CONF-9506201--6

EVENT IDENTIFICATION BY ACOUSTIC SIGNATURE RECOGNITION*

William B. Dress and Stephen W. Kercel
Oak Ridge National Laboratory
P.O. Box 2008
Oak Ridge, Tennessee 37831-6011

To be presented at the
11th Annual Security Technology Symposium
Virginia Beach, Virginia
June 19-22, 1995

*Research sponsored by the U.S. Department of Energy under contract DE-AC05-84OR21400 with Lockheed Martin Energy Systems, Inc.

EVENT IDENTIFICATION BY ACOUSTIC SIGNATURE RECOGNITION*

William B. Dress and Stephen W. Kercel
Oak Ridge National Laboratory
P.O. Box 2008
Oak Ridge, Tennessee 37831-6011
(615) 574-4801

ABSTRACT

Many events of interest to the security community produce acoustic emissions that are, in principle, identifiable as to cause. Some obvious examples are gunshots, breaking glass, takeoffs and landings of small aircraft, vehicular engine noises, footsteps (high frequencies when on gravel, very low frequencies when on soil), and voices (whispers to shouts). We are investigating wavelet-based methods to extract unique features of such events for classification and identification. We also discuss methods of classification and pattern recognition specifically tailored for acoustic signatures obtained by wavelet analysis. The paper is divided into three parts: completed work, work in progress, and future applications.

The completed phase has led to the successful recognition of aircraft types on landing and takeoff. Both small aircraft (twin-engine turboprop) and large (commercial airliners) were included in the study. The project considered the design of a small, field-deployable, inexpensive device. The techniques developed during the aircraft identification phase were then adapted to a multispectral electromagnetic interference monitoring device now deployed in a nuclear power plant. This is a general-purpose wavelet analysis engine, spanning 14 octaves, and can be adapted for other specific tasks.

Work in progress is focused on applying the methods previously developed to speaker identification. Some of the problems to be overcome include recognition of sounds as voice patterns and as distinct from possible background noises (e.g., music), as well as identification of the speaker from a short-duration voice sample.

A generalization of the completed work and the work in progress is a device capable of classifying any number of acoustic events—particularly quasi-stationary events such as engine noises and voices and singular events such as gunshots and breaking glass. We will show examples of both kinds of events and discuss their recognition likelihood.

1. INTRODUCTION

The human attribute essential to many security functions that can be successfully automated at the present state of the art is the mind's ability to classify events based on limited sensory data.[1] To use a human operative in a security situation is highly effective, but it can be costly to the security program and dangerous to the operative. A smart electronic device can serve as a low-cost, low-risk replacement for a guard or observer. It can also perform repeatable high-speed interpretation of surveillance data, often replacing a human analyst. Replacing personnel, when practical, with hardware can result in a sharp reduction both in the cost of security programs and in the personal risk to operatives.

---

*Research sponsored by the U.S. Department of Energy under contract DE-AC05-84OR21400 with Lockheed Martin Energy Systems, Inc.

Many security functions can be facilitated by hardware that classifies signatures at audio frequencies. An obvious application is the covert monitoring of activity at remote airstrips in operations such as counter narcotics.[2] Signature classification hardware can also be used to detect breaches in secure facilities; it can be programmed to detect events such as gunshots, while not being confused by similar events such as thunder. In prisons, small, unobtrusive and inexpensive devices could be used to monitor weak points in security for distinctive sounds such as digging or fence climbing. In access control, a gate security system can be programmed to grant access only to a vehicle that emits an allowable acoustic signature. Other transportation security systems can be based on audio-frequency signature identification.[3-5]

Pattern recognition hardware has not been widely used because several inherent problems have remained unsolved. The most difficult technical problem is the identification of a suitable feature space.[6] A feature space is a mathematical space in which the attributes of samples of a given class occupy a limited region, while attributes of samples of other classes occupy other limited, yet distinct, regions.[7]

Although the Fourier frequency domain is often used in acoustic signal analysis, it is not a good feature space for classifying acoustic signatures. The discrete Fourier transform washes out time variations in the spectrum and introduces artifacts into the spectrum that are not present in the underlying signal.[8] In contrast, the wavelet transform resolves a signal into both scale and time components and provides good localization in both dimensions.[9] Wavelet scale is much easier to treat mathematically than is Fourier frequency. Wavelet time-scale space is highly successful as a feature space for acoustic signature classification.

The other problem inherent in pattern recognition is realization in inexpensive hardware. We have built and operated several prototype hardware wavelet engines using off-the-shelf digital signal processing (DSP) chips. The wavelet engine is nothing more than a set of finite impulse response (FIR) filters arranged in a straightforward structure.[10] Therefore, it is feasible to replace the DSP chips with dedicated FIR chips,[11] resulting in a wavelet engine that is small, inexpensive to produce in quantity, and simple to program.

2. COMPLETED WORK

2.1 Airport Monitor

Typical results of wavelet analysis of airplane acoustic signatures are shown in Fig. 1. Time series acoustic signatures for four different airplanes taking off from McGhee Tyson Airport (Knoxville, Tennessee) are shown projected onto the eighth level of a 12-level Daubechies wavelet. The first level

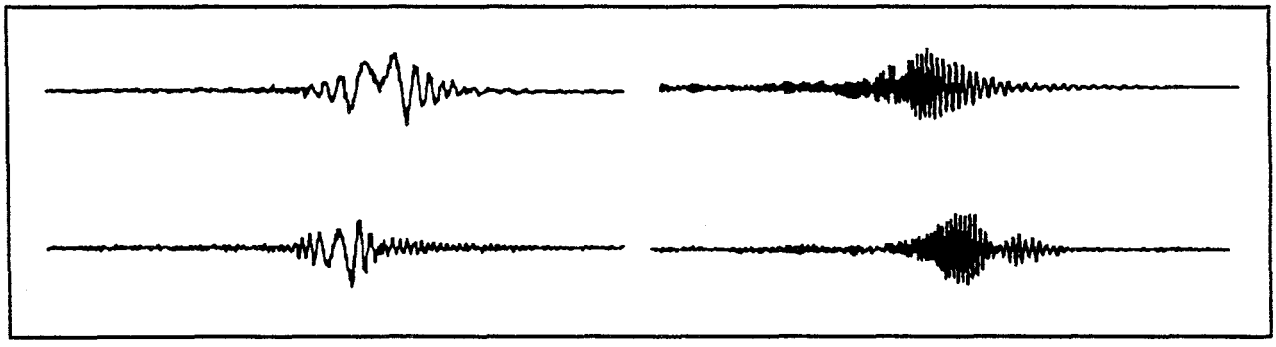


Fig. 1. Wavelet signatures of four airplanes in two classes.

is the finest scale; the twelfth is the coarsest. The Daubechies wavelet is implemented with a dyadic tree of 12 pairs of 16-coefficient FIR filters.[12]

The plots clearly show the similarities and differences in the signatures. Two different specimens of the same class of airplane, the Saab two-engine turboprop, have similar looking signatures. Two different specimens of another class of airplane, the Jetstream two-engine turboprop, have similar looking signatures. However, the Saab signatures look different from the Jetstream signatures despite the fact that the classes are similar, being two-engine turboprops of roughly the same size.

To make the pattern classification problem tractable on simple hardware, it is necessary to compress the thousands of data points that constitute the wavelet transform into a feature vector of reasonable size. Since most of the energy is in wavelet levels 5 to 8, an effective way to generate a feature vector is to use the discrete Fourier transform to obtain the spatial frequency spectrum at each level and to accumulate the results in ⅓-octave bins. This leads to eight bins per level, or a 32-dimension feature vector to characterize the entire signature.

We found that fuzzy logic was the most reliable method for classifying the resulting feature vectors. Each class constitutes a fuzzy set and is defined as the fuzzy union of the feature vectors of the training samples that typify the set.[13] To classify an unknown specimen, the degree of possibility of membership in each fuzzy set is computed. The fuzzy set to which the sample has the strongest possibility of membership is the class with which it is identified. Even with a sparse training set, this scheme resulted in better than a 90% rate of correct classifications.

2.2 Magnetic Spectral Receiver

We have developed a magnetic spectral receiver and deployed it at several nuclear power plant sites to perform unattended long-term monitoring of ambient magnetic fields.[14] This is not a pattern recognition application, but the specialized nature of the job required wavelet methods. Conventional field measurement techniques take one of two extremes in the trade-off between time and frequency resolution. Spectrum analyzers give extremely fine frequency resolution but eliminate all time information within the interval during which the signal was collected. Dosimeters respond to transient effects extremely localized in time but eliminate all frequency information within the band covered. Neither device provides an adequate description of ambient magnetic fields.

To characterize ambient magnetic fields in a nuclear power plant requires simultaneous time and frequency localization. The wavelet engine (Fig. 2) in the magnetic spectral receiver fulfills this function, implementing the minimum cost trade-off between frequency and time resolution. In frequency, it splits the signal into 14 one-octave-wide bands. In time, it responds to transient effects with a resolution of 820 μs in the lowest octave (305–610 Hz), 420 μs in the next highest octave (610–1220 Hz), and so on.

The multiresolution structure incidentally implements a fast wavelet transform in hardware. It is made up of 13 pairs of half-band Daubechies wavelet FIR filters with filter outputs decimated by two. The mathematical interpretation of the 14 outputs is that each constitutes a list of time-shifted wavelet coefficients at the 14 different scales. The input signal is a time series describing an event; the outputs constitute the wavelet transform. The circuit can be programmed for various functions, including feature extraction for pattern recognition.

3. WORK IN PROGRESS

Speaker identification and verification by scoring similarity of test phrases against a database of known speakers has been a longstanding problem. Limited success has been obtained by using Fourier-based methods coupled with hidden Markov models. Our most notable success has been in the limited problem area of comparing a known speaker speaking a known phrase of sufficient length to obtain adequate statistics for positive identification. We are extending the identification and verification effort

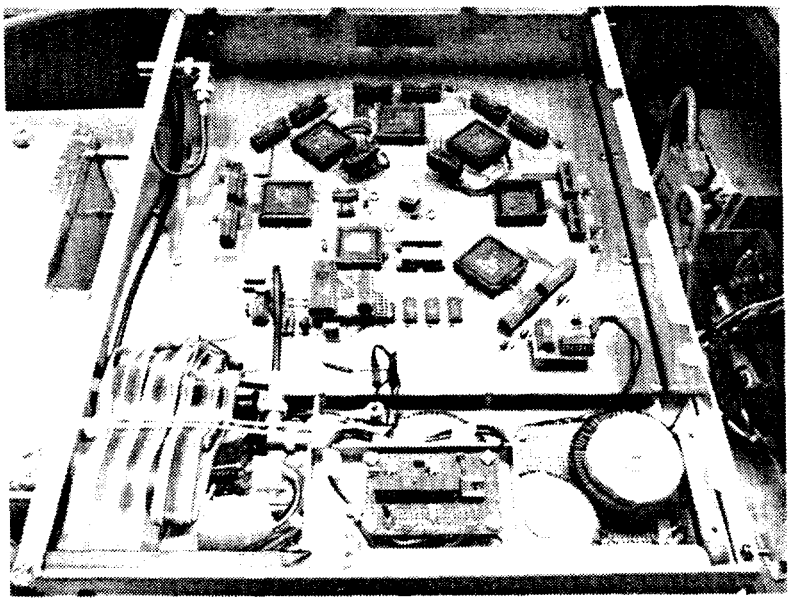


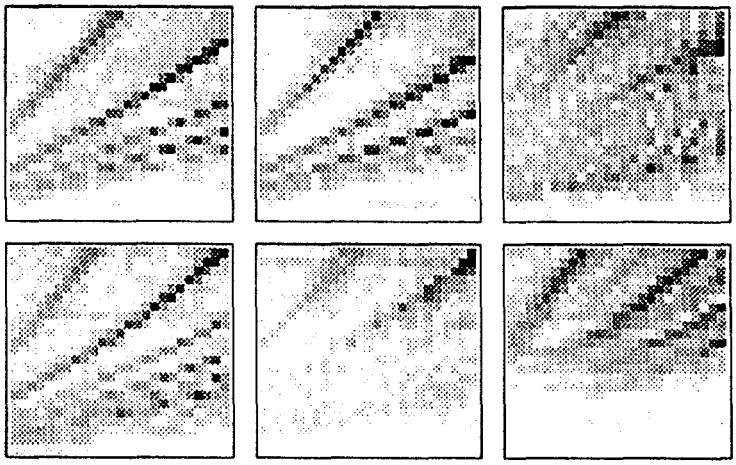
Fig. 2. Prototype wavelet board.

to identification of the voice independent of the phrase spoken with a goal of high-probability identification based on less than 2 s of speech.

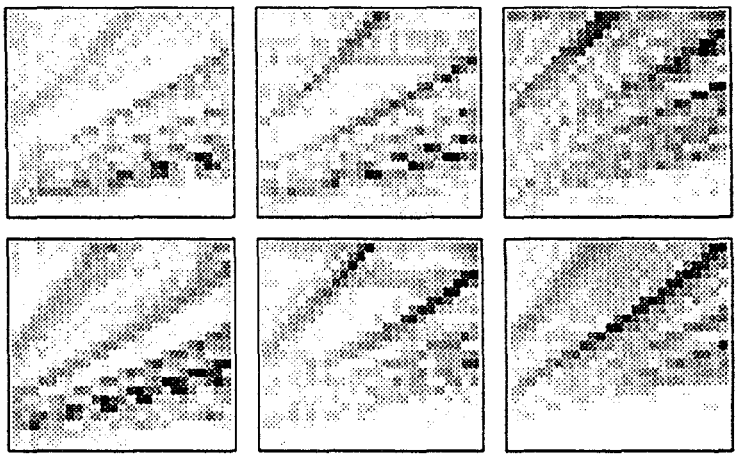
Figure 3 shows sets of ridge plots of a different phrase from each of two similar-sounding male speakers. These plots show intensity of Fourier component varying with wavelet scale (vertical axis) and wavelet wave number (horizontal axis). The upper-left corner in each of the subplots is at the lowest scale and lowest wave number used in the analysis. (The wave number is the number of oscillations in the mother wavelet used to analyze the sample.) Ridges at a high angle (e.g., 60°) to the horizontal axis correspond to low-frequency formants in the voice. A high-pitched voice (i.e., female) shows most of the ridge activity below the diagonal running from lower left to upper right. These voice-characteristic ridges resemble a set of "spokes" radiating from a center located off to the left of each figure. This spoke feature of these voice plots seems to be independent of the characteristic pitch of the voice being analyzed; indeed, the same phone by the same speaker at a different pitch produces a rotation of the spoke pattern (to first order), allowing easy adjustment for pitch and prosody.

A simple method of classification based on activity above and below the diagonal separates the entire speaker set (630 speakers) into two categories. Less than 1% of the speakers were misclassified as to gender based on this simple method. A categorization based on the average number of dominant ridges in a phrase can be used to correct the 1% classification error: male voices tend to have three to five strong ridges, while female voices exhibit one to three strong ridges.

Note the similarity between the sets of ridge plots. In particular, compare the first frame in the bottom row of Speaker A with the third frame in the bottom row of Speaker B. There is a single strong ridge approximately on the diagonal of the figure and a weaker ridge at a low pitch (high angle) above the diagonal. Below the diagonal are at least three weaker ridges in each, occurring about the same relative strengths and at nearly the same angles, indicating voices of very close pitches and similar quality. Although it requires careful listening to be able to state that these two samples are indeed taken from two different individuals, a statistical comparison of each set with the (much larger) identification sets shows statistical differences between these two samples and correctly identifies each speaker.



Speaker A. "Swing your arm as high as you can." Six 128-ms regions were selected out of 2 s of voice as being energetic enough to characterize this speaker.



Speaker B. "The saw is broken, so chop the wood instead." As above, six 128-ms regions were selected out of 2.9 s of voice for this speaker.

Fig. 3. Voice signatures.

Music exhibits the same basic ridge patterns as voices do, but the details of the ridge shape and position provide a means to discriminate between the two categories of sound. A limited number of musical phrases were analyzed and compared to both individual speakers and composite voice signatures consisting of major groupings in the 4000-phrase database. First indications show clear distinctions between music and voices in that a composite music pattern is significantly different from the composite American-voice pattern. This gives us confidence that machine separation of voice segments from music segments is feasible.

4. FUTURE WORK

4.1 Quasi-Stationary Events

The spectra of signatures of quasi-stationary events vary with time, but not sensitively. The time variation often constitutes the distinction that makes the classification of quasi-stationary signatures practical. The relative slowness of the time variation makes the classification process potentially cheap. Voices and vehicle sounds are typical quasi-stationary processes.

The hardware and algorithms described in this paper are well suited to quasi-stationary processes. They can be reasonably expected to produce consistently good classification results. Even with statistically sparse data, correct classification rates above 90% are observed.

The major practical problem in deploying these techniques in security system hardware is the collection of statistically significant numbers of signatures. A reasonable set consists of at least 40 training signatures and 40 test signatures per class. A prediction based on fewer data is more an expression of hope than confidence. A good set of signatures is relatively expensive to obtain. For example, it cost an average of two person-hours per signature to collect the data for the Airport Monitor project.

4.2 Singular Events

Singular events are impulsive in nature. There is not much periodicity in their signatures, and spectral analysis techniques do not lend much insight into their information content. Examples might include gunshots or breaking glass.

Singular events are often most effectively handled by time-domain techniques. Typical systems include time-domain reflectometry and broadband pulse radar. These produce a time series signature that is analyzed by direct comparison with signatures of known events. Mathematically, this comparison is accomplished by computing the correlation of the unknown signature (or perhaps time-resolved pieces of the signature) with known signatures for each of the possible events. Alternatively, a digital filtering process that is computationally cheaper but mathematically equivalent to correlation can be done. The signature that produces the highest correlation value identifies the class.

We are investigating precrash restraint actuation for the National Highway Traffic Safety Administration. Because a classification error can lead directly to loss of life, the error rate must be extremely low. A preliminary look at time-domain signature identification is encouraging, but it is premature to make any predictions about error rate.

5. CONCLUSIONS

We have successfully developed wavelet-based acoustic signature classification algorithms and the hardware on which to run them. The Airport Monitor demonstrates the success of the wavelet time-scale domain as a feature space for classifying airplanes from takeoff or landing acoustic signatures. The magnetic spectral receiver is a field-deployed 14-level wavelet engine. While the present application does not need to implement pattern recognition, the wavelet board could easily be reprogrammed to extract pattern recognition features. The research currently under way demonstrates a cheap and effective method to identify human speakers and to distinguish voices from other sounds. Its hardware is easily reprogrammable to other pattern recognition applications.

Fourier-based acoustic signature identification schemes have been tried for many years without much success.[15] The wavelet algorithms and devices described in this paper use real-world data and operate in real time or with "sample-and-hold" processes in near real time. The wavelet transform implemented on dedicated hardware appears to be the breakthrough needed to solve many previously intractable problems in acoustic signature classification.

For application to practical security problems, two tasks remain. First, a statistically significant number of signatures from each of the relevant classes of events must be collected to train the device for the desired application. This is a large expense but needs to be done only once per application. Second, an application-specific integrated circuit implementing the wavelet engine is needed. The initial development cost is high, but it would lead to a wavelet engine chip available for a few dollars per copy. This would clear the way for a whole family of small, inexpensive, and smart security devices.

REFERENCES

1. J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles*, Addison-Wesley, Reading, Mass., 1974, p. 5.

2. W. B. Dress and S. W. Kercel, "Wavelet Based Acoustic Recognition of Aircraft," pp. 778–791 in *Wavelet Applications*, Harold H. Szu, ed., Proc. SPIE 2242, 1994.

3. G. O. Allgood, R. K. Ferrell, S. W. Kercel, R. A. Abston, P. I. Moynihan, and C. L. Carnal, *Traffic Flow Wide-Area Surveillance System Definition*, ORNL/TM-12827, Oak Ridge National Laboratory, November 1994, pp. 39–41; available from National Technical Information Service, Springfield, Va.

4. H. Roe and G. S. Hobson, "Classification of Road Vehicles from Microwave Profiles," *IEE Sixth International Conference on Road Traffic Monitoring and Control, London, April 29–30, 1992.*

5. *Vehicle Detector Field Test Specifications and Field Test Plan, Task Report F for Detection Technology for IVHS Contract Number DTFH61-91-C00076*, Rev. 5, Hughes Aircraft Corp. and JHK Associates, February 1994.

6. Ref. 1, p. 244.

7. R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, 1973, pp. 17–20.

8. F. J. Harris, "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform," *Proceedings of IEEE* **66**(1), 51–83 (January 1978).

9. S. L. Robinson and P. F Ryczek, "Wavelets," *Mathematica J.* **5**(1), 74–81 (Winter 1995).

10. A. N. Akansu and R. A. Haddad, *Multiresolution Signal Decomposition*, Academic Press, San Diego, 1992, p. 133.

11. T. Lin, *Architectures and Circuits for High-Performance Multirate Digital Filters with Applications to Tunable Modulator/Demodulator/Bandpass Filter Systems*, Ph.D. Dissertation, University of California, Los Angeles, 1993.

12. I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992, pp. 194–195.

13. B. Kosko, *Neural Networks and Fuzzy Systems*, Prentice-Hall, Englewood Cliffs, N.J., 1992, pp. 269–289.

14. W. B. Dress and S. W. Kercel, "A Hardware Implementation of Multiresolution Filtering for Broadband Instrumentation," in *Wavelet Applications for Dual Use*, Harold H. Szu, ed., Proc. SPIE 2491, in press.

15. S. W. Kercel and W. B. Dress, *Pattern Recognition Features in Near-Runway Aircraft Acoustic Signatures*, K/ITP-482, Special Projects Program Office, Martin Marietta Energy Systems, Inc., 1992.