5V but the circuits were all designed to run at 3V for simpler battery operation.
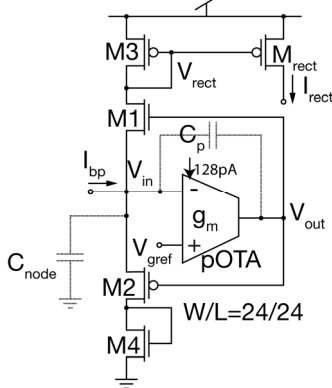

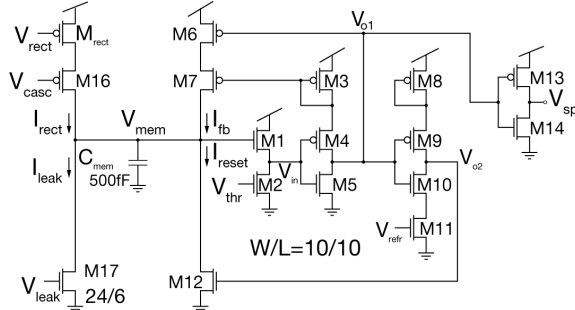Fig. 5 Current rectifier. Capacitors are parasitic.
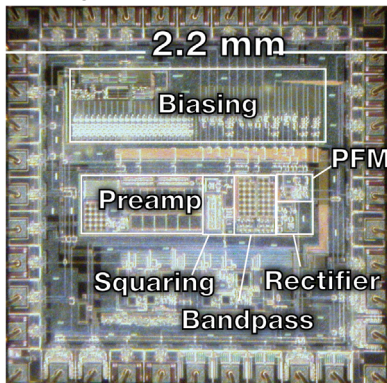

Fig. 6 PFM spiking neuron circuit.

Fig. 7 Die photo.

3. CHARACTERIZATION RESULTS

A standard electret capsule microphone with integrated JFET and transducer gain of 7.9mV/Pa (Monacor MCE-401) was attached to V_{mic} (Fig. 2). Circuits along the processing chain were measured separately by test outputs which drove unity gain analog voltage buffers; internal currents were inferred from the subthreshold gate voltages, assuming back gate coefficient of $\kappa=0.8$. Intermediate nodes were also characterized by disabling the microphone preamp and driving V_{micref} externally.

Fig. 8 shows the response of the squaring circuit to an AM-modulated sound. The output signal current (inferred from the measured diode-connected output voltage) increases like the square of the input amplitude. Fig. 9 shows the bandpass filter operation. The input is a square-wave-modulated sinusoidal AM signal modulated by 30% and the outputs are from the squaring circuit, the rectifier, and the spiking neuron. The neuron spikes only after rising edges of the input. In this measurement the PFM neuron leak was increased to reduce the background spike rate. The currents were inferred from the measured gate voltages. Fig. 10 shows the measured modulation frequency transfer function. The input was a 30%-modulation AM modulated signal with 800Hz carrier frequency and the output is the events per second; the gain A_{bp} of the bandpass was increased to increase the spike rate, to speed up the measurement.
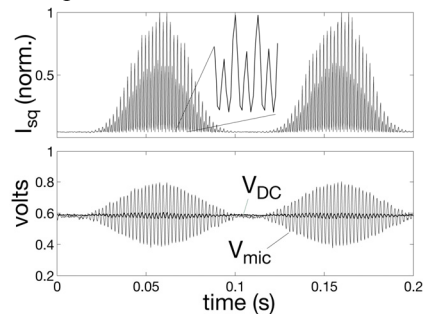

Fig. 8 Measured squaring circuit response for varying input amplitude.
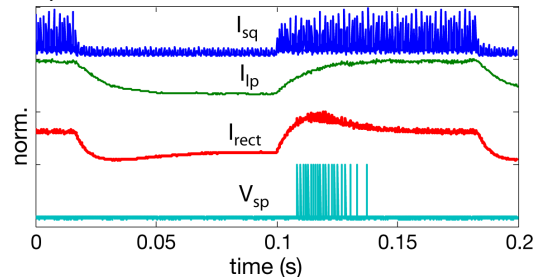

Fig. 9 Measured bandpass step response.

The SD chip as a whole was characterized by measuring PFM output rate as a function of recorded speech input (Fig. 11). The sound sample consisted of speech recorded on a noisy street and was played on a loudspeaker at speech level for person-to-person communication on a noisy street at the microphone input (70dB LAF SPL, measured with Bruel & Kjaer 2250). The sound starts with a truck driving up the street. The first words are "If I could detect speech in this scene that would be pretty good." The segments of speech were identified manually and are marked as Speech. The spike output (V_{sp}) clearly

peaks during segments of speech. The output spike event rate is about 1.2Hz during nonspeech and is about 20Hz during speech. Fig. 11 also shows intermediate signals which are inverted as needed to show increases in current upwards. V_{sq} is the log of the audio power current from the squaring circuit; V_{lp} is the log of the bandpass internal lowpass current node. V_{rect} is the log of the neuron input current which is the half-rectified bandpass filter output.
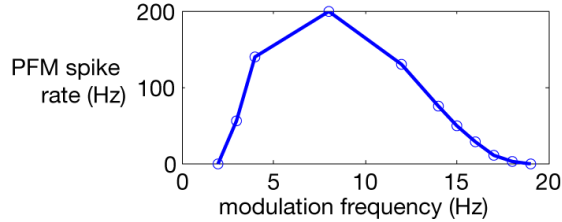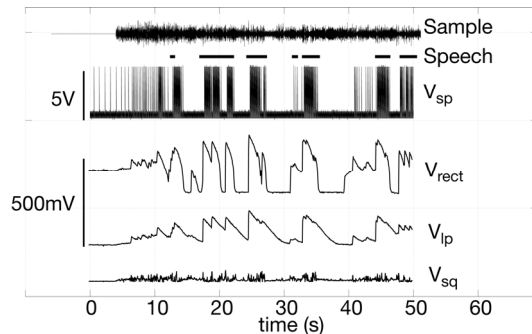

Fig. 10 Measured modulation transfer function.

Fig. 11 Measured spike raster response to input speech-non-speech sound sequence at average 75dB SPL. The manually marked speech segments are shown.

4. DISCUSSION

Fig. 12 is a table of specifications. The background PFM output rate of about 1Hz could leave a microcontroller almost completely asleep: Assuming a wakeup time of 5us and a processing time of another 20us per event, the active duty cycle would be 25us/s=25e-6. If the controller burned 5uA sleeping and 20mA active, the average controller current consumption would be only 5.5uA. The SD circuit at 155uA would dwarf the controller consumption. Still, a pair of 2Ah AA batteries would power the system for more than a year.

Although this functional silicon is encouraging, innovations are required to reach sub-100uW power levels. As in much prior work, the microphone preamplifier still consumes the majority of the total power. Also, the simulation-based estimation of bias currents which are then hardwired is too inflexible for prototype silicon and calls for some degree of programmability. In future work, the SD wakeup chip will be integrated with a microcontroller to explore the feasibility of a complete speech detector system.

6. REFERENCES

[1] D. Goldberg, *et al.*, "VLSI implementation of an energy-aware wake-up detector for an acoustic surveillance sensor network," *ACM Transactions on Sensor Networks (TOSN),* vol. 2, pp. 594-611, 2006.

[2] P. Julian, *et al.*, "Experimental results for cascadable micropower time delay estimator," *Electronics Letters,* vol. 42, pp. 1218-1219, 2006.

[3] G. Cauwenberghs, *et al.*, "A miniature low-power intelligent sensor node for persistent acoustic surveillance [5796-41]," in *Proc. SPIE. Vol. SPIE-5796*, 2005, pp. 294-305.

[4] P. Julian, *et al.*, "Field test results for low power bearing estimator sensor nodes," in *ISCAS 2005*, 2005, pp. 4205-4208.

[5] P. Julian, *et al.*, "A comparative study of sound localization algorithms for energy aware sensor network nodes," *IEEE Transactions on Circuits and Systems I-Regular Papers,* vol. 51, pp. 640-648, Apr 2004.

[6] S. Chakrabarty and A. Gore, "Sigma-Delta Analog to LPC Feature Converters for Portable Recognition Interfaces," in *ISCAS 2009*, Taipei, 2009.

[7] T. Koch, "Design of micropower microphone and speech detector circuits," Masters, D-ITET, ETH, Zurich, 2008.

[8] M. C. Büchler, "Algorithms for Sound Classification in Hearing Instruments," PhD, ETH Zurich, Zürich, 2002.

[9] M. W. Baker and R. Sarpeshkar, "A low-power high-PSRR current-mode microphone preamplifier," *IEEE Journal of Solid-State Circuits,* vol. 38, pp. 1671-1678, Oct 2003.

[10] T. Delbruck, ""Bump" circuits for computing similarity and dissimilarity of analog voltages," in *Proceedings of the International Joint Conference on Neural Networks*, Seattle WA, 1991, pp. 475-479.

[11] B. Linares-Barranco, *et al.*, "Current mode techniques for sub-pico-ampere circuit design," *Analog Integrated Circuits and Signal Processing,* vol. 38, pp. 103-119, 2004.

[12] S. M. Zhak, *et al.*, "A low-power wide dynamic range envelope detector," *IEEE Journal of Solid-State Circuits,* vol. 38, pp. 1750-1753, Oct 2003.

[13] G. Indiveri, *et al.*, "A VLSI array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity," *IEEE Transactions on Neural Networks,* vol. 17, pp. 211-221, 2006.

[14] T. Delbruck and A. van Schaik, "Bias current generators with wide dynamic range," *Analog Integrated Circuits and Signal Processing,* vol. 43, pp. 247-268, Jun 2005.

Process	1.6um 2M 2P CMOS
Supply voltage	3V
Die size	2.2x2.2mm^2
SD circuit size, incl. biasing	1.6mm^2
Supply current	155uA (SD circuit, 48uA, Microphone 107uA)
Power consumption	465uW
Microphone preamp output with 65dB SPL sound input and Rf=220kOhm	100mVRMS
Zero modulation PFM rate	1.2Hz
Peak PFM rate during speech	23Hz

Fig. 12 Specifications.