Anawake: Signal-Based Power Management For Digital Signal Processing Systems

John Lazzaro and John Wawrzynek
CS Division, UC Berkeley, Soda Hall
Berkeley, CA 94720-1776
lazzaro@cs.berkeley.edu, johnw@cs.berkeley.edu

Richard Lippmann
MIT Lincoln Laboratory
Room S4–121, 244 Wood Street
Lexington, MA 02173-0073
rpl@sst.ll.mit.edu

Abstract

Single-chip, low-power, programmable digital signal processing systems are capable of hosting complete speech processing applications, while consuming a few milliwatts of average power. We present a power management architecture that decreases the average power consumption of these systems to 3–10 microwatts, in applications where speech signals are present with a sufficiently low duty cycle. In this architecture, a micropower analog signal processing system, *Anawake,* continuously analyzes the incoming signal, and controls the power consumption of the DSP system in a signal-dependent way. We estimate system power consumption for Anawake designs optimized for different peak-speech-signal to average-background-noise ratios.

1. Introduction

Single-chip programmable digital signal processing (DSP) systems (Figure 1), with integrated RAM, ROM, and A/D and D/A converters, have developed into suitable platforms for demanding applications such as small-vocabulary, speaker-independent speech recognition. An example is a single-chip DSP system developed for cost-sensitive toy products, that supports moderate-quality, isolated-word speaker-independent recognition with a 10-word vocabulary [1]. This device provides 4 MIPS of programmable processing power (in addition to dedicated filtering circuits for speech pre-processing) and consumes 15 mW at 3V. Improvements in low-power digital design [2], combined with process scaling, have led to predictions [3] that single-chip low-cost DSP systems early in the next decade will offer

0.25mW/MIPS power dissipation, with dynamically-settable performance up to 400 MIPS.

These future systems will support 20-MIPS applications while consuming a few milliwatts of power. This power specification supports battery operation on AAA cells, *if* the application requires a relatively low duty cycle of operation (standard AAA Nickel-Cadmium cells have a capacity in the range of 220 mA H). For a speech recognition system, the low duty cycle condition usually is implemented as a mechanical switch that the user presses to initiate a session. In some speech recognition applications, however, this "press to begin" model is a distinct liability: instead, a system that listens continuously for speech input is desired. In some cases, user interface issues dictate a continuously-listening model; for other applications (wordspotting systems for surveillance, speech control for the physically disabled) continuous listening is inherent in the problem definition.

A battery-operated, continuously-listening speech recognition system would require an implementation with an *average* power consumption in the 1-10 microwatt range. Microwatt analog signal processing techniques [4] have been developed for niche applications such as hearing aids. Microwatt analog pattern recognition systems have also been developed for implantable medical devices; one recent example, an intracardiac arrhythmia classifier [5], includes a statistical classifier similar to those used in speech recognition architectures. Recent research has focused on analog micropower implementations of other structures used in speech recognition, include speech feature extraction [6,7] and hidden Markov model state decoding [8]. These processing systems offer *peak* microwatt power consumption, permitting the use of energy sources with low maximum current ratings, such as lithium button batteries.

However, hard-wired analog implementations are best suited to product classes that do not require extensive customization for each application. Portable speech recognition systems, in contrast, are highly application-specific in nature: the speech vocabulary and the product behavior are different in each application. These systems require the flexibility of programmable digital systems.

This paper presents a system architecture for implementing continuously-listening speech recognition systems that consume milliwatts of peak power, but microwatts of average power. The architecture combines a single-chip programmable DSP system with an application-independent micropower analog system. Although we present this architecture in the context of speech recognition, it can be generalized to support to other signal processing and pattern recognition applications.

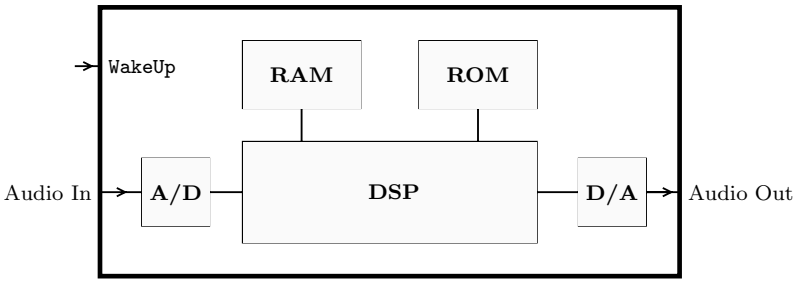

Figure 1. Block diagram for a single-chip programmable DSP system, with integrated ROM, RAM, and speech codec.

2. Anawake

Single-chip DSP systems, as shown in Figure 1, typically have a micropower sleep mode; in this mode, processing and A/D and D/A conversion are halted, but the internal state of the system is preserved. Typically, the processor switches between normal operation and sleep mode under the control of an external input, (labeled `WakeUp` in Figure 1). In systems that include a mechanical switch to start speech input, the `WakeUp` signal is toggled by the switch. The processor can also enter sleep mode under program control.

In the architecture shown in Figure 2, a micropower analog system, *Anawake,* acts as a signal-based power management system for a single-chip DSP system. Anawake has two functional blocks: a speech detector and a 100ms analog delay. The speech detector monitors the audio input and activates the `WakeUp` input of the DSP if a speech signal is present. The analog delay postpones the presentation of audio input to the A/D converter of DSP chip. Without this delay the beginning of the speech utterance would be clipped off, seriously degrading speech recognition accuracy (speech detection requires the examination of about 100ms of a speech sound).

The operation of the system is straightforward. After a suitable period of time has elapsed without speech input, the processor enters sleep mode under program control, switching from milliwatt power consumption to microwatt power consumption. Anawake continues to function, consuming microwatts. The speech detector constantly monitors the input signal for renewed speech activity, and the analog delay preserves the last 100ms of audio input. If the speech detection block senses new speech activity, Anawake toggles the `WakeUp` signal of the DSP, and the DSP reverts to its normal processing mode. Upon wakeup, the DSP processes the audio input of the previous 100ms (held in the analog delay) that includes the start of the speech utterance.

If speech activity is present in the audio input with a duty cycle D $(0 \leq D \leq 1)$, the power consumption of the architecture shown in Figure 2 is

$$P_a + (1 - D)P_s + DP_n, \tag{1}$$

where P_a is the power consumption of Anawake, P_s is the power consumption of the DSP chip in sleep mode, and P_n is the power consumption of the DSP chip while

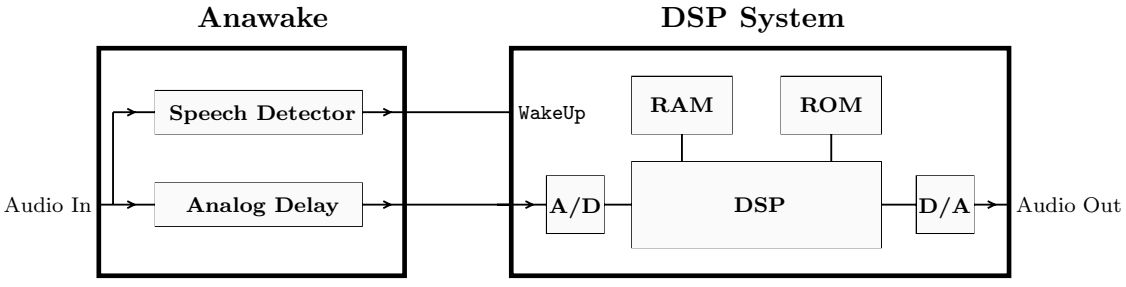

Figure 2. The Anawake architecture: a micropower analog chip performs signal-based power management for a single-chip DSP system.

performing speech recognition. For sufficiently low duty cycles ($D \ll (P_a + Ps)/P_n$), the system power consumption is approximately equal to the sum of the Anawake power consumption and the sleep-mode power consumption ($P_a + P_s$).

The function of the Anawake system is independent of the vocabulary and language used in the application; all application-specific processing occurs in the programmable DSP chip. A single Anawake design, implemented as a low-pin-count chip, or as a macroblock in a standard-cell library, can be used in different speech applications without alteration.

3. Evaluating Anawake

A modified form of the system architecture shown in Figure 2 could be implemented entirely in software, dispensing of the Anawake system. A software implementation is possible if the DSP system has the ability to dynamically reduce its clock rate under program control, and realize a power savings from this speed reduction.

In this scheme, the functionality of the Anawake system (speech detection and speech buffering) is implemented as a program running on the DSP. Since this program would be simpler than the main speech recognition application, the DSP system could be set to a lower clock rate to run the Anawake emulation. We define the power consumption of the DSP system while running the Anawake emulation as P_e.

If the Anawake software emulation detects the presence of speech, it transfers program control to the main speech recognition software system. This transfer requires returning the processing speed of the DSP to its normal rate. As in Section 2, we define the power consumption of the DSP during speech recognition processing as P_n. The average power consumption of this system is

$$(1 - D)P_e + DP_n, \tag{2}$$

where D is the duty cycle of speech input as in Equation 1. If $D \ll P_e/P_n$, the average system power consumption is P_e. In contrast, the average system power consumption for the architecture that use a hardware implementation of Anawake is approximately $P_a + P_s$ (for $D \ll (P_a + Ps)/P_n$). Recall that P_a is the power consumption of Anawake, and P_s is the power consumption of the DSP chip in sleep mode.

An architecture that includes a hardware implementation of Anawake offers clear benefits if $P_a + P_s \ll P_e$. In the following sections, we compare the software and hardware implementations of Anawake functionality, in the context of this inequality. In this analysis, we model the single-chip DSP system as predicted in [3], with 0.25mW/MIPS power dissipation during normal processing. We assume the sleep-mode power dissipation P_s of this single-chip DSP is $1\mu W$. We also assume the on-chip analog-to-digital conversion system requires $100\mu W$ to condition and sample to speech signal. In the following analysis, we refer to a single-chip DSP with these specifications as the "reference platform."

4. Speech Detection

A variety of speech detection algorithms have been developed for use in voice-processing applications. The difficulty of the speech detection task is a function of

the characteristics of the background noise that is mixed with the speech signal, as well as the signal-to-noise ratio of the speech signal and the background noise.

4.1 Energy Contour Methods

Energy-contour approaches to speech detection work well in applications where the peak-speech-signal to average-background-noise ratio is at least 30dB, and where the background noise is quasi-stationary in nature [9]. Energy-contour methods are based on the assumption that a speech signal varies in energy at the syllabic rate, but background noise varies at a much slower rate. These methods estimate the energy contour of a sound signal using a fast, syllabic time constant, then normalize this contour by an energy estimate taken with a much longer time constant. This normalized energy contour is relatively independent of the level of background noise. Simple decision techniques, using thresholding (for on-line applications) or histograms (for off-line applications), can be applied to the normalized energy contour to accurately predict the start of a speech utterance [9].

A software implementation of an on-line energy contour speech detection algorithm requires about 0.1 MIPS of processing, for an 8 Khz speech sampling rate. Running this implementation on the reference platform would consume $125\mu W$ of power; power consumption is dominated by the A/D conversion.

Analog implementations of energy-contour speech detectors are used in speakerphone systems and other analog voice processing systems. Typically, these designs include a precision rectification circuit for energy detection, two energy-averaging circuits operating at different time constants, and on-line decision circuitry to flag the start of speech utterances. These circuits could be implemented using MOS transconductance amplifiers operating in the weak-inversion regime as building blocks, with an expected power consumption of under 100 nW.

4.2 Spectrally-Enhanced Energy Contour Methods

More sophisticated speech detection algorithms are needed for environments with higher level of background noise; energy-contour methods perform poorly under these conditions. One approach to noise-robust speech detection is to compute the power spectrum of the audio signal, and apply the signal-processing operations used in energy-contour approaches (time-averaging at multiple time constants, normalization, histograms) in a frequency-dependent way. A DSP implementation of this approach [10] uses about 6 MIPS of processing; this speech detection algorithm would consume 1.6 mW of power on the reference platform.

Spectrally-enhanced energy-contour algorithms are also amenable to micropower analog implementation. A micropower wavelet analog filterbank [6,7] can be used to provide the initial spectral decomposition; a 64-channel, logarthimic-scale filterbank, covering frequencies up to 4 Khz with 60dB dynamic range, consumes about $5\mu W$ of power. A variety of weak-inversion building blocks for array and scalar signal processing, as reviewed in [4], can be used for spectral post-processing and decision making. As the time constants of these secondary operations are in the millisecond range, the filterbank is the dominant source of power dissipation.

4.3 HMM-based Methods

The most sophisticated speech detection algorithms integrate speech detection into a hidden Markov model (HMM) based speech recognition system [9]. In HMM-based speech recognizers, a state machine formalism is used to represent the words in the vocabulary; to add speech detection functionality, an extra state is added to the grammar to represent silence between words. By computing the best path through this grammar for a given audio input, speech and non-speech segments can be accurately labeled.

This approach works well if speech and silence are the only types of sounds present in the audio input. To handle background noise, the front-end of the speech recognition system, which converts the audio waveform into a spectrally-based feature vector, is enhanced to generate similar feature vectors for silence and quasi-stationary noise [11].

As this speech detection approach requires running a complete speech recognition system continuously (i.e. $P_n = P_e$, using the notation of Equation 2), it is not a good candidate for use in an Anawake software emulation. In its original form, this approach is also unsuitable for a hardware Anawake implementation: an Anawake speech detector should be application-independent, but this method requires detailed vocabulary knowledge.

However, an application-independent grammar, that uses phonetic classes instead of vocabulary words to model speech, could be used in this speech detection architecture. Wordless grammars have been used to model arbitrary speech in a similar manner in language-identification systems [12].

An HMM-based speech-detection system, using an application-independent grammar, would be a challenging, but plausible, analog micropower design project. The main building blocks of a such a system (a speech recognition front-end [6,7], a hidden Markov model state decoder [8] and a probability generation system [5]) have been successfully implemented in micropower analog circuits. These implementations would suggest a realistic estimate of $8\mu W$ for a complete HMM-based speech-detection system. Integrating these blocks into a coherent, robust system would be the primary design challenge.

4.4 Analog Delay

A complete power analysis of Anawake must include the power dissipation of the analog delay. For a given sampling rate, the power dissipation of a sampled analog delay circuit is a function of the dynamic range of the delay. In the simple Anawake architecture shown in Figure 2, the audio signal path always passes through the analog delay; the dynamic range of the analog delay limits the performance of the system at all times. By including a DSP-controlled switch to selectively bypass the delay (Figure 3), the dynamic range of the analog delay becomes less critical.

In this scenario, the DSP samples both the delayed signal and the direct signal for the first 100 ms after returning to normal operation from sleep mode; subsequently, it only samples the direct signal. The minimal use of the delayed signal makes the limited dynamic range of a micropower analog delay circuit tolerable. An analog

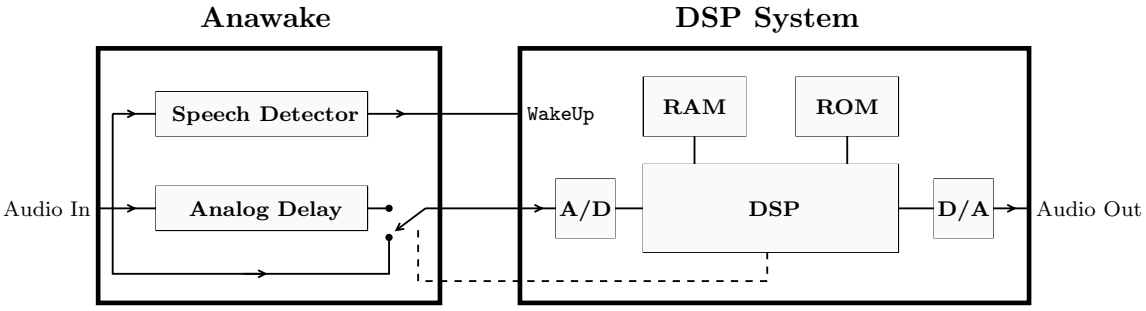

Figure 3. Modified Anawake design, to support analog delay systems with a reduced dynamic range.

delay with a $1\mu W$ power budget, organized as an analog memory array under digital control, would yield a dynamic range of about 40 dB at an 8 KHz sampling frequency. This estimate assumes the use of simple clocked-voltage-follower delay elements in the memory array. Advanced voltage-mode circuits, or a switched-current approach [13], may result in a better dynamic range figure.

4.5 Summary

Figure 4 summarizes the power consumption estimates in this section. For both simple and complex speech detection systems, an Anawake analog implementation offers a significant power advantage (factors of 50-500) over a software emulation on the reference platform.

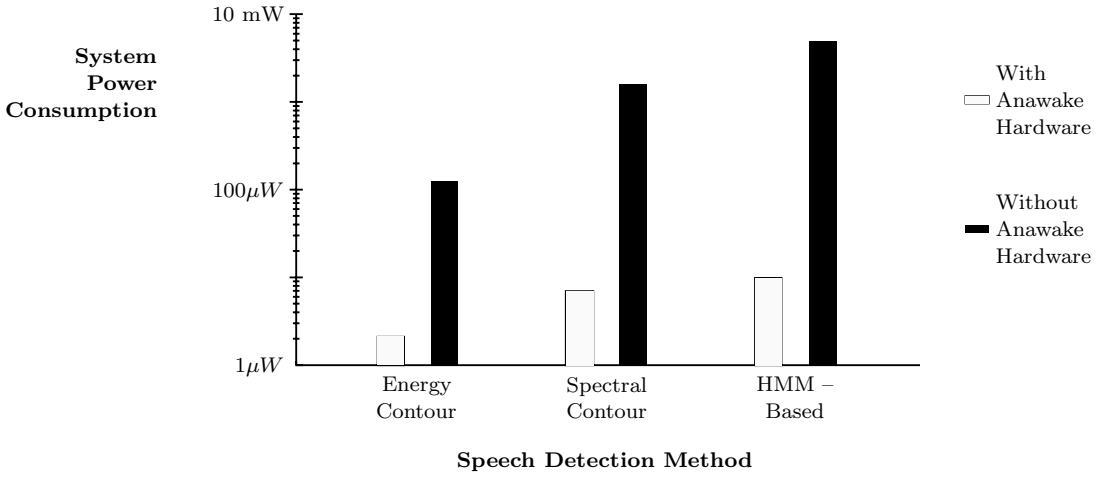

Figure 4. Bar chart showing the power consumption for systems using Anawake hardware (shaded bars, representing P_a+P_s) and software Anawake emulation (black bars, representing P_e). Note logarithmic scale. See text for details on speech detection methods.

5. Other Applications

Anawake applications are not limited to speech recognition systems. Battery-operated solid-state speech recorders could use Anawake to support a voice-actuated record mode. Another potential application for Anawake is power management for digital hearing aids. Digital hearing aids convert an analog microphone input to digital form for patient-specific signal processing; the processed signal is reconverted to an analog signal to drive an earphone. Hearing aids require very low power consumption, and are principally worn to aid voice communications; blocking out non-speech sounds to lengthen battery life is an acceptable tradeoff. The latency imposed by a 100 ms delay, however, is problematical. Modified speech detection systems, that shorten speech detection time in exchange for a higher false alarm rate, may make Anawake suitable for this application.

Other candidate Anawake applications involve the processing of non-speech (and perhaps non-audio) signals. For these applications, the speech detector is replaced by a detection system for the appropriate signal class. An example application is a remote, battery-operated seismic telemetry system; spurious vibrations should be ignored, but rare, earthquake events should be analyzed and reported. For this application, an Anawake system with a "probable seismic event" detector is needed.

6. Summary

We have presented a system architecture for signal-based power management. This architecture controls a programmable DSP system with a hardwired micropower analog system. Using the continuously-listening, battery-powered speech recognition task as an example, we showed how this architecture results in a system that consumes a few microwatts of average power, while maintaining the flexibility of a programmable DSP system.

Acknowledgments

We thank Dan Hammerstrom for his contributions to this work. We thank Xavier Arreguit, Chris Toumazou, and Michael Godfrey for comments on this work. This work was sponsored by the Office of Naval Research (URI-N00014-92-J-1672) and by the Department of Defense Advanced Research Projects Agency. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Air Force.

References

[1] D. Bursky, "Single chip employs neural-net algorithms to recognizer, synthesize speech," *Electronic Design,* p. 232, September 19, 1994.

[2] A. Chandrakasan and B. Brodersen, *Low power digital CMOS design,* Kluwer Academic Publishers, Boston, 1995.

[3] M. Gold and A. Bindra, "Confab: get ready for 400-Mips DSP," *Electronic Engineering Times,* issue 849, p. 1, May 22, 1995.

[4] E. A. Vittoz, "Analog VLSI signal processing - why, where and how," *Analog Integrated Circuits and Signal Processing,* vol. 6, no 1, pp. 27–44, 1994.

[5] R. Coggins, M. Jabri, B. Flower, and S. Pickard, "A low-power network for on-line diagnosis of heart patients," *IEEE Micro,* vol. 15, no. 3, pp. 8-25, 1995.

[6] L. Watts, D. A. Kerns, R. F. Lyon, and C. Mead, "Improved implementation of the silicon cochlea," *IEEE Journal Solid State Circuits,* vol. 27, no. 3, pp. 692-700, 1992.

[7] A. van Schaik, E. Fragniere, and E. Vittoz, "Improved silicon cochlea using compatible lateral bipolar transistors," in *Advances in Neural Information Processing Systems 8,* D. Tourestzky, M. C. Mozer, and M. E. Hasselmo, Eds., Cambridge, Mass: MIT Press, 1996.

[8] J. Lazzaro, J. Wawrzynek, and R. Lippmann, " A micropower analog VLSI HMM state decoder for wordspotting, " in *Advances in Neural Information Processing Systems 9,* M. C. Mozer, M. Jordan, and T. Petsche, Eds., Cambridge, Mass: MIT Press, 1997.

[9] J. G. Wilpon and L. R. Rabiner, "Application of hidden Markov models to automatic speech endpoint detection," *Computer, Speech and Language,* vol. 2, pp. 321-341, 1987.

[10] J. Pencak and D. Nelson, "The NP speech activity detection algorithm," *International Conference on Acoustics, Speech, and Signal Processing,* vol. 1, pp. 381-384, 1995.

[11] H. Hermansky and N. Morgan, "RASTA processing of speech," *IEEE Transactions of Speech and Audio Processing,* vol. 2, no. 4, pp. 578-589, 1994.

[12] Y. Yonghong, E. Barnard, and R. Cole, "Development of an approach to automatic language identification based on phone recognition," *Computer Speech and Language,* vol. 10, no. 1, pp. 37-54, 1996.

[13] P. Shah and C. Toumazou, "Trading off speed versus dynamic range in switched current circuits," *Proceedings International Symposium in Circuits and Systems,* vol. 5, pp. 297-300, 1994.