A Wake-Up Detector for an Acoustic Surveillance Sensor Network: Algorithm and VLSI Implementation

David H. Goldberg[*]
Dept. of Electrical &
Computer Engineering
Johns Hopkins University
Baltimore, MD 21218 USA

Andreas G. Andreou
Dept. of Electrical &
Computer Engineering
Johns Hopkins University
Baltimore, MD 21218 USA

Pedro Julián
Departamento de Ingeniería
Eléctrica y Computadoras
Universidad Nacional del Sur
Bahía Blanca, CP 8000,
Argentina

Philippe O. Pouliquen
Dept. of Electrical &
Computer Engineering
Johns Hopkins University
Baltimore, MD 21218 USA

Laurence Riddle
Signal Systems Corporation
Severna Park, MD 21146 USA

Rich Rosasco
Signal Systems Corporation
Severna Park, MD 21146 USA

ABSTRACT

We describe a low-power VLSI wake-up detector for use in an acoustic surveillance sensor network. The detection criterion is based on the degree of low-frequency periodicity in the acoustic signal. To this end, we have developed a periodicity estimation algorithm that maps particularly well to a low-power VLSI implementation. The time-domain algorithm is based on the "bumpiness" of the autocorrelation of one-bit version of the signal. We discuss the relationship of this algorithm to the maximum-likelihood estimator for periodicity. We then describe a full-custom CMOS ASIC that implements this algorithm. This ASIC is fully functional and its core consumes 835 nano-Watts. The ASIC was integrated into an acoustic enclosure and tested outdoors on synthesized sounds. This unit was also deployed in a three-node sensor network and tested on ground-based vehicles.

Categories and Subject Descriptors

B.7.1 [**Integrated Circuits**]: Hardware Types and Design Styles—*Algorithms implemented in hardware*; C.3 [**Special-Purpose and Application-Based Systems**]: Signal processing systems

General Terms

Algorithms, Performance, Design

[*]Present address: Dept. of Physiology & Biophysics, Weill Medical College of Cornell University, New York, NY 10021.

Keywords

Wake-up detection, power management, acoustic surveillance, sensor networks, maximum likelihood estimation, periodicity, VLSI implementation

1. INTRODUCTION

Sensor network nodes are subject to strict power budgets, as dictated by the need to prolong battery life. This requirement necessitates power-conscious design from high-level algorithms down to the circuit implementation. One way to reduce power consumption is to employ a power management strategy. In the case of a surveillance application, the system may only be required to operate at full functionality in the presence of a novel object. For the remainder of the time, the system can persist in a "sleep" state, where the only required functionality is to detect the presence of the novel object. This power management scheme requires a "wake-up" front-end—a subsystem that detects the novel object and arouses the surveillance system to full functionality. It is crucial that the wake-up subsystem consume very little power relative to the system as a whole if the power management strategy is to be effective.

In this paper, we describe a wake-up detector for an acoustic surveillance sensor network. This sensor network detects and localizes ground-based vehicles such as jeeps and tanks. The wake-up criterion is based on the presence of low-frequency periodicity in the acoustic signal, a feature that is characteristic of sounds generated by vehicle engines. In Section 2, we summarize a maximum-likelihood approach to periodicity estimation and detection. In Section 3, we describe an algorithm for periodicity estimation that maps to a low-power VLSI implementation and a wake-up criterion based on this algorithm. We relate this algorithm to the maximum likelihood approach. In Section 4, we highlight the main features of the ASIC that implements the wake-up detection algorithm. In Section 5, we show experimental results from field tests with synthesized sounds and actual ground-based vehicles. In Section 6, we summarize the paper and discuss the factors that limit the performance of the system.

2. MAXIMUM LIKELIHOOD APPROACH

In this section, we derive the maximum likelihood estimator (MLE) for periodicity. The MLE can be combined with a threshold to make a periodicity detector. In the next section, we will present an alternative periodicity measure that maps to a low-power VLSI implementation. In general, the performance of an estimation algorithm can be gauged by comparison with the MLE, which can be thought of as a "gold standard".

The problem of periodicity estimation is closely related to that of pitch estimation. Pitch estimation asks the question, "What is the period of the signal", whereas periodicity estimation asks the question, "What is the degree of periodicity in the signal?" Much work has been done on pitch estimation, particularly in the context of speech recognition applications [4]. A maximum likelihood approach has been developed by several authors [3, 7, 1]. We will briefly summarize that approach, and then we will relate it to maximum likelihood periodicity estimation.

For the problem of pitch estimation, it is assumed that the sound consists of a distinct repeating pattern and the goal is to extract the period of this pattern. We start with a received signal $x[k]$ that is K samples long. This signal can be decomposed into a periodic signal $s[k]$ plus noise $n[k]$:

$$x[k] = s[k] + n[k] \tag{1}$$

The signal $s[k]$ can be considered a repeating version of a subsegment $q[k]$, which is P samples long and repeated $L = K/P$ times[1] such that

$$s[k] = q[k \bmod P] \tag{2}$$

where $q[k]$ and P are unknown and must be estimated by maximizing the likelihood of the received signal. For a given value of P, $q_P[k]$ is given by

$$q_P[k] = \frac{1}{L} \sum_{l=0}^{L} x[k + lP] \tag{3}$$

The energy of $q_P[k]$ is given by

$$\mathcal{E}_Q(P) = L \sum_{k=0}^{P-1} (q_P[k])^2 \tag{4}$$

The MLE for the period $\hat{P}$ is given by the value of P that maximizes $\mathcal{E}_Q(P)$.

Once we obtain the MLE for the period, we only need go one step further to find the MLE for periodicity. Our estimate of the underlying repeating sequence is given by

$$\hat{q}[k] = q_{\hat{P}}[k] \tag{5}$$

We can use Equation 2 to obtain

$$\hat{s}[k] = \hat{q}[k \bmod \hat{P}] \tag{6}$$

The degree of periodicity, in a maximum likelihood sense, is given by the estimated signal-to-noise ratio:

$$\mathcal{P}_{\text{MLE}} = \frac{\mathcal{E}_{\hat{S}}}{\mathcal{E}_{\hat{N}}} = \frac{\sum_{k=0}^{K-1} (\hat{s}[k])^2}{\sum_{k=0}^{K-1} (x[k] - \hat{s}[k])^2} \tag{7}$$

Figure 1 shows the analysis of two typical signals with the MLE method (ambient recording in (a) and (b) and vehicle recording in (c) and (d)). Figure 1(a) and (c) show the periodicity (in dB SNR) as a function of time (x-axis; 1 second-blocks, corresponding to

1024 samples per block) and candidate periods (y-axis). During each 1 *second* block, the period which maximizes the SNR is the MLE for the pitch period. The maximum SNR is the MLE for periodicity, which is plotted in Figure 1(c) and (d). We see that the MLE for periodicity is sufficient for distinguishing the two sounds.

To rigorously quantify the discrimination ability of the detector, we constructed a receiver operating characteristic (ROC) curve (Figure 1(e)). This curve was compiled from 16 vehicle recordings and one ambient recording. Each recording was approximately 200 seconds long. For this detector, the area under the ROC curve is 0.977. The minimum probability of error $P[\text{miss}] + P[\text{false alarm}]$ is 11.9% and the corresponding threshold is -7.71 dB.[2]

3. ALTERNATIVE ALGORITHM

The MLE for periodicity is computationally intensive, as it requires the estimation of the period of the underlying signal P, which in turn requires the computation of $q_P[k]$ for each value of P. We propose an alternative periodicity measure (PM) that does not require an estimation of P. It is based on the autocorrelation function (ACF), given by

$$R_{xx}[n] = \frac{1}{K} \sum_{j=0}^{K-1} x[k] x[k+n] \tag{8}$$

We will also refer to the normalized ACF:

$$\bar{R}_{xx}[n] = R_{xx}[n] / R_{xx}[0] \tag{9}$$

Figure 2 depicts the normalized ACF for 1 second segments of ambient and vehicle recordings. The ACF of the noise-like ambient recording has low energy at higher lag values, whereas the ACF of the periodic vehicle recording has high energy at higher lag values. We can formulate and alternative periodicity measure (PM) that quantifies this. We define a range of lags of interest, $[N_{\min}, N_{\max}]$, and we sum the squares of the ACF over this range:

$$\mathcal{P}_{\text{alt}} = \sum_{n=N_{\min}}^{N_{\max}} (\bar{R}_{xx}[n])^2 \tag{10}$$

Our proposed periodicity measure can be related to the MLE for periodicity in the following way. Wise et al. [7] shows that power of the MLE estimate of the underlying periodic signal $\hat{s}[k]$, which is maximized to find $\hat{P}$, is equivalent to the following expression:

$$\mathcal{E}_{\hat{s}}(P) = \frac{\hat{P}}{K} \left(R'_{xx}[0] + 2 \sum_{l=1}^{N-1} R'_{xx}[l\hat{P}] \right) \tag{11}$$

where $R'_{xx}[n]$ is alternative version of the ACF:

$$R'_{xx}[n] = \sum_{k=0}^{K-1-n} x[k] x[k+n] \tag{12}$$

In words, the power of the estimate of the signal is related to the value of the ACF at its peaks, which occur every $\hat{P}$ samples. For the class of signals that we are interested in, an ACF that has large values at the (non-zero) peaks also tends to have large energy. Therefore, our algorithm is sufficient for estimating the periodicity.

For practical reasons, our PM requires a modification. In reality, the signal is not stationary, enabling low-frequency fluctuations to introduce additional peaks into the ACF, as shown in Figure 2(a). If the peaks corresponding to these fluctuations are far above the

---

[1]For simplicity, we will only describe the case where $x[k]$ is an integer number number of periods ($K = LP$).

[2]It should be kept in mind that minimizing the probability of error will not necessarily correspond to optimal performance. The true optimal operating point will be determined by the relative costs of misses and false alarms.
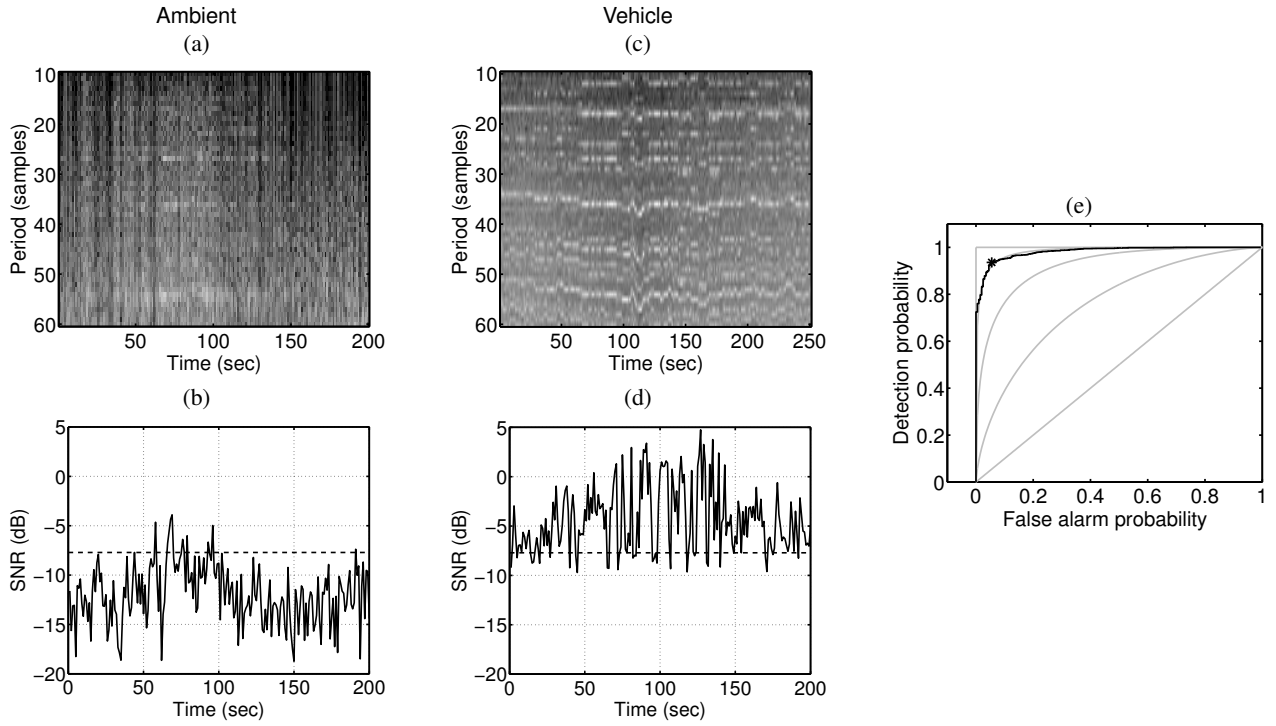
Figure 1: Detection results for the measure based on the maximum likelihood estimator for periodicity. The periodicity measure was computed for 1 second blocks ($K = 1024$ samples). Periods covering the range of $[10, 60]$ samples were examined. (a) Periodicity estimates over all periods for an ambient recording. White corresponds to 5 dB and black corresponds to -40 dB. (b) MLE for periodicity (the periodicity estimate at $\hat{P}$) for an ambient recording. The dotted line corresponds to the threshold that gives the minimum probability of error. (c) Periodicity estimates over all periods for a vehicle recording. (d) MLE for periodicity for a vehicle recording. (e) ROC curve for MLE periodicity measure generated from 16 vehicle recordings and one ambient recording. Lines of constant d' are shown, corresponding to $d' = [0\ 1\ 2\ 3\ \infty]$. The point on the ROC curve which minimizes the probability of error is given by the $*$.

range of lags, they will appear as an offset in the ACF, causing the measure described above to indicate periodicity where none is present. We can eliminate this offset by first computing an approximate time derivative of the ACF, and then computing the sum of the squares. The expression for the PM becomes

$$\mathscr{P}_{\text{alt}} = \sum_{n=N_{\text{min}}}^{N_{\text{max}}-1} (\bar{R}_{xx}[n+1] - \bar{R}_{xx}[n])^2 \tag{13}$$

This can be thought of as a measure of the "bumpiness" of the ACF. Because the derivative is a high-pass operation, this can be thought of as emphasizing higher-frequency components of the signal, and de-emphasizing lower-frequency components of the signal.

Figure 3 shows the analysis of the two typical signals with the alternative PM. Figure 3(a) and (c) show the ACF as a function of time (x-axis; 1 second-blocks) and lags (y-axis). Figure 3(b) and (d) show the PM as a function of time. This measure is clearly also sufficient for distinguishing the two sounds. The ROC curve for this measure is shown in Figure 3(e)). In this case, the area under the ROC curve is 0.970. The minimum probability of error is 10.1%, corresponding to a threshold of 8.34×10^{-3}. This performance is comparable to the MLE detector. It should be noted that the ROC curve in this case is somewhat skewed, enabling very high detection probabilities with essentially no false alarms.

Algorithm simplifications for implementation

The algorithm we have just described can be simplified a great deal and still give satisfactory results for the detection task. These simplifications have been made with a hardware implementation in mind, whether it be in an embedded processor or in a full-custom ASIC.

The signals used in the previous section were quantized with 16-bit precision. It turns out that one-bit precision is sufficient to give acceptable results for the detection task. This is because an infinitely clipped version of the signal retains the periodic structure of the original signal. In fact, the ACF of an infinitely clipped signal $\tilde{x}[n]$ is related to the ACF of the original signal $x[n]$ by [6]

$$R_{\tilde{x}\tilde{x}}[n] = \frac{2}{\pi} \sin^{-1}(R_{xx}[n]) \tag{14}$$

Provided that we encode the signal with zeros and ones (as opposed to -1s and $+1$s), the use of one-bit input signals greatly simplifies the hardware implementation. For example, the multiply operation in the correlation computation is reduced to an XNOR. The one-bit signal has the additional advantage that the normalization step in the ACF computation (the division by $R_{xx}[0]$) is eliminated, as the amplitude information in the signal has already been discarded. One caveat is that we must ensure that the signal is centered around the quantizer threshold point. The PM computation can be further simplified by computing the sum of the absolute value of the discrete differences rather than the sum of the squares.
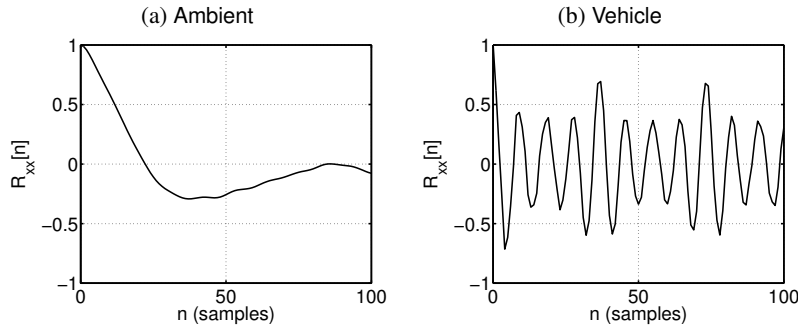
Figure 2: Normalized autocorrelation functions for the two recorded sounds. (a) Ambient recording. (b) Vehicle recording.

The PM equations become

$$\tilde{R}_{\tilde{x}\tilde{x}}[n] = \sum_{k=0}^{K-1} \overline{x[k] \oplus x[k+n]} \qquad (15)$$

$$\mathscr{P}_{\text{simp}} = \sum_{n=N_{\min}}^{N_{\max}} |\tilde{R}_{\tilde{x}\tilde{x}}[n+1] - \tilde{R}_{\tilde{x}\tilde{x}}[n]| \qquad (16)$$

where $\oplus$ represents the XOR operation and the bar represents negation.

Figure 4 shows the results of the simplified PM. The format of the figure is the same as that of Figure 3. In this case, the area under the ROC curve is 0.964. The minimum probability of error is 10.3%, corresponding to threshold of 551. We see that the use of the one-bit signal does not affect the discriminability of the signals in any way.

4. VLSI IMPLEMENTATION

We designed a full-custom CMOS ASIC to implement the simplified wake-up detection algorithm described in the previous section. The chip was fabricated on a 3 mm × 1.5 mm die in a 0.5 μm-process available from the MOSIS service. A micrograph of the chip is shown in Figure 5.

Figure 6 shows a block diagram of the detection circuitry. The 1024 clock cycle detection operation consists of two main phases: an 984 clock cycle ACF phase and a 40 clock cycle PM phase. During both phases, the chip takes as input a one-bit stream of audio data. A 52-sample history of the input is stored in an input register INP. The 40 most recent samples (INP[0 : 39]) in the register are correlated with the oldest sample (INP[51]); this realizes a lag range of [13, 52]. As discussed in the previous section, the correlation consists of an XNOR operation. During the ACF phase, the output of each correlation operation goes to a 10-bit accumulator ACC. After the ACF has been computed for 984 clock cycles, the PM phase begins. A state machine generates select signals such that one even ACF lag and one odd ACF lag are driven onto their respective busses in the proper sequence, enabling the computation of the discrete derivative of the ACF. The results are accumulated in the PM ACC register. At the end of the PM phase, the PM is compared to a user-settable detection threshold, which generates a detection signal. At this point, the ACF registers are reset, and the detection operation begins anew.

The wake-up ASIC was connected to a microphone and signal conditioning circuitry (including a comparator for one-bit A/D conversion) and tested in a laboratory setting. The entire chip consumes 6.3 μW during operation. Because we use separate power supply pins for the I/O pads and the core, we can divide the total power consumption into its constituent parts. The I/O pads con-

sume 5.5 μW and the core consumes 835 nW. The large power consumption of the pads is attributable to two factors: First, the clock input switches at 32 kHz, and is subsequently internally divided down to 1 kHz. This design decision was made in order to facilitate integration with a COTS oscillator. Secondly, approximately 40 non-essential pads were included for debugging purposes. Therefore, it is reasonable to assume that a next generation chip in the same process would consume 1 μW. This power consumption level is far smaller than that of the microphones and signal conditioning circuitry, which draw 300 μW. At these levels, the system will run for well over a year on 3 AA batteries.

5. FIELD TEST RESULTS

The wake-up ASIC was integrated into acoustic surveillance unit (ASU) enclosure along with an array of four Knowles SiSonic MEMS microphones, signal conditioning circuitry, and two bearing estimation ASICs [5, 2]. The ASU enclosure is depicted in Figure 7. We conducted two field tests, one in a public park in Severna Park, Md. with synthesized sounds (field test 1), and another at Aberdeen Proving Ground, Aberdeen, Md. with a selection of ground-based military vehicles (field test 2). In both tests, the detection threshold was set to 1024. This setting corresponds to the point at which the ROC curve in Fig. 4(e) departs from the y-axis and attains the highest value of P[detection] for which P[false alarm] $= 0$.

5.1 Field test 1: Synthesized sounds

In the first field test, synthesized sounds were played from a sub-woofer placed in an open field. In this setting, we had strict control of the frequency content and amplitude of the sounds. In all trials, the ASU was placed 30 feet from the sub-woofer at an angle of 90°. We first performed a series of trials with a signal consisting of three simultaneous time-varying, harmonically related tones (125 Hz, 150 Hz, 175 Hz). This signal is often used as a model for sound generated by a vehicle. The wake-up detector was reliably triggered down to a narrowband SNR of 13 dB (Figure 8). In trials with broadband white noise, the loudest possible volume (50 dB SPL) did not elicit a trigger.

5.2 Field test 2: Ground-based vehicles

In the second field test, vehicles were driven around a 662 m × 108 m oval-shaped track, and three ASUs were placed at various points around the oval. One ASU, at one end of the oval, contained the wake-up detector. When the detector was triggered, the bearing estimation circuit on all of the ASUs localized and tracked the sound source. Table 1 summarizes the wake-up results on an assortment of vehicles in terms of the maximum distance that elicited a sustained detection. During the duration of the test, the false
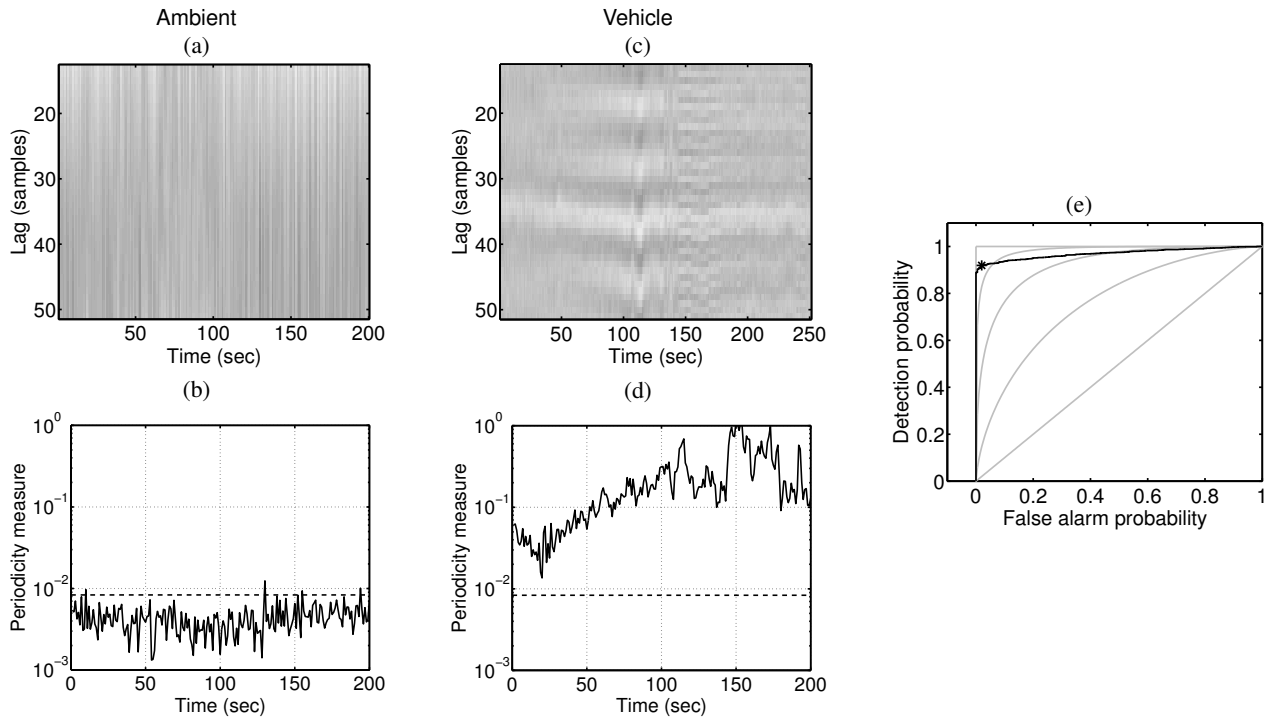
Ambient

(a)

(c)

Vehicle

(b)

(d)

(e)

Figure 3: Detection results for the alternative periodicity measure. $N_{\min} = 13$ **samples and** $N_{\max} = 51$ **samples. (a) Autocorrelation function for an ambient recording. White corresponds to** $+1$ **and black corresponds to** -1**. (b) Periodicity measure for an ambient recording. The detection threshold value was set to the value that minimizes the probability of error. (c) Autocorrelation function for vehicle recording. (d) Periodicity measure for vehicle recording. (e) ROC curve for simplified periodicity measure generated from 16 vehicle recordings and one ambient recording. Same format as Figure. 1(e).**

alarm rate was less than 2 per hour. Also, numerous unscripted targets were detected, such as helicopters, powerboats, and trucks. At one instance, an F/A-18 fighter jet flew overhead at approximately $10,000$ ft and a detection was not elicited.

Vehicle	Description	Distance
M60	Heavy Tracked	> 500 m
HEMET	Heavy Wheeled	~ 250 m
M548	Light Tracked	~ 400 m
HMMWV	Light Wheeled	~ 55 m

Table 1: Performance of the wake-up system on ground-based vehicles.

6. CONCLUSION

We have presented a wake-up detector algorithm based on the degree of periodicity in acoustic signals generated by ground-based vehicles. This algorithm was developed with a low-power VLSI implementation in mind. We designed and tested a CMOS ASIC that implements this algorithm. The core of the ASIC consumes 835 nW, and we expect the next-generation entire chip to consume on the order of 1 μW. The wake-up detector has been field tested and its performance is robust.

The performance of the wake-up system is limited by the signal acquisition and conditioning subsystem rather than the algorithm itself. Because the algorithm requires a one-bit input, it relies on a high-quality comparator at the output of the microphone. In order to provide a useful signal, the comparator threshold must be able

to track the DC level of the acoustic signal. In the case of low-amplitude sounds, the microphone output is on the order of the comparator offset, and the comparator fails to trigger. The design of a low-power, low-offset comparator for integration with a MEMS microphone is a focus of our current research in this area.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] D. H. Friedman. Pseudo-maximum-likelihood speech pitch extraction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-25(3):213–221, June 1977.

[2] P. Julián, A. Andreou, P. Mandolesi, and D. Goldberg. A low-power CMOS integrated circuit for bearing estimation. In *Proceedings of the 2003 International Symposium on Circuits and Systems*, volume 5, pages 305–308, 2003.

[3] A. M. Noll. Pitch determination of human speech by the harmonic product spectrum, the harmonic sum spectrum, and a maximum likelihood estimate. In J. Fox, editor, *Proceedings of the Symposium on Computer Processing in Communications*, volume 19, pages 779–797, Chichester, Sussex, UK, 1969. Wiley-Interscience.

[4] L. R. Rabiner, M. J. Cheng, A. E. Rosenberg, and C. A. McGonegal. A comparative performance study of several pitch detection algorithms. *IEEE Transactions on Acoustics,*
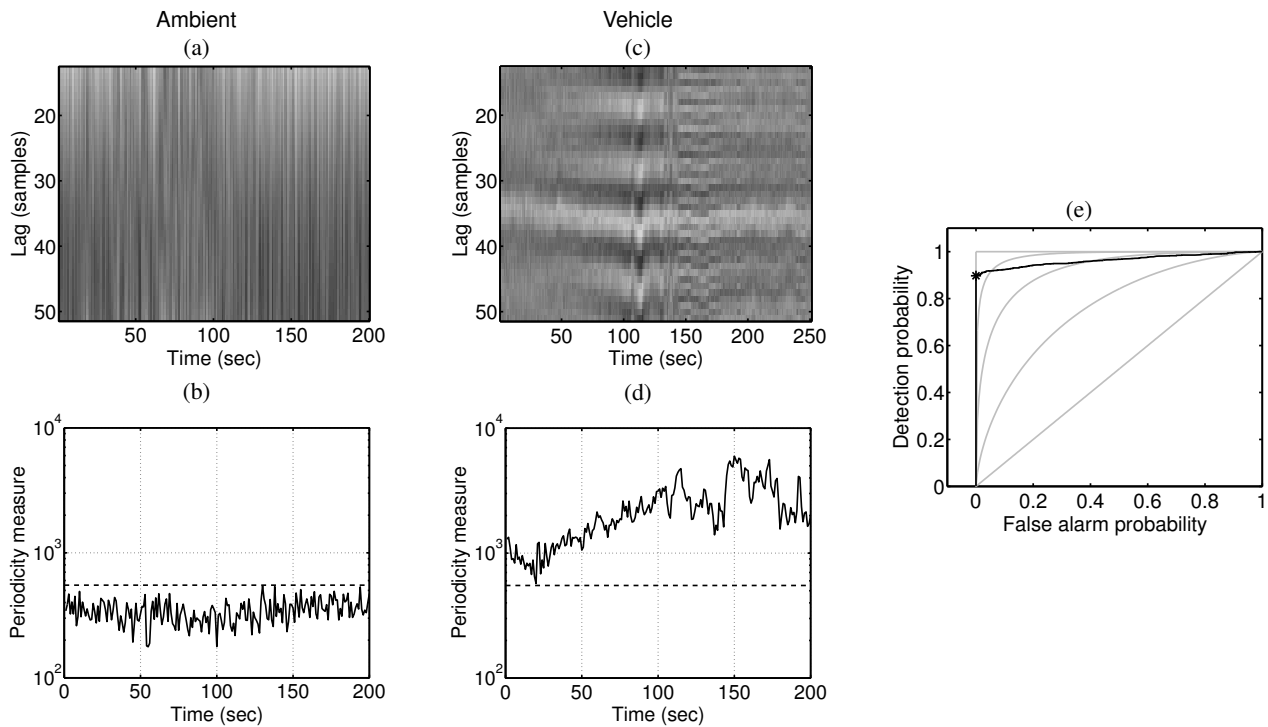
Figure 4: Detection results for the simplified periodicity measure. Same parameters as Figure 3. (a) Autocorrelation function for an ambient recording. White corresponds to 1024 **and black corresponds to** 0. **(b) Periodicity measure for an ambient recording. (c) Autocorrelation for vehicle recording. The detection threshold value was set to** 551. **(d) Periodicity measure for vehicle recording. (e) ROC curve for simplified periodicity measure generated from 16 vehicle recordings and one ambient recording. Same format as Figure. 1(e).**

Speech, and Signal Processing, ASSP-24(5):399–418, October 1976.

[5] M. Stanacevic and G. Cauwenberghs. Mixed-signal gradient flow bearing estimation. In *Proceedings of the 2003 International Symposium on Circuits and Systems*, volume 1, pages 777–780, 2003.

[6] J. H. Van Vleck and D. Middleton. The spectrum of clipped noise. *Proceedings of the IEEE*, 54(1):2–19, January 1966.

[7] J. D. Wise, J. R. Caprio, and T. W. Parks. Maximum likelihood pitch estimation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-24(5):418–423, October 1976.
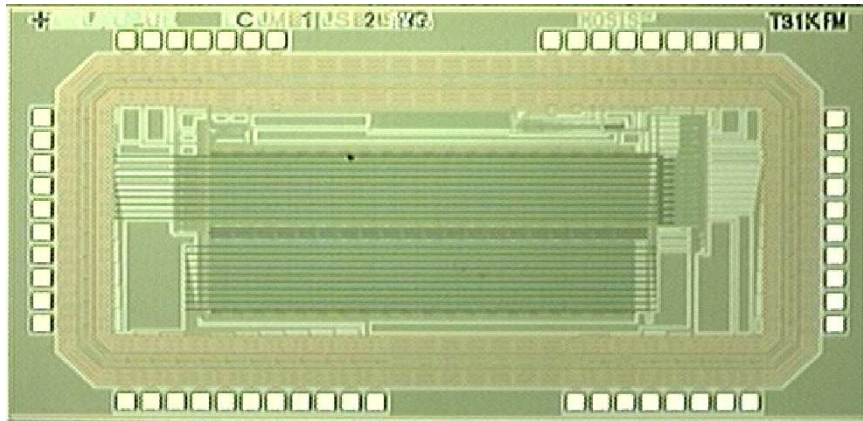
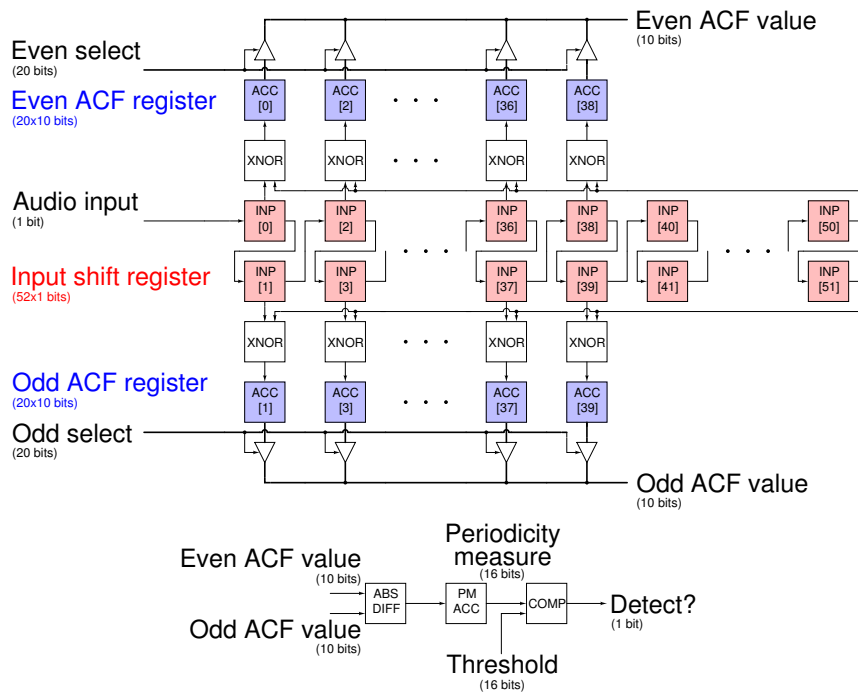**Figure 5: Micrograph of the wake-up detector CMOS ASIC.**

Figure 6: Block diagram of the VLSI implementation of the wake-up detection algorithm. INP **represents the input register.** ACC **represents a 10-bit accumulator and corresponding register. The** ABS DIFF **block computes** $|a - b|$**. The register** PM ACC **stores the periodicity measure, and** COMP **represents a 16-bit digital comparator.**
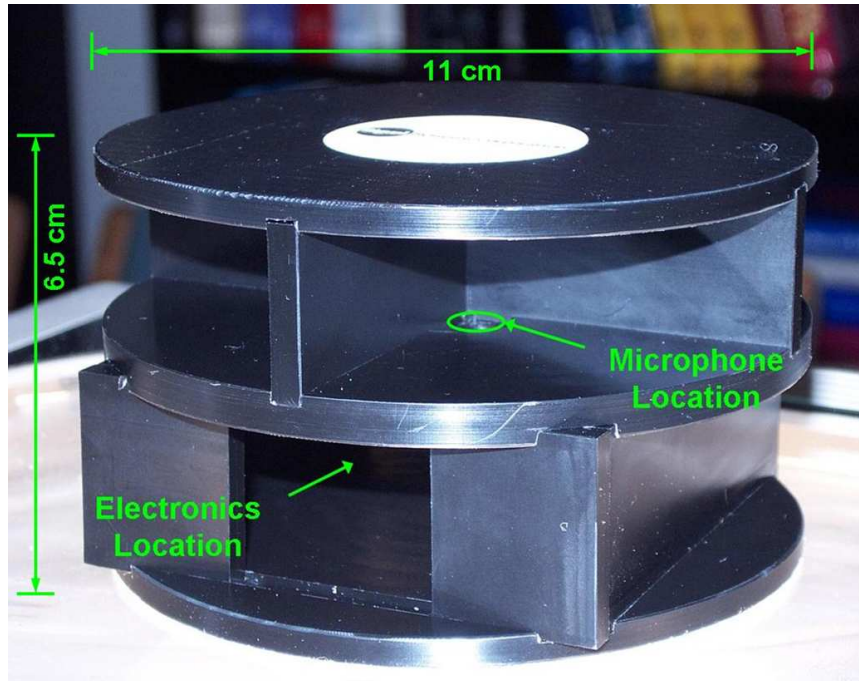
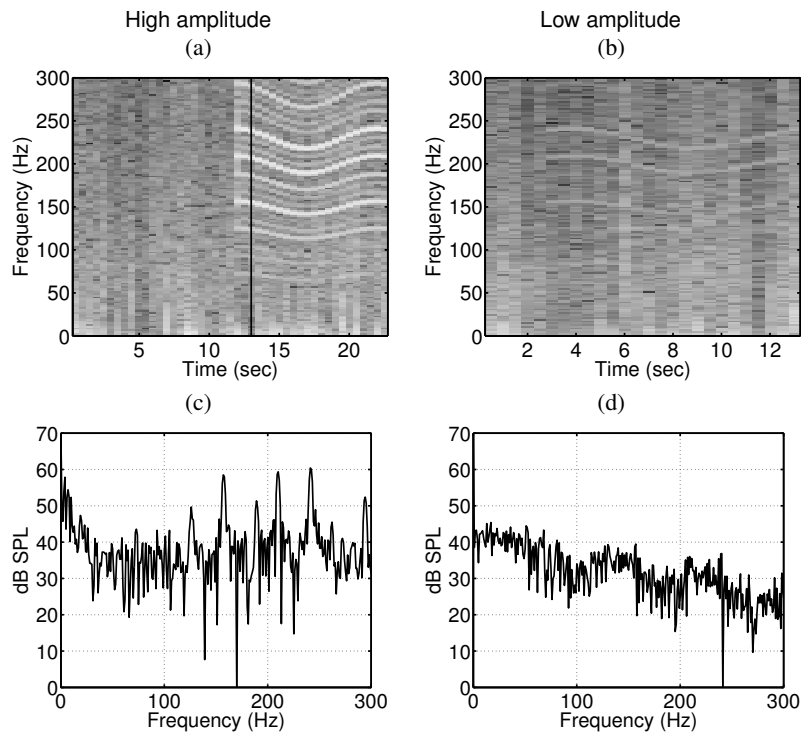**Figure 7: Photograph of acoustic surveillance unit enclosure.**

Figure 8: Field test 1 results for a signal three simultaneous time-varying, harmonically related tones (125 **Hz,** 150 **Hz,** 175 **Hz). (a) Spectrogram representation of a loud sound. The line indicates the moment that the detection signal is triggered. (b) Spectrogram representation of a lower amplitude sound, corresponding to the limits of the system's operation. (c) Power spectral density for loud sound from seconds 11 to 12. (d) Power spectral density for lower amplitude sound from seconds 2 to 3.**