An overview of the use of remote embedded sensors for audio acquisition and processing

Lewis Girod
Computer Science and AI Lab
Massachusetts Institute of Technology
32-G918, 32 Vassar St.
Cambridge, MA, 02139
girod@nms.csail.mit.edu

Marie A. Roch
Dept. of Computer Science
San Diego State University
5500 Campanile Dr
San Diego, California, 92182-7720
marie.roch@sdsu.edu

Abstract

In recent decades, the cost of acoustic technologies has declined dramatically. Advances in networks, storage devices, and power management have made it practical to consider the remote location of sensors. Nonetheless, many challenges remain for the fabrication, deployment, and use of remote sensors.

This paper provides an overview of the issues involved in developing remote acoustic sensors. We discuss physical design and the integration of components, data storage and communication issues, signal acquisition and classification, and the relationship of these issues to power usage requirements.

1 Introduction

The design, implementation, and deployment of remote acoustic sensors is a broad area with many cross discipline aspects. Remote acoustic sensors come in all sizes and shapes, ranging from simple stand alone recording devices to integrated sensor networks where acoustics may be only one of many modalities. In many cases, remote acoustic sensors have stringent energy requirements which can place restrictions on the storage, communication, processing, and signal acquisition properties of the devices.

In locations with limited infrastructure, power management and the ability for the user to access or retrieve the data are paramount. In some situations, the need for localization or improved signal to noise ratio may dictate the use of microphone arrays. Deployment in hostile environments such as arctic or deep sea conditions requires additional considerations.

Remote sensors are capable of generating large acoustic or mixed media data sets. With these large corpora, the need for automated processing becomes critical as the staffing requirements for human analysis are both cost and labor prohibitive. The development of automated analysis can yield valuable data such as seasonal or diel behavioral patterns of animals, surveillance, or perimeter intrusion detection.

While many of the aspects of remote acoustic sensors are interdependent, this overview paper is organized into sections which discuss individual facets of these devices. Section 2 describes choices related to the physical design of the sensor and the integration of components. Sections 3 and 4 discuss power and communication/data management issues. Section 5 describes design choices and issues related to data acquisition, and classification of the data is covered in section 6. Finally, we conclude with a summary of relevant trends and research directions.

2 Packaging and System Integration

Packaging refers to all of the mechanical aspects of an embeddable node, while *system integration* refers to the process of selecting appropriate components and modules and integrating them into a working system. Because these tend to be less technical issues, they are often glossed over in research reports, despite the fact that they comprise some of the most time-consuming elements of building an embedded sensing system. In addition, the decisions made in this process ultimately determine the scope of features that a system can support.

When developing a new design for remote deployment, there are a number of design requirements to take into consideration:

Form Factor. The size and weight of the system, as well as the number of separate components, all make a tremendous difference in the amount of effort required to deploy the system. This is especially important for systems that must be deployed and removed rapidly.

Environmental Conditions. Difficult weather conditions such as moisture, extreme temperatures, wind, dust, immersion in water, or high pressure will require special packaging. The activity of animals can present problems; *e.g.*, animals destroying cabling, antennas, or microphone wind screens. For many remote applications, protection from human tampering or theft is also required.

Mechanical Sensor Configuration. Holding the sensors in a solid mechanical frame will provide higher reliability and enable more repeatable experiments. For sensor arrays, the ideal size and shape may depend on the application.

Lifetime. Remote systems are often costly to visit for maintenance, and in some cases maintenance visits are impossible [31]. Routine maintenance such as replacement of batteries and local storage media can be planned based on maintenance costs. On-site maintenance caused by failures is generally difficult to predict without experience and extensive testing to work out bugs.

Computational Capacity. Local computation can be used to implement real-time responses at the sensor, as well as data reduction to maximize the utility of limited storage and network capacity. Some designs, such as LEAP [20] support a low-power always-on "preprocessor" that can wake a main processor when events of interest occur. The specific algorithms to be used determine the requirements in terms of RAM and processing power. Local processing increases the software complexity of the system as well as energy and heat dissipation requirements.

Storage Capacity. Large-capacity persistent storage can archive raw data, eliminating the need for local data reduction or real-time network transfer. Flash drives have smaller capacity than hard drives, but are more mechanically robust and energy-efficient.

We will see throughout this paper that these choices have a significant impact across many aspects of building a remote embedded sensing system. In practice, many of these choices are dictated by component availability and resources, both in terms of time and funding. The task of assembling a platform is a difficult compromise between cost, capability, and flexibility in the face of unknown future requirements.

3 Energy Management

Energy is one of the most difficult challenges in developing a remote embedded acoustic system. Much of the prior work in embedded sensing has focused on low-rate sensing of environmental parameters, such as temperature and humidity [34] [12]. While low-rate sensors can conserve energy by sleeping much of the time, this problem is much more challenging in acoustic sensing. In addition, the much higher data rates associated with acoustic sampling typically require more processing, storage, and network resources, in general leading to higher energy requirements.

3.1 Factors Contributing to Energy Consumption in Acoustic Sensing

There are several primary contributors to the energy requirements of an acoustic sensor system, including: the analog front end, sampling, digital processing, storage, and network transmission.

Analog Front End and Sampling. As a rule, the circuits required to amplify and sample the acoustic sensors draw a significant amount of power, at best between 10 and 100 mW, and often considerably more.[1] This is partly due to theoretical lower bounds on energy requirements [35] and partly due to inefficiencies imposed by practical implementations [22]. Achieving low power performance often requires custom designs, since off-the-shelf solutions are often not designed for energy saving.

Some applications may only need to be active at certain specified times when the phenomenon of interest occurs, and can "duty cycle" by shutting the system down when it is not needed. But for those applications that require continuous digital monitoring, there is no way around this cost. In some cases, a continuously running low power analog filter can be used to trigger digital sampling and processing. However, if the trigger circuit incurs latency, this type of design will typically lose some part of the triggering event in the interim before the sampling process starts.

Digital Processing. The energy requirements of a main processor board vary widely depending on the architecture, clock rate, and the amount and type of RAM. Processor architecture and feature sets have a wide-ranging impact, including bit width, virtual memory support, floating point support, "MMX" instruction sets, support for special digital signal processing (DSP) features, and voltage and clock scaling. Often specialized DSP chips can yield performance gains, although these solutions tend to be more customized and typically do not support a full-featured operating system, which makes building a sophisticated system more difficult.

The cost of refreshing dynamic RAM (DRAM) is one of the most significant power costs after the CPU itself, and scales with the size of the RAM. This cost dominates in situations where the processor is idle or in "standby mode." Saving the RAM to persistent storage and shutting down the CPU completely eliminates that cost, but incurs a long startup latency. One promising new technology is "Unified Memory," in which a small, fast SRAM is used as a write

[1] Our systems are based on the 4-channel Digigram VXPocket 440, which draws 2 W, and a custom preamplifier that draws 40 mW [11].

cache for a larger flash memory. Such a solution can avoid both the power cost of DRAM and the limited speed and write cycles of flash.

Persistent Storage. Persistent storage technologies include flash, hard drives, and new "hybrid" drives that are just starting to become available from Samsung. Flash has a low power cost and high mechanical robustness, but has many disadvantages; in particular, the slow speed of the erase cycle, the limited number of permitted write-erase cycles, and the smaller capacities relative to hard drives. Hard drives are very costly in power and must be protected against shock, although smaller drives tend to be more shock resistant. The lowest power and most durable hard drives currently available are "microdrives," which typically require about 45 mW standby power and 800 mW when writing.[2] Hybrid drives are a forthcoming technology that couple a hard drive to a large flash write buffer. Internal software writes new data to flash, only spinning up the drive when necessary to clear space in the flash. The flash write buffer eliminates the standby cost and optimizes write costs.

Networking. Wireless networking is another significant energy cost. Since wireless systems must emit energy to communicate, there are theoretical lower bounds on the energy costs [7]. In addition, inter-node coordination, channel estimation and negotiation, and recovery from message loss introduce significant additional overhead. In general, wireless networks impose a significant energy cost when idle, waiting to receive, and higher costs to actually send and receive data. For example, the SMC 2532W 802.11B card draws 0.5 W in "power save" mode and 3 W when transmitting.[3] To avoid this idle cost, the system must completely shut down the radio, which requires additional coordination in the network to determine when to bring the network up. In some instances, a low-power paging channel can be used for coordination, as in the LEAP platform [20] and related work.

3.2 Energy Technologies

For some remote embedded sensing applications, infrastructure-supplied energy may be available. For the rest, the system will need to store energy for its use, and possibly recharge that energy supply. In this section we discuss different alternatives for this.

Batteries. Battery technologies vary in energy density and in ease of use. The highest energy densities are currently achieved by the Lithium Ion (Li+) battery technology, commonly used in laptops. These batteries are very lightweight and have high energy density, but are also expensive and easily damaged. In addition, specialized charging circuits are required to use Li+ batteries without damage, and most providers of off-the-shelf solar power solutions do not support them. However, Li+ batteries are ideal for fielded systems where the batteries are periodically swapped, or for temporary deployments.

Deep-cycle lead-acid batteries are the most commonly used type of battery for solar charging applications. These batteries are heavy but tend to be inexpensive and durable. Charge controllers that integrate with wind generators or solar cells are readily available off-the-shelf.

Solar Panels. Depending on the deployment environment, different energy sources may be available: wind, water, and sunlight are the most common. Solar panels are desirable because they do not require any moving parts. However, solar panels can perform poorly for several reasons. Dirt on the panel from animals or from wind-carried dust can block the sun from the panel. In addition, the entire panel must have access to the sun. Because the panel is composed of many cells wired in series, any cell that is not in sunlight will *consume* rather than generate energy. This means that a panel in shadows or with partial coverage will have greatly reduced output.

4 Communication Environment and Data Reduction

Networking remote embedded sensors varies greatly across different deployments, from cases where there is very limited or no connectivity to cases where direct Internet access is possible. Given a data rate and local data reduction policy, the capabilities of the network, combined with local storage capacity, determine the duration of the experiment. There are several networking technologies to consider:

No Connectivity. Deployments on the sea floor, in caves, and other locations that have no RF connectivity may simply rely on mass storage to collect data, perhaps with an alarm mechanism or very low bandwidth link for control and status monitoring. Such sites may also be served by "data mules" that physically visit and download data.

Satellite Internet. Satellite Internet can provide high speed access to remote locations, although uplink speeds are generally limited to 56 Kbps. However, with limited sample rates and data compression, it may be possible to stream back complete data. Satellite phones are also a possibility, although they are costly and provide only 2.4 Kbps data rates.

Long-range Wireless. Long-range wireless links may be useful for reaching nearby infrastructure, although installation of such links requires both time and monetary in-

[2] Technical specifications from Hitachi 3K6 Microdrive.

[3] From specifications on the SMC website, `http://www.smc.com`.

vestment. Off-the-shelf 802.11 radios can be coupled with amplifiers and directional antennas to provide a low-cost long-range link. Work at CENS on the MASE seismic array has deployed a multi-hop 802.11 network with links as long as 10 Km.[4] Multi-hop wireless networks are difficult to implement; for this deployment a significant investment in custom routing software was required. High speed connections are also feasible. HPWREN[5] deploys a number of embedded sensors (including audio) on a wireless regional network which spans large portions of the county of San Diego, California with a 45 Mbps wireless backbone. When audio is deployed on a large network, quality of service issues become relevant for some applications.

Short-range Wireless. Compared with long-range multi-hop networks, short range networks are relatively easy to implement; many routing solutions exist off-the-shelf, *e.g.*, [1]. Using a short-range network, a single point can act as a local server or a gateway to the Internet for a collection of rapidly deployable nodes. For remote deployments, the server can be a single node with enhanced storage, processing, and user interface capabilities. Short-range networks can also provide time synchronization services. Systems that implement cooperative signal processing often rely on the Global Positioning System (GPS) to provide tight time synchronization, but GPS is not available in all locations, *e.g.*, inside buildings or under dense canopies of foliage. Short-range networks can be used to propagate GPS time references to nodes in GPS-denied areas, enabling them to participate in cooperative signal processing algorithms [6] [17].

In consideration of these different technologies, there are clearly limits to the amount of acoustic data that can be carried. A single channel at 16 bit / 24 KHz amounts to 384 Kbps, already larger than the low-end Digital Subscriber Line (DSL) uplink speed, and significantly more than what can be achieved with satellite hardware. For many applications, higher sample rates and sensor arrays are desirable; these requirements multiply the raw data rates.

Local data reduction can address these issues. Lossless compression using the Free Lossless Audio Codec[6] or WavPack[7] can often achieve lossless reduction of 50% or more. Selecting the minimum sample rate required, or applying a decimation filter in software also cuts down the data rate. In cases where the phenomenon is well-understood, lossy reduction algorithms can be used, such as wavelet compression or event detection and segmentation.

Lossy techniques can be used to reserve the limited network bandwidth only for events of interest, while locally storing all data, or can be used to extend the lifetime of the experiment by locally storing only a subset of the raw data. In the latter case, a tiered data retention policy should be formed in which complete raw data are stored for portions of the experiment, and the remaining space is used to store reduced data covering a longer time period.

5 Signal Acquisition

In Sections 2 and 3 we discussed some of the factors concerning signal acquisition. Solid packaging and secure wiring are critical to getting good results from an acoustic sensing system. In this section we expand on this topic, discussing a number of issues relevant to signal acquisition.

5.1 Minimizing Electronics Noise

Commercial off-the-shelf products are usually well-designed, but it is often the case that custom hardware must be assembled, either to provide specialized functionality or for reasons of cost. When building custom acquisition hardware, there are several key design points to consider.

First, it is best to design or purchase an ADC that supports a *differential* or *balanced* input. A differential input amplifies the voltage difference between two lines (see [14]), independent of a ground reference. This approach has two advantages: 1) it ignores any differences in the ground reference on opposite ends of the signal line, and 2) any pickup from EMI will affect both lines equally and will have no effect on the *difference*.

Second, the signal lines should be *impedance matched* and if run over a long distance, they should be a twisted pair. Impedance matching is critical for noise property (2) above to hold true. This can be accomplished by placing closely matched resistors in series with both signal lines.

Third, it is important to provide sufficient power supply filtering on the supply for the sensor biasing and preamp electronics. Typically these filters are made up of several different sizes of capacitor that bypass different ranges of noise frequency, as in [15]. It may also be best to use a linear regulator rather than a switching power supply, to cut down on supply noise.

5.2 Minimizing Environmental Noise

Noise in the environment presents problems for an acoustic sensing system. In cases where the noise is limited and restricted to a specific band, it can be readily filtered out after sampling. However, if the noise level is high enough to saturate the ADC, then it will not be possible to filter.

Environmental noise can be filtered before sampling in two ways: 1) using an analog electronic filter on the output

[4] See `http://www.cens.ucla.edu/Project-Descriptions/Seismology/index.html`.

[5] See `http://hpwren.ucsd.edu`

[6] See `http://flac.sourceforge.net`.

[7] See `http://www.wavpack.com`.

of the preamp, and 2) using mechanical filters that attenuate noise. Analog filters are typically designed with a fixed rather than configurable transfer function. They add to the complexity of the analog front end design, but are a good idea in cases where the nature of the noise is well-known ahead of time.

Mechanical acoustic filters are more commonly used to address environmental noise problems. Wind filters can be constructed to shield microphones. The operational theory of wind filters is to shunt wind around the microphone using a material that is relatively transparent to sound. The filters are typically composed of a soft, fur-like material suspended on a mesh around the microphone. In marine environments, the mechanical coupling between a surface float and a hydrophone can cause the hydrophone to move up and down in the water column, producing noise due to turbulence. Alternatively, when the hydrophone is floated from an anchor, a current can produce strum on the line. In both cases, this problem is typically addressed by adding one or more acoustic dampers to reduce the coupling effect between the hydrophone and float.

5.3 Signal Enhancement

Once the signal is sampled and recorded digitally, many signal enhancement options present themselves. With sufficient local processing capacity, many of these techniques can be applied directly at the sensor, to help with data reduction. In this section we describe several methods of signal enhancement.

Filtering. The simplest form of enhancement is filtering to select portions of the signal based on frequency content. There are many different types of filter implementation. The most general method of filtering is to use a Discrete Fourier Transform (DFT) to represent the signal in the frequency domain. In the frequency domain representation, individual frequency ranges can be zeroed out, attenuated, or enhanced arbitrarily. Then, after transforming back to the time domain, the resulting signal will be a filtered version of the original [32]. While the DFT is very useful for filtering finite segments of signals, it is more difficult to apply continuously, and is computationally heavier than other filtering algorithms.

The Finite Impulse Response (FIR) filter is a second type of filtering algorithm. The FIR produces each output value based on a linear combination of the previous N input values, called "taps." An FIR filter can be designed using a number of different design tools. FIR filters can represent low-pass, high-pass, or band-pass filters, and can provide a variety of "roll-off" characteristics (*i.e.,* how quickly the filter transitions from the pass to stop band). Sharper roll-off requires more taps, hence more state in the filter and more work to compute the linear combination.

A concern for filtering systems is the amount of time needed to reach steady state. Until samples are available for all of the N taps, the filter output is ill defined. For systems triggered by analog events (see Section 3.1), this adds to the delay in acquiring the signal. The number of taps required to implement a given FIR filter is inversely proportional to the width of the transition between pass and stop bands, and filter designers should take care to avoid making transition bands narrower than needed.

The Exponentially Weighted Moving Average (EWMA) smoothing filter is a third type. An EWMA filter is easy to implement and computationally cheap: at each step, the new output is a weighted sum of the input and the previous output. While EWMA may be a good choice for certain types of application, it suffers from "phase distortion," in which phase delays vary as a function of frequency.

Beamforming. When a sensor array is available, *beamforming* techniques can be used to enhance a signal [36]. In beamforming, signals from multiple sensors are combined to produce a single, enhanced version of the signal arriving from a particular direction.

Beamforming is based on the principle that signals arriving from a particular direction arrive at the sensors at different times, as a function of the geometry of the array, the direction of arrival, and the propagation speed of the signals. If the direction of arrival (DOA) of a signal is known, the components coming from that direction can be combined together to emphasize the signals from that direction, while de-emphasizing other signals.

In the simplest form of this technique, the signals from several microphones are shifted in phase and added together. The phase shift is performed in the frequency domain so that the phase shift need not be an integral number of samples. By adding the shifted signals, signals from the desired direction are emphasized, while other signals will tend be more likely to cancel each other out.

Beamforming is often combined with an algorithm to estimate direction of arrival; see for example algorithms based on maximum likelihood estimates [3]. These algorithms will typically estimate the direction of arrival for energy found in a specific set of frequencies. The resulting DOA estimate can be used as the input to a beamforming algorithm to enhance the signal, and can be combined across nodes to localize the source.

6 Signal Detection and Classification

Signal detection and classification are necessary to provide useful information about large acoustic datasets which cannot be effectively summarized by human staff due to cost and time constraints. System designers must make choices about where the processing will occur, the feature set to use, and the types of classification schemes that will be implemented.

Location of Processing. In most instances, the ideal situation is to perform all processing on the embedded sensor and to report classification results as part of the metadata for the acoustic data. For some applications, only the metadata itself need be transmitted, greatly reducing bandwidth requirements. An example of this type of system is one where tactical observations are reported to a command and control center [9]. In other cases, such as for bioacoustic studies, the acoustic data must be reported in addition to the metadata as it may be archived for further study or verification. For bioacousticians, it is frequently desirable to archive the entire acoustic signal as the data may be used for new studies at later dates[8].

For many embedded sensors, the power, computational, and storage requirements necessary to implement classifiers are prohibitive and the acoustic sensor either does limited or no preprocessing. In these cases, sampled or parameterized acoustic data is typically sent to a general purpose computer or cluster where detailed analysis is performed.

Signal Detection. In most applications, it is desirable to have a low cost algorithm which can identify potential events of interest. This is typically accomplished using either a rule- or classifier-based signal detector. Rule-based classifiers are typically finite state machines with transition rules between signal and noise based upon a measure of the energy in the signal. Li et. al. [19] provide a good example of this type of system. In contrast, classifier-based systems such as [27] use machine learning techniques to discriminate between the signal and the noise. The advantage of the classifier-based systems is that they are usually more easily adapted to new environments and do not require as much tweaking of the parameters to obtain good results. Unfortunately, many classifier-based algorithms have higher computational cost and may not be as well suited for sensors with limited computational capacity. When multiple networked sensors detect the same event, it is possible to have the detectors collaborate to determine bearing and report the acoustic data with the highest SNR.

Feature Extraction. For complex signals such as human speech, a representation of the short-time spectrum is typically desirable. Cepstral coefficients (CC) are derived either from the output of perceptual filterbanks such as the so-called Mel filters or from the signal's linear prediction coefficients (LPC) [24] (perceptual weighting is possible for LPC cepstra [13]). The homomorphic transform used to create the cepstrum has the advantage of transforming signals that have been convolved into ones that are added, making it possible to approximate the vocal tract filter or compensate for convolutional noise. It is common to provide information about the rate of spectral change by appending the first and second differences of the CCs to the feature vector.

While perceptually weighted CC are well accepted for speech technologies, their use is not appropriate in many non-speech applications. For bioacousticians working with species whose hearing is radically different from humans, the use of human centric filters is inappropriate although it is still possible to obtain good results in certain cases [5, 18]. Recently, Clemins et al. [4] have proposed an extension to the work of [13] to permit perceptual filters for species with known hearing characteristics. For non-biological noises, perceptual filters are not necessarily appropriate.

Classification. A wide variety of classifiers have been applied to acoustic data. An issue that is likely to have an influence on classifier choice is whether or not the data has an expected structure over time such as speech or stereotyped animal calls. When the audio has unknown temporal characteristics (e.g. recognizing a speaker without requiring a predetermined phrase), classifiers that do not typically capture time domain structure have been shown to be successful. Examples of such classifiers include Gaussian mixture models (GMMs) [30, 29], classification and regression trees (CARTs) [23], support vector machines (SVMs) [2], and neural networks [8].

Data with structure over time is typically recognized by other methods. When the acoustic pattern is relatively consistent from one production to the next, matched filters [33], correlated spectrograms [21], or dynamic time warping [26] can be used. For more variable acoustic data, hidden Markov models (HMMs) [26, 16] are frequently used classifiers. HMMs provide a set of state dependent distributions and distributions which govern the transitions between states. As each observation is processed, the HMM has the possibility of transitioning to another state. The state sequence is considered to be hidden, and dynamic programming techniques are used to determine the likelihood of either all possible state transition sequences or the best one. Most current systems use GMMs to model the state dependent distributions, but many of the aforementioned time-independent classifiers [16] have been used in place of GMMs with excellent results.

The implementations of classifiers cited throughout this discussion all assume that data has been sent to a centralized processing location and do not typically address the problem of distributed learning where decisions are based upon input from multiple sensors. While less studied, there are examples in the literature of classification systems which fuse multiple sensors. Garg et al. [10] demonstrate a system which uses an ensemble method to improve a Bayesian network which fuses the outputs of visual detectors and a simple audio signal detector. The interested reader is referred

[8] John A. Hildebrand, personal communication 2005.

to [25] for a general introduction to distributed learning on a sensor network.

7 Directions for Research

It is always a difficult task to determine the issues whose resolution will result in the greatest advancement to the field, and any attempt at soothsaying will of necessity be biased by the experiences of the researchers making the prediction. We identify three areas that we believe will permit significant improvements with respect to both the capabilities and deployment of embedded sensors.

Throughout this article, we have made consistent references to sensor limitations which can be traced to power requirements. We believe that advances in both power sources and power management can yield significant improvements for distributed sensor technologies and should thus be a significant focus of research for the future.

Secondly, further work needs to be done on classification in noisy environments. While techniques such as beam forming and means subtraction [16] can help enhance signals, the contributions of additive and convolutional noise still remain significant challenges for recognition. Recent work has addressed the issue of missing or corrupted data, and proposed techniques for either estimating or ignoring portions of the short-time spectrum which are determined to be unlikely to be associated with the signal (see [28] for a summary). These techniques show promise and should continue to be explored.

Finally, the majority of systems described in the literature are experimental in nature and require a significant learning curve for scientists in other disciplines to use. As these systems begin to mature, further consideration needs to be given to user interfaces and how to package complex concepts into tools that can be easily learned and deployed.

Acknowledgements

The authors would like to thank the HPWREN project and the NSF for their support of this work through NSF grants 0426879 and CNS-0520032, and the NSF Cooperative Agreement CCR-0120778. We also thank the reviewers for their thoughtful comments and suggestions.

References

[1] J. Bicket, D. Aguayo, S. Biswas, and R. Morris. Architecture and evaluation of an unplanned 802.11b mesh network. In *MobiCom '05)*, pages 31–42, August 2005.

[2] W. M. Campbell, J. P. Campbell, D. A. Reynolds, E. Singer, and P. A. Torres-Carrasquillo. Support vector machines for speaker and language recognition. *Computer Speech and Language*, 20(2-3):210–229, April-July 2006.

[3] J. Chen, K. Yao, and R. Hudson. Maximum-likelihood source localization and unknown source localization estimation for wideband signals in the near-field. *IEEE Transactions on Signal Processing*, (8):1843–1854, 2002.

[4] P. J. Clemins and M. T. Johnson. Generalized perceptual linear prediction features for animal vocalization analysis. *J. Acous. Soc. of Am.*, 120(1):527–534, 2006.

[5] P. J. Clemins, M. T. Johnson, K. M. Leong, and A. Savage. Automatic classification and speaker identification of African elephant (*Loxodonta africana*) vocalizations. *J. Acous. Soc. of Am.*, 117(2):956–963, February 2005.

[6] J. Elson, L. Girod, and D. Estrin. Fine-grained network time synchronization using reference broadcasts. In *OSDI*, pages 147–163, Boston, MA, December 2002.

[7] D. Estrin, L. Girod, G. Pottie, and M. Srivastava. Instrumenting the world with wireless sensor networks. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2001)*, pages 2033–2036, 2001.

[8] K. R. Farrell, R. J. Mammone, and K. T. Assaleh. Speaker recognition using neural networks and conventional classifiers. *IEEE Trans. Acoust., Speech, Signal Processing*, 2(1):194–205, January 1994.

[9] D. Friedlander, C. Griffin, N. Jacobson, S. Phoha, and R. R. Brooks. Dynamic agent classification and tracking using an ad hoc mobile acoustics sensor network. *EURASIP Journal on Applied Signal Processing*, 4:371–377, March 2003.

[10] A. Garg, V. Pavlović, and J. M. Rehg. Booseted learning in dynamic bayesian networks for multimodal speaker detection. *Proc. IEEE*, 91(9):1355–1369, September 2003.

[11] L. Girod, M. Lukac, V. Trifa, and D. Estrin. The design and implementation of a self-calibrating distributed acoustic sensing platform. In *ACM SenSys*, Boulder, CO, Nov 2006.

[12] R. Guy, B. Greenstein, J. Hicks, R. Kapur, N. Ramanathan, T. Schoellhammer, T. Stathopoulos, K. Weeks, K. Chang, L. Girod, and D. Estrin. Experiences with the extensible sensing system (ESS). Technical Report 61, Center for Embedded Networked Sensing (CENS), March 2006.

[13] H. Hermansky. Perceptual linear predictive (plp) analysis of speech. *Journal of the Acoustical Society of America*, 87(4):1738–1752, April 1990.

[14] P. Horowitz and W. Hill. *The Art of Electronics*, pages 98,184. Cambridge University Press, 1989.

[15] P. Horowitz and W. Hill. *The Art of Electronics*, page 45. Cambridge University Press, 1989.

[16] X. Huang, A. Acero, and H.-W. Hon. *Spoken Language Processing*. Prentice Hall PTR, Upper Saddle River, NJ, 2001.

[17] B. Kusy, P. Dutta, P. Levis, M. Mar, A. Ledeczi, and D. Culler. Elapsed time on arrival: A simple and versatile primitive for canonical time synchronization services, in press, 2006.

[18] J. A. Kéç-Kogan and D. Margoliash. Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study. *J. Acous. Soc. of Am.*, 103(4):2185–2196, April 1998.

[19] Q. Li, J. Zheng, A. Tsai, and Q. Zhou. Robust endpoint detection and energy normalization for real-time speech and speaker recognition. *IEEE Trans. Speech Audio Processing*, 10(3):146–157, March 2002.

[20] D. McIntire, K. Ho, B. Yip, A. Singh, W. Wu, and W. J. Kaiser. The low power energy aware processing (LEAP) embedded networked sensor system. In *IPSN '06: Proceedings of the fifth international conference on Information processing in sensor networks*, pages 449–457, New York, NY, USA, 2006. ACM Press.

[21] D. K. Mellinger and C. W. Clark. Recognizing transient low-frequency whale sounds by spectrogram correlation. *J. Acous. Soc. of Am.*, 107(6):3518–3529, June 2000.

[22] B. Nauta. Analog CMOS low-power design considerations. In *Digital Low Power-Low Voltage Workshop at ESSCIRC 96*, pages 1–16, 1996.

[23] M. Padmanabhan, L. R. Bahl, and D. Nahamoo. Partitioning the feature space of a classifier with linear hyperplanes. *IEEE Trans. Speech Audio Processing*, 7(3):282–288, May 1999.

[24] J. W. Picone. Signal modeling techniques in speech recognition. *Proceedings of the IEEE*, 81(9):1215–1247, September 1993.

[25] J. B. Predd and H. V. Kulkarni, Sanjeev R.and Poor. Distributed learning in wireless sensor networks. *IEEE Signal Processing Mag.*, 23(4):56–69, July 2006.

[26] L. R. Rabiner and B.-H. Juang. *Fundamentals of Speech Recognition*. Signal Processing Series. Prentice-Hall, Englewood Cliffs, NJ 07632, 1993.

[27] B. Raj and R. Singh. Classifier-based non-linear projection for adaptive endpointing of continuous speech. *Computer Speech and Language*, 17(1):5–26, January 2003.

[28] B. Raj and R. M. Stern. Missing-feature approaches in speech recognition. *IEEE Signal Processing Mag.*, 22(5):101–116, September 2005.

[29] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn. Speaker verification using adapted gaussian mixture models. *Digital Signal Processing*, 10(1/2/3):19–41, Jan/Apr/Jul 2000.

[30] D. A. Reynolds and R. C. Rose. Robust text-independent speaker identification using gaussian mixture speaker models. *IEEE Trans. Acoust., Speech, Signal Processing*, 3(1):72–83, January 1995.

[31] P. Schenker, E. Baumgartner, P. Backes, H. Aghazarian, L. Dorsky, J. Norris, T. Huntsberger, Y. Cheng, A. Trebi-Ollennu, M. Garrett, B. Kennedy, A. Ganino, R. Arvidson, and S. Squyres. FIDO: A field integrated design and operations rover for mars surface exploration. In *The 6th International Symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS-'01)*, number AM012, 2001.

[32] W. Siebert. *Circuits, Signals, and Systems*. MIT Press, 1985.

[33] A. Sirovic, J. A. Hildebrand, S. M. Wiggins, M. A. McDonald, S. E. Moore, and D. Thiele. Seasonality of blue and fin whale calls and the influence of sea lee in the Western Antarctic Peninsula. *Deep Sea Research Part II Topical Studies in Oceanography*, 51(17-19):2327–44, 2004.

[34] G. Tolle, J. Polastre, R. Szewczyk, D. Culler, N. Turner, K. Tu, S. Burgess, T. Dawson, P. Buonadonna, D. Gay, and W. Hong. A macroscope in the redwoods. In *Proceedings of the 3rd International Conference on Embedded Networked Sensor Systems (SenSys05)*, pages 51–63. ACM, Nov. 2005.

[35] E. Vittoz. Future of analog in the VLSI environment. In *ISCAS'90 Int. Symp. Circuits Systems*, pages 1372–1375, 1990.

[36] K. Yao, R. Hudson, C. Reed, D. Chen, and F. Lorenzelli. Blind beamforming on a randomly distributed sensor array system. *IEEE JSAC*, 16(8):1555–1567, Oct. 1998.