

ornl

ORNL/TM-13050

**OAK RIDGE
NATIONAL
LABORATORY**

MARTIN MARIETTA

RECEIVED

OCT 13 1995

OSTI

**Anticipatory Precrash Restraint
Sensor Feasibility Study:
Final Report**

Stephen W. Kercel
William B. Dress



DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MANAGED BY
MARTIN MARIETTA ENERGY SYSTEMS, INC.
FOR THE UNITED STATES
DEPARTMENT OF ENERGY

kg
MASTER

This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from the Office of Scientific and Technical Information, P.O. Box 62, Oak Ridge, TN 37831; prices available from (615) 576-8401, FTS 626-8401.

Available to the public from the National Technical Information Service, U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161.

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

**Portions of this document may be illegible
in electronic image products. Images are
produced from the best available original
document.**

Instrumentation and Controls Division

**ANTICIPATORY PRECRASH RESTRAINT
SENSOR FEASIBILITY STUDY:
FINAL REPORT***

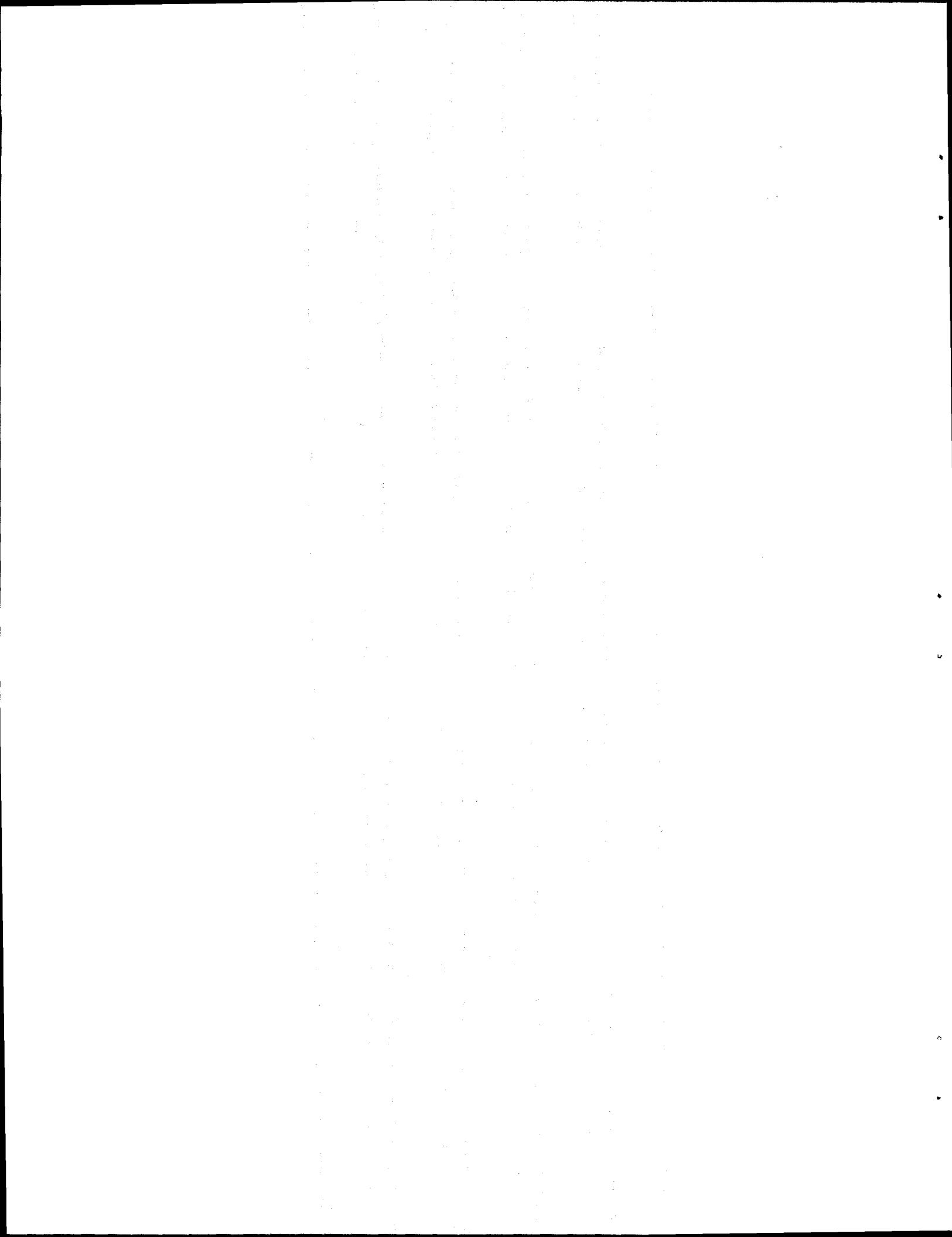
Stephen W. Kercel
William B. Dress

Date Published—August 1995

*This report was part of the Anticipatory Airbag Deployment project sponsored by the U.S. Department of Transportation's National Highway Traffic Safety Administration.

Prepared for the
NATIONAL HIGHWAY TRAFFIC SAFETY ADMINISTRATION

Prepared by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37831-6285
managed by
LOCKHEED MARTIN ENERGY SYSTEMS, INC.
for the
U.S. DEPARTMENT OF ENERGY
under contract DE-AC05-84OR21400



CONTENTS

List of Figures	v
List of Tables	vii
Executive Summary	ix
Abbreviations and Acronymns.....	xi
1. PROBLEM STATEMENT	1
2. OBJECTIVES	3
3. SURVEY OF THE LITERATURE	7
What does "anticipatory" mean?	7
Can (and should) targets be classified?	9
Can signal processing extract good dynamical data about targets?	11
Can a vehicle sense its own dynamical state?	11
Can precrash restraints leverage from related technologies?	12
4. PRACTICAL CONSTRAINTS	15
5. EXPERIMENT	17
Van Data: Collision Course.....	21
Modified Cart Data: Collision and Noncollision Courses	21
Car Data: Noncollision Course	24
6. MATHEMATICAL ANALYSIS.....	27
Data Analysis: General Comments	27
Fourier Analysis.....	27
Discrete Wavelet Analysis.....	27
"Van Slowing Down".....	27
Data Analysis: Some Results.....	28
Simple Analog Filter Model	28
FFT Extraction of Velocity	29
Discrete Wavelet Analysis.....	29
Chirplet Analysis	30
Nonlinear Model of a Single Cycle.....	32
Karhunen-Loeve Transform.....	34
Singular Value Decomposition	37
Auditory Analysis	39

Discussion: Danger Zones and Safety Zones.....	39
Beam-Forming Approach.....	40
Data Analysis: Comparison of Techniques	41
Techniques Not Explored in This Study	41
7. PROSPECTS FOR THE FUTURE	43
8. CONCLUSIONS	45
9. THE NEXT PHASE: A TASK LIST.....	47
Task 1: General Data Collection.....	47
Task 2: Dynamical Information Extraction	49
Task 3: Beam-Forming Proof of Principle.....	49
Task 4: Identification of a Sparse On-Vehicle Sensor Array	51
Task 5: System Definition of the Anticipation Engine	51
Task 6: Kinematic Simulation.....	52
Task 7: Survey of Existing Crash Data	52
Task 8: Target Classification	52
10. RECOMMENDATIONS	55
References	57
Appendix: Tutorial on Orthonormal Wavelets	61

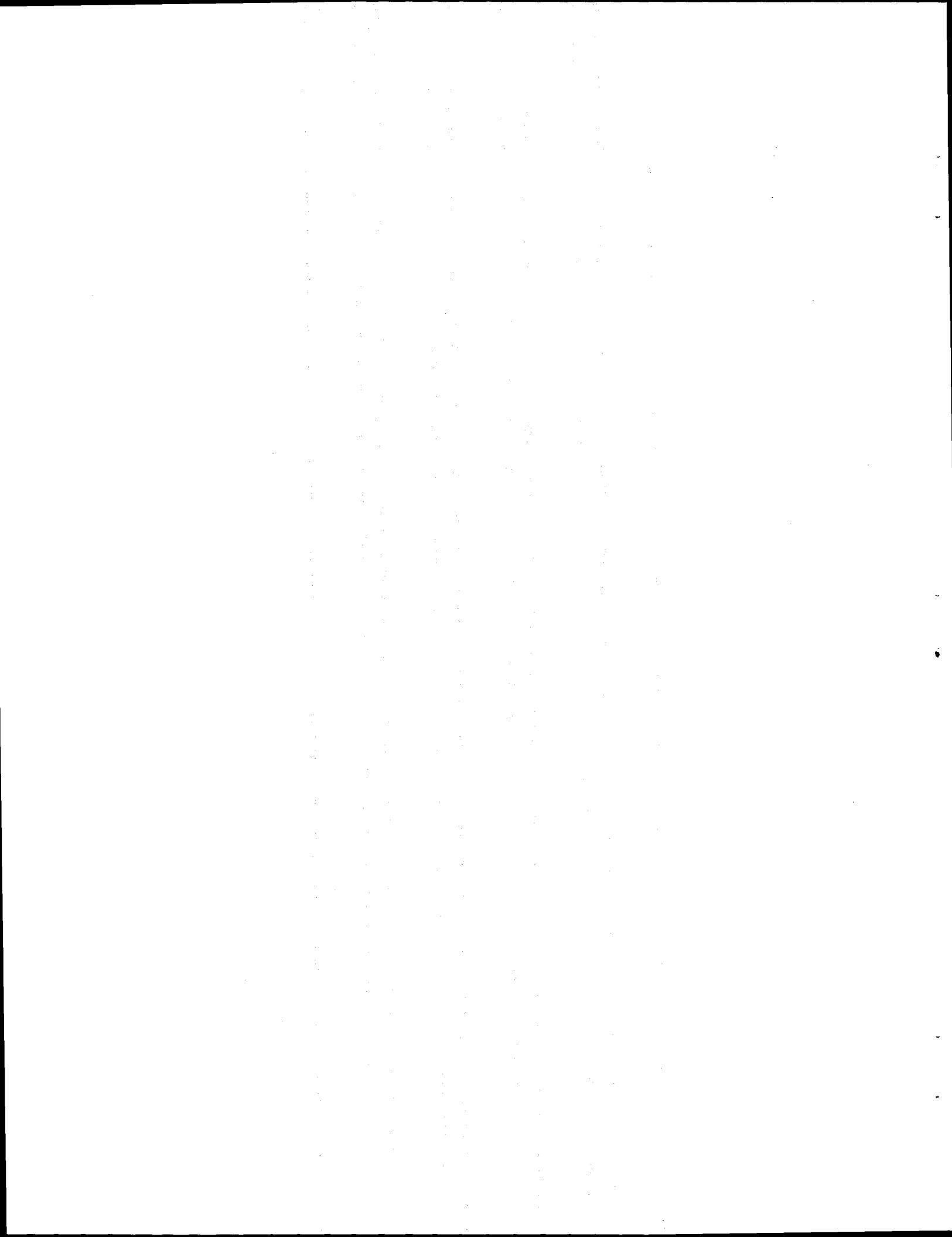
FIGURES

2.1	Schematic for a formal anticipatory system.	4
2.2	Schematic for extracting system information from sensor data.	5
2.3	Schematic for extracting target information from sensor data.	5
5.1	Ultrasonic ranging system.	18
5.2	Doppler radar system.	18
5.3	Doppler radar unit used in experiments.	18
5.4	Head-on geometry.	19
5.5	Pass-by geometry.	19
5.6	Side-swipe geometry.	19
5.7	Cart with reflecting screen.	20
5.8	Reflecting screen at 45° angle.	21
5.9	Screen supporting structure.	22
5.10	Doppler signature of van.	22
5.11	Doppler signatures of cart with screen.	23
5.12	Passenger car drive-by Doppler signature.	24
5.13	Receding passenger car Doppler signature.	25
6.1	Data plots for van slowing down.	28
6.2	Analog filtered output.	28
6.3	Fourier analysis of van slowing down.	30
6.4	Extraction of velocity from Fourier analysis.	31
6.5	Wavelet analysis of van slowing down.	31
6.6	Fourier analysis of wavelet level 2.	31
6.7	One row of the chirplet sets.	32
6.8	Fixed-speed signal correlation with fixed-speed chirplet.	33
6.9	Deceleration signal correlation with deceleration chirplet.	33
6.10	Nonlinear curve fitting.	33
6.11	Velocity obtained by nonlinear curve fitting.	35
6.12	Acceleration obtained by nonlinear curve fitting.	35
6.13	Signature of possible multiple reflections.	36
6.14	Eigenvalues of covariance matrix.	36
6.15	Projection onto the first three eigenvectors.	36
6.16	Resolution of projection into distinct surfaces.	37

6.17	Singular values of the covariance matrix.	38
6.18	Projection of acceleration signature onto SVD basis.	38
6.19	Projection of constant speed signature onto SVD basis.	38

TABLES

5.1	Data collection geometries	17
6.1	Comparison of analytical methods	42
7.1	Best-case costs for a future precrash restraint sensor system	44



EXECUTIVE SUMMARY

This report explores feasibility of an anticipatory precrash restraint sensor. The foundation principle is the anticipation mechanism found at a primitive level of biological intelligence and originally formalized by the mathematical biologist Robert Rosen. A system based on formal anticipatory principles should significantly outperform conventional technologies. It offers the prospect of high payoff in prevention of death and injury.

This study includes a survey of previous research. The formal anticipation engine is realizable, but the only past development has been for large-scale nuclear power plant control systems. An anticipation engine has never been attempted for a small instrument. Previous research in predictive crash detection has not used the formal anticipation paradigm. Sensors and processors are available to provide a good, fast, and inexpensive description of the present dynamical state of the vehicle to the embedded system model in the anticipation engine.

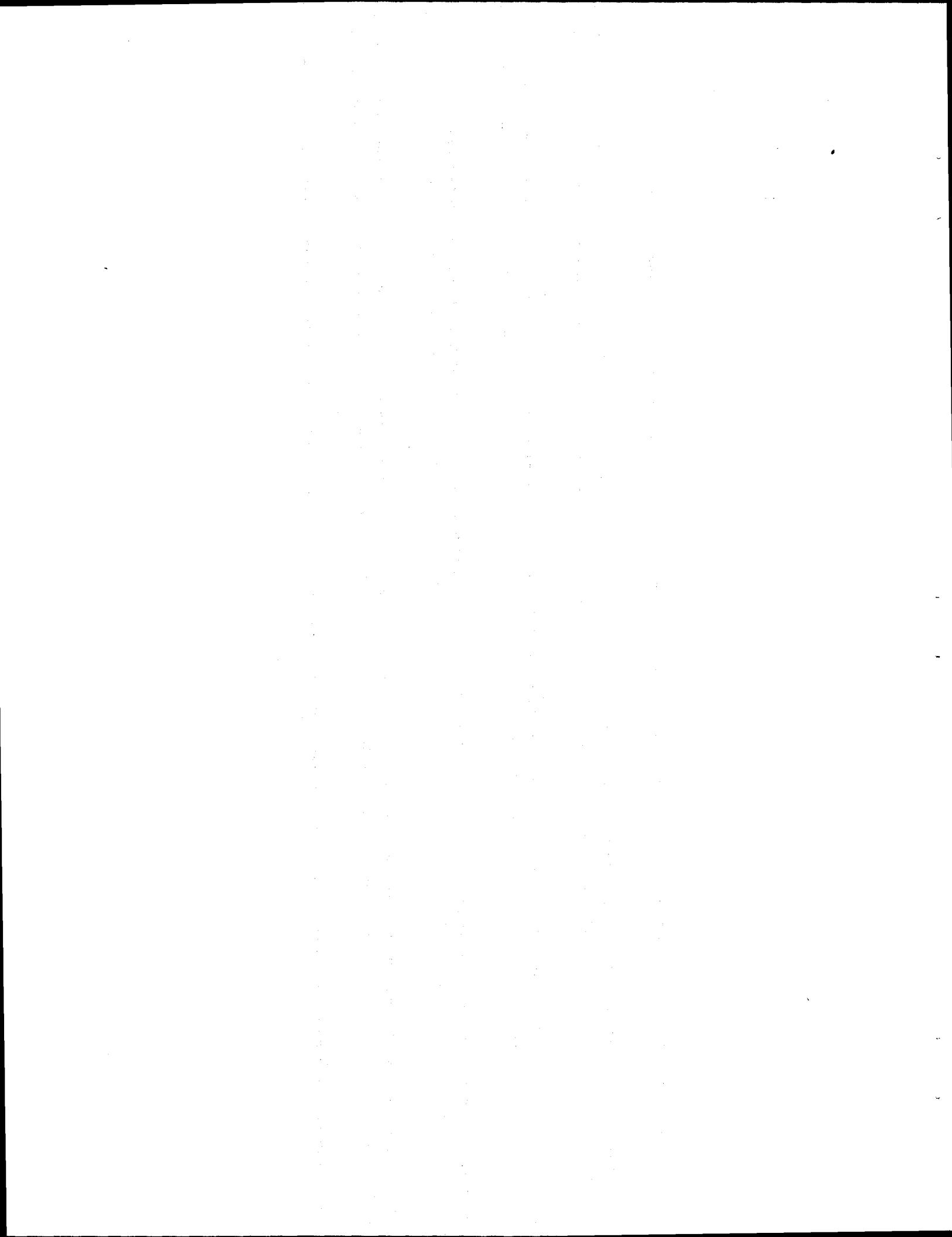
What is not discussed in the literature is whether or not there are sensors and processors that will provide a good, fast, and inexpensive description of the present dynamical state of the targets to the embedded environment model in the anticipation engine. The unavailability of dynamical information on targets would make the anticipatory system infeasible. The object of the experimental and analytical part of this study was to determine if target dynamical

information can conveniently be made available.

The experimental part of this study found that inexpensive radar in a "real-world" setting does return useful data on target dynamics. The velocity and acceleration of the target were extracted from a Doppler radar signal by seven different signal-processing methods. In addition, Doppler radar shows distinct signatures for distinct approach geometries.

The data produced by a radar system can be converted to target dynamical information by good, fast and inexpensive signal-processing techniques. The Karhunen-Loeve transform (KLT) is used to examine the upper limit on what information is likely to be revealed by sophisticated signal-processing methods. While too computationally costly to be used in a practical instrument, KLT is the best available algorithm for uncovering the information buried in a signal. The KLT of the Doppler radar time series appears to simultaneously resolve velocity and acceleration of multiple surfaces. It demonstrates that, in principle, all this information is extractable. Several other techniques, such as chirped wavelets, performed almost as well.

Not only is the anticipatory sensor feasible, but further development under the sponsorship of the National Highway Traffic Safety Administration is necessary and desirable. There are a number of possible lines of follow-on investigation. The level of effort and expected benefits of various alternatives are discussed.



ABBREVIATIONS AND ACRONYMS

ABS	antilock braking system
FFT	fast Fourier transform
FT	Fourier transform
FMCW	frequency-modulated continuous wave
ITS	intelligent transportation system
KLT	Karhunen-Loeve transform
NHTSA	National Highway Traffic Safety Administration
DSP	digital signal processing
STFT	short-term Fourier transform
SVD	singular value decomposition

1. PROBLEM STATEMENT

The ultimate goal of the anticipatory precrash restraint sensor project is the development of a sensory system that decides at the present moment, with great accuracy, whether or not to actuate a system of restraint devices such that the restraint is properly deployed a short time into the future. This is more than merely the prediction of whether or not a crash will occur. It is an estimate that the expected crash will be of sufficient severity to warrant device deployment. The vehicle will have a suite of restraint devices, driver and passenger airbags, side stiffeners in the right and left doors, and possibly other devices, each of which allows for a range of degrees of deployment. It is not desirable to fully deploy all restraint devices for every crash, and a context-sensitive decision must be made as to which devices to deploy, and to what degree.

Contrast the anticipatory sensor with an accelerometer-based sensor. The accelerometer indicates two things: whether or not a crash is already in progress, and if so, the severity of the crash. This is a purely reactive system. It knows nothing until the crash actually starts. The information it provides to the restraint system requires that the restraint system deal with a situation that is already occurring. For many types of crashes, there is not sufficient time to determine that a crash is already in progress, decide what to deploy, and complete the deployment before the crash energy is transmitted to the occupants of the vehicle. To obtain the necessary extra milliseconds, an anticipatory system is needed.

An anticipatory system is a formal mathematical scheme based on interacting predictive models. One model takes information about the past and present state of the vehicle and makes a prediction of its dynamical state in the near future. Another model takes information about the past and present state of the environment (i.e., likely targets) and predicts the dynamical state of the targets in the near future. Based on these sets of predictions, the system forms an expectation of whether a crash will occur, and if so, how bad will it be.

For a model to operate, it needs real-time data streams to provide the "present-state" information on the vehicle and the targets. This requires a suite of sensing transducers that provides "data," and real-time signal processing that extracts "information" from the data. Thus, the anticipatory system is not a sensor with some incidental built-in intelligence. Rather, it is an integrated intelligent system that incidentally uses an array of sensing elements, signal processors, and an anticipation engine.

The major unsolved problem is the intelligence, not the hardware. Specific questions are as follows: Can a formal anticipatory system consisting of interacting nested models reliably predict the onset and severity of a crash? What are the computationally cheapest models that provide an adequate prediction? What input information does the system model need, and how sparse can the information be without significantly diminishing the performance of the system model? What input information

does the environment model require? Does the output data of available sensor elements contain the information needed by the models? What are the computationally cheapest signal processing methods for extracting the information needed by the models from the data produced by the sensors?

The reason for using the anticipatory paradigm is that it has the potential to produce decisions that are simultaneously good, fast and inexpensive, and to produce them far more effectively than do conventional techniques. It does so by providing a method to make maximal use of minimal data. Development of the intelligence is expensive, but it is a one time-cost. The cost of producing copies of the developed intelligence for hardware systems is trivial by comparison.

The anticipation process is conceptually well founded. Anticipation is one of the most primitive functions of biological intelligence. Unlike higher cognitive functions, it is deterministic. It does not require that the creature doing the anticipating make a volitional choice to do so. Being deterministic, the rigorous development of a theoretical foundation for a mathematical description of anticipation is straightforward, albeit tedious.

Theoretical rigor has a profound practical consequence. With its foundations properly laid, the philosophical debate as to whether or not it is possible to mathematically describe anticipation is settled; it is. This is in stark contrast with machine cognition, which seeks to mathematically emulate volition and consciousness and which is the subject of heated and legitimate philosophical debate. It is truly unknown whether or not a machine can be programmed to form a concept, nor is it known whether we would recognize an instance of machine cognition if we saw one.

A major priority that remains to be defined by the National Highway Traffic Safety Administration (NHTSA) is whether the ultimate objective of this research is to use an anticipatory system to arm a restraint

device, or to fire a restraint device. There is a risk-reward tradeoff to be considered in this decision. The arming function is less expensive and less risky than the firing function, but it leads to a lower potential payoff.

It is tempting to consider the arming function to the exclusion of the firing function. Arming requires information about target dynamics, which can be obtained from radar returns. It does not require target class or mass (Najm 1995). The consequences of error may be tolerable. A false positive means that the system arms, but no crash occurs. If a crash does not occur within some prescribed amount of time after arming, the system can disarm, probably with no harm done. A false negative means that a crash occurs, but the system does not arm until forced to do so by the crash detection accelerometers; in that instance, the advantages of the anticipation are lost. Swihart and Lawrence have produced experimental data that show at least a 10% reduction in firing time for accelerometers supplemented by radar-based crash prediction for arming (Swihart and Lawrence 1995).

The firing function is costlier and riskier but leads to a greater potential payoff than the arming function. In addition to target dynamics, the firing function requires information about target class and mass. It has not been proven that class and mass data are obtainable by a device that must meet the constraints of an automotive precrash restraint sensory system, nor was it an objective of this study to explore the question. The consequences of error are quite severe for an anticipatory firing system. A false positive is very likely to cause a crash where none would have occurred otherwise. A false negative might cause the restraint to fail to actuate during a crash, causing more severe injury (or loss of life) than would have otherwise occurred. The payoff for an anticipatory firing system is that it would deploy the restraint device several tens of milliseconds faster than a contact-based firing system.

2. OBJECTIVES

This report covers only the feasibility study phase of the project. The objective of any feasibility study is to answer the question, Can the thing being studied really be done? A reasonable way to answer the question for this case is to determine whether or not there are "showstoppers" that might preclude the implementation of an anticipatory sensory system for the proper deployment of precrash restraints.

Anticipatory does not merely mean predictive. It means a system of nested predictive models that emulates biological anticipation and is generally consistent with formalism laid down by Rosen (1985). The scheme is shown in Fig. 2.1. The anticipation engine does not appear to be a showstopper. Anticipation is rigorously justified in mathematical theory.

The dynamical models for crash prediction are based on deterministic classical physics. The development of the specific models themselves would be tedious but straightforward. The models should be implementable in no more than a few thousand machine language instructions. Given present-day microprocessor clock rates, it should be possible to transform the initial dynamical conditions into a prediction in less than a millisecond.

Rosen identifies five necessary attributes that distinguish an anticipatory system. The first is that an anticipatory system, S_2 , must contain a model, M , of another system, S_1 . Second, the anticipatory system, S_2 , contains a set of observable quantities that can be linked mathematically to S_1 and an orthogonal

set of observables that cannot. Third, the predictions of the model, M , can cause an observable change in the state of S_2 . Fourth, there must be some observable difference in the interaction between S_1 and S_2 when the model is present and when the model is not. Finally, M must be a predictive model; based on the present conditions, M must change state faster than S_1 (operate faster than real time) such that M 's changed state constitutes a prediction about S_1 .

Since the vehicle and its environment (the set of objects into which the vehicle might crash) are everyday sized objects moving at ordinary speeds, the equations of classical dynamics should provide an adequate model:

$$\frac{dx_i}{dt} = f_i(x_1, \dots, x_n), \quad i = 1, \dots, n$$

$$\frac{dy_j}{dt} = g_j(y_1, \dots, y_m), \quad j = 1, \dots, m.$$

Here, the vector x_i characterizes the state of the vehicle at an instant, and the vector dx_i/dt , characterizes the time rate of change of the state of the vehicle at an instant, where the state has i degrees of freedom, and f_i is a mapping function that is not necessarily linear and not necessarily conservative. The vector y_j characterizes the state of the environment at an instant, and the vector dy_j/dt characterizes the time rate of change of the state of the environment at an instant, where the state has j degrees of freedom, and g_j is a

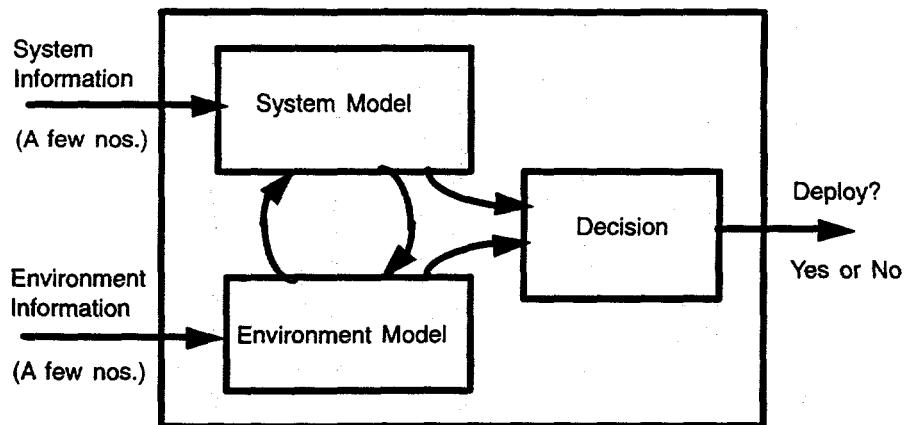


Fig. 2.1. Schematic for a formal anticipatory system.

mapping function that is not necessarily linear and not necessarily conservative.

Since they must operate faster than real time, neither model is an exact characterization of the system it describes. A major task of a subsequent phase of the development of a practical airbag actuation sensor is to determine the minimal set of vectors x_i and y_j that account for the necessary observables, and the functions f_i and g_j that adequately characterize the system for this particular application.

How these models and the interaction between them are to be implemented in the sensory system hardware is a computational detail. One promising method would be to implement the two dynamical models as two different cellular automata, each on its own set of simple dedicated massively parallel very-large-scale integrated circuit (VLSI) hardware. The interaction between them could be modeled as fuzzy set membership; there already exists a dedicated fuzzy logic chip that is probably suitable to the task. However, this example illustrates only one possible method. Others might turn out to be faster, better, or less expensive.

Sensing the state of the vehicle and extracting the information from the sensor data should not be a showstopper. On-vehicle sensors for velocity, acceleration, strain, etc., represent a mature technology, and processing can be done in real time with cheap dedicated digital-signal-processing (DSP) chips. The

scheme is shown in Fig. 2.2. In fact, it is the abundance of inexpensive sensor data describing the state of the vehicle that defines the engineering problem for this part of the system. How sparse can the data set be made to still provide an adequate description?

If there is a showstopper, it is in developing the information about the state of the environment. Optical techniques are not practical; they are too easily disrupted by environmental effects. Radio-frequency (RF) radar offers the proper range and resolution but, historically, has been extremely expensive and has required absurdly fast processing times. If RF radar is used in an impending crash detector, are there ways of implementing it within the cost constraints of consumer electronics? Does RF radar generate the information that the environment model needs? Can target dynamics be extracted from the data fast enough for it to be practical? Can target class be extracted at all? The scheme for extracting target information from sensor data is shown in Fig. 2.3.

Identification of target class may be as important as measurement of target dynamics, and it is a fundamentally more difficult problem. There is no general theory of pattern recognition against which conceptual level questions can be tested. Notwithstanding some of the claims reported in the literature, an accurate, reliable, and statistically valid ground vehicle classification system has never

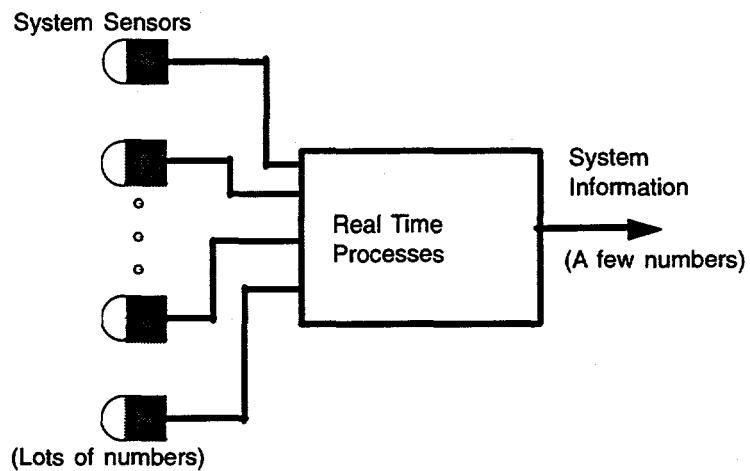


Fig. 2.2. Schematic for extracting system information from sensor data.

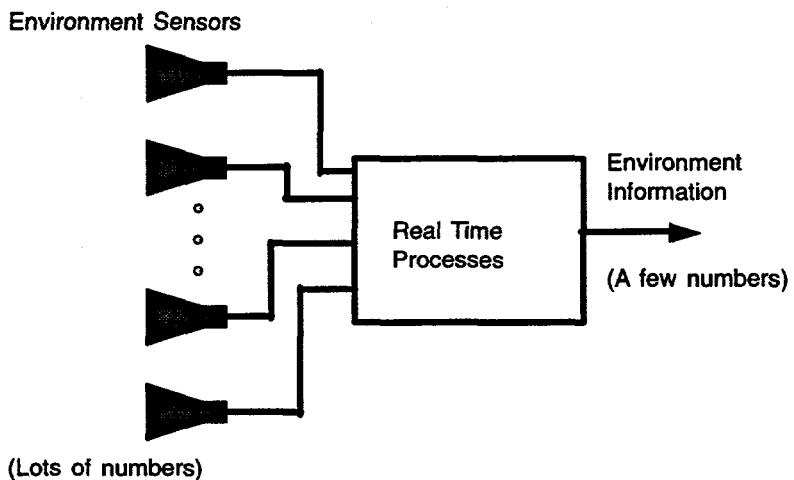


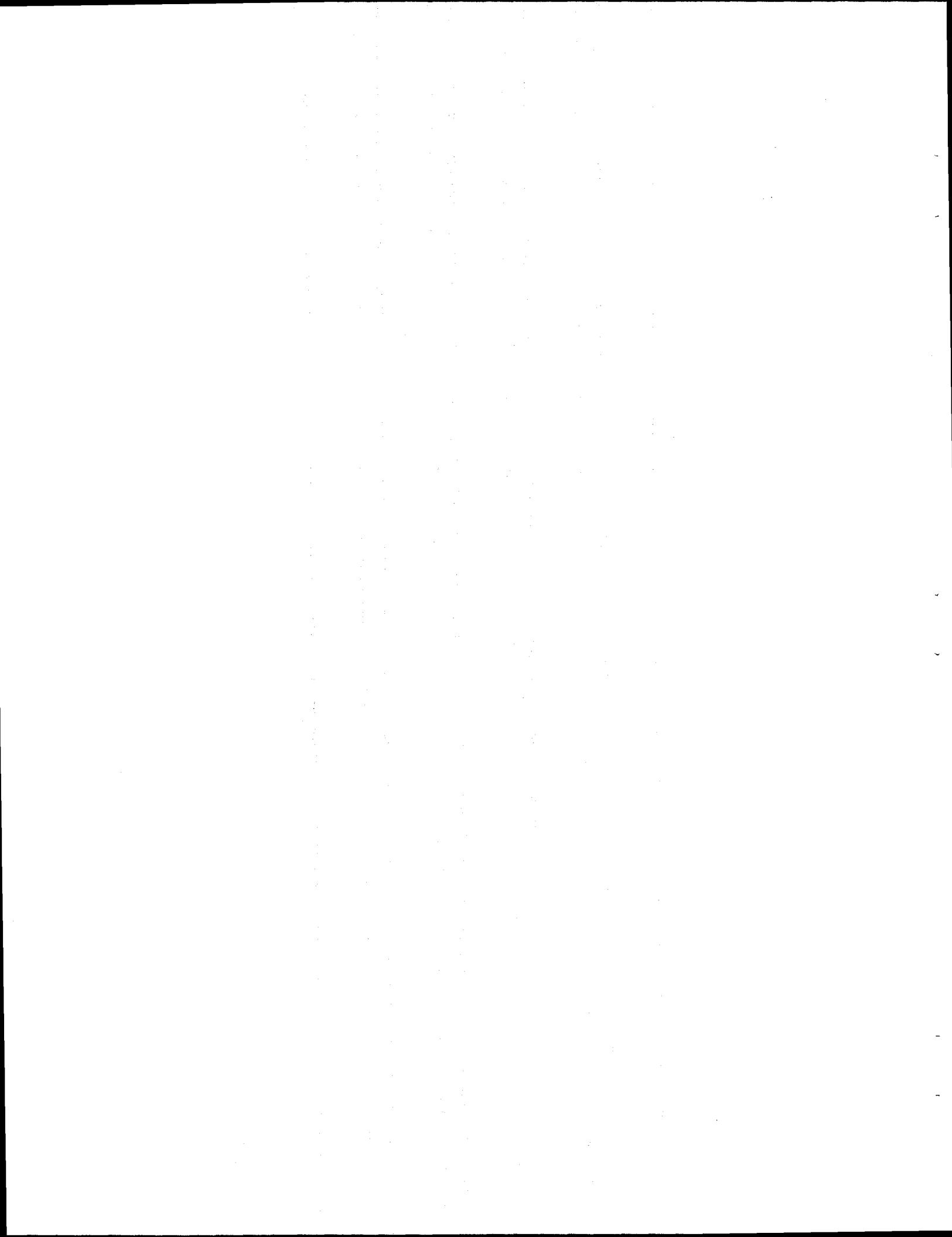
Fig. 2.3. Schematic for extracting target information from sensor data.

been demonstrated. The optimistic assessment of the literature is that several proof-of-principle studies have been published, and none of them precludes the possibility that vehicle class might be extracted from either acoustic signatures or radar returns.¹

To show that the anticipatory airbag sensing system is feasible, the experimental and analytical research effort in this project has concentrated on extracting information on

target dynamics from the output data of cheap sensors. This project does not confuse the system with its transducers; the global objective remains the development of an anticipatory system. Rather, the feasibility study focused its primary effort on analysis of environment transducer outputs because uncertainty about the availability of environment information constitutes the most likely showstopper.

¹By mutual agreement between NHTSA and the authors, experimental and mathematical study of vehicle classification was specifically excluded from the scope of this feasibility study. All comments on vehicle classification appearing in this document are based on a review of the recent literature, and on the authors' experience in pattern recognition and signature analysis research for other projects.



3. SURVEY OF THE LITERATURE

What does "anticipatory" mean?

A formal anticipatory system is based on mathematically formalized principles of anticipation. The term generalizes the notion that the system can take present action based on an expectation of a future state. It refers to a system in which decisions are made by an algorithm that emulates the anticipation process found at a primitive level of biological intelligence. The principles of anticipatory systems were rigorously derived by Robert Rosen, a mathematical biologist (Rosen 1985).

The most important use of the anticipation mechanism in nature is to preserve the safety of the creature doing the anticipating. Based on an extremely sparse set of percepts describing the present state, a creature performs the remarkable feat of recognizing, with sufficient time to take corrective action, whether or not the future state constitutes a danger. Since eons of natural selection have caused the anticipation mechanism to abound in nature, not only must it be effective, but also it must have superior survival value compared to other paradigms for identifying threats.

If we learned how to perform the seemingly impossible task of heavier-than-air flight by observing how it was done in nature, does it not make sense to try the same thing in the development of robust safety systems? This was precisely the approach taken by Tsoukalas in the development of anticipatory controls for large systems such as nuclear power plants (Tsoukalas 1989). Since the anticipatory paradigm has proved itself in

the ongoing struggle for the survival of the fittest, it is reasonable to expect that a safety-system technology based on anticipation should significantly outperform conventional technologies.

The implementation of a formal anticipatory system in a small instrument has never been reported in the literature. The most likely reason for this is that, before the recent development of high-performance processors and algorithms, it was probably impractical. However, as this feasibility study indicates, at the present level of technology, the development of an anticipation engine in a small system is the next logical step in the progression. It is reasonable to expect that the development of an anticipatory system applied to vehicular safety should lead to a high payoff in reduction of accidental death and injury.

NHTSA-sponsored research in predictive crash sensing goes back to the work in the early seventies by Holstrom and associates (1993). These researchers defined active restraints as devices requiring some action on the part of the vehicle occupant, such as seatbelts, and defined passive restraints as devices not requiring action by the vehicle occupants, such as airbags. Airbags were treated as a system under development, to be initially installed at some time in the near future. They envisioned that airbags would be actuated by accelerometers mounted on the vehicle firewall and observed that the long actuation time of accelerometers (20–40 ms after crash contact first occurs) greatly diminishes the potential benefit of airbags.

Holstrom and associates used the term "anticipatory sensor" simply to mean a device that provides some sort of warning of an impending crash before contact occurs. Nothing in their work suggests the idea of a formal anticipatory system based on nested models. There is no reason to expect that they should have discussed formal anticipatory systems. It would be nearly 15 years before Rosen would lay down the principles (Rosen 1985). Even if they had known the theory, they would not have considered it for on-vehicle use. A practical implementation of a formal anticipatory system for vehicular safety would have been utterly beyond the reach of the computer technology of a quarter century ago.

Their objective was to provide crash warning when the target is 50–100 cm from the victim. The reasoning was that the extremely short range minimizes the false alarm problem, while the 25 ms or so gained by firing the airbag without waiting for the crash to start dramatically enhances the effectiveness of the airbag. They were generally more willing to tolerate false negatives (failure to predict a crash that actually occurs) than false positives (airbag actuation when no crash occurs).

Holstrom and associates did not prescribe an acceptable false alarm rate but speculated that it should fall in the range of one occurrence in 4 years (probability of accident involvement) to one occurrence in 2500 years (probability of fatal injury). Their rates of occurrence were based on statistics for the early seventies. Statistics for 1990 result in averages of one police reported crash involvement per 14.8 years of driving and one fatal crash involvement in 2910 years of driving (Knipling 1995).

The decision that Holstrom et al. were seeking is logically crisp. The system either says that a crash is about to occur (fire everything), or that no crash is impending (do not act). There is no provision for actuating some devices but not others, and no provision for variable actuation, such as context-

sensitive control of the degree of inflation of an airbag.

Their cost constraint seems reasonable. They asserted that the maximum that an original equipment manufacturer is willing to pay for such a sensor is \$10–\$20. Taking a quarter century of inflation into account, this is consistent with our estimate of \$50–\$100 per sensor system. In both cases it is assumed that the production volume would be 1 million to 10 million per year.

These researchers investigated several sensing technologies, seeking presence, closing rate, and the nature of the target. They examined and rejected several candidate technologies, finally settling on RF radar as the most promising possibility. They used a 10-GHz Doppler radar with a Gunn diode, not dissimilar to our RF setup, but they separated receiving and transmitting antennas such that the sensitivity included a highly localized "hot spot" along the centerline of the car and 1 m in front of the radar system. Distance information was implied by the presence of a target in the hot spot. Velocity was extracted from the Doppler return by straightforward analog processing.

They considered target classification to be crucial to a good crash sensor. They sought to distinguish between two classes, hazardous and innocuous targets. Although they did not resort to modern pattern recognition jargon, they attempted to do so by classifying the Doppler radar signatures by visual inspection of Doppler returns in time domain in hopes of finding a distinguishing threshold. They were not especially happy with the results; some innocuous targets gave bigger Doppler bursts than some hazardous targets. They did not seek distinguishing features in any transform feature space.

They concluded by recommending a hybrid system. They asserted that accelerometer type actuators are adequate below 30 mph but that precrash warning is needed at higher speeds and suggested a system with both accelerometers and radar. Below 30 mph only the accelerometers would

be used, in conventional fashion, and the radar would not operate. Above 30 mph, the radar would provide advance warning, and the accelerometers operating at a low threshold would provide confirmation. In the 30- to 60-mph range, they did not recommend making a decision to deploy based solely on the radar data.

Given the processing technology available at the time, the proposed system seems quite reasonable. It is remarkable that it never caught on.

The same research is recounted in much greater detail in a Department of Transportation (DOT) report (Hopkins et al. 1974). It is worth noting that these researchers expected the sensor to detect 60 to 80% of major targets but to have an extremely low rate of false positives. The report is a more detailed proof of the claims asserted in Holstrom's IEEE paper. Most importantly, as with the IEEE paper the detailed report uses "anticipatory" to mean "precrash"; it does not describe a formal anticipatory system.

More recent literature changes the descriptive adjective from "anticipatory" to "predictive," but the emphasis is still on the sensing hardware element and not the processing. It does not discuss formal anticipatory systems. Swihart and Lawrence (1995) are concerned with the same result as the early NHTSA work: precontact warning of an impending crash. Their application is more demanding, requiring information for context-sensitive deployment. Their system would use radar-based prediction to perform the arming function, and not the firing function. Their experimental data shows at least a 10% reduction in firing time for accelerometers supplemented by radar-based crash prediction for arming. They note that a system that fires an airbag solely on the basis of a noncontact prediction requires not only much better target dynamics than their system provides but also target-class information.

Swihart and Lawrence's objective is to extract crash-warning information from the radar returns of intelligent cruise-control

systems. They define a false alarm as strictly a false positive. They see three tasks for the crash warning system: (1) target identification (presumably, this means deciding whether or not the sensor has a target in view, rather than identifying the kind of target), (2) trajectory computation (presumably the extraction of target dynamics), and (3) threat/no-threat classification. They acknowledge the possibility that knowledge of the victim's dynamics, extracted from onboard sensors, could enhance the reliability of the crash prediction, but they do not visualize it in terms of nested interacting models of the system and its environment.

Curiously, while their paper implicitly recognizes the need for a prediction of the severity of the impending crash, they do not see target class as an important datum. They are extremely concerned with the intelligent deployment of restraint devices so as to minimize injury. However, they do not list target classification as a necessary task in their operating paradigm.

Can (and should) targets be classified?

Both past and recent research considers the class of the target to be a significant datum to be produced by an impending collision detector (Holstrum et al. 1973; James and Sampan 1995). A possible basis for classification is acoustic signature. It is instructive to consider a recent paper on vehicle acoustic signature classification (James and Sampan 1995). The paper illustrates what can go wrong when the attempt to solve a pattern recognition problem becomes confused with the analysis of the internal workings of a neural network.

The fundamental error in (ab)using neural networks is the notion that it eliminates the need to understand the physical process that produces the samples to be classified. According to Dr. Harold Szu, of the Naval Surface Warfare Center, one of the world's authorities on neural networks, just about any set of labeled numbers thrown at a neural network will result in a classification rate that is 80% correct (Szu 1995). By itself,

increasing the size of the training sets does not improve the classification rate.

In order for any classifier, including a neural network, to give a higher classification rate, the signatures to be classified must be projected onto a proper feature space. 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 (Duda and Hart 1973). The most difficult part of any pattern recognition problem is the identification of the feature space (Tou and Gonzalez 1974). Establishing a feature space requires the projection of representative signatures of each class onto sets of basis vectors until a basis set is found that shows good localization by class. Selecting the basis sets to use for classification inevitably requires an appeal to the physical process producing the signatures.

The scheme reported by James and Sampan starts with the idea that a long time-series vector can be compressed into a short vector by an algorithm that somehow looks reasonable but is chosen without recourse to physical reality; there is no indication as to whether or not useful classification features are being discarded. A long list of compressed signatures, each labeled by class, is used to train the classifier. Despite the fact that each of the four classes uses hundreds of training samples, a close reading of the paper shows a correct classification rate of about 80% for four classes.

For pattern classification, an acoustic signature containing thousands of data points must be compressed to a vector of no more than a few dozen dimensions. However, the compression scheme cannot be arbitrary. The idea that acoustic signals have their energy unevenly distributed in various mathematical spaces is well known. The principle for selecting a compression algorithm consists in projecting it onto various spaces until one is found in which there is considerable localization of energy and vast regions of the space that contain little of the signal's energy.

The uninhabited regions can be discarded. What remains is a compressed version of the original with considerable reduction of data but little reduction of information. The selection of spaces must be based on some understanding of the physical process generating the signal.

The standard of comparison for supervised classifiers is the Bayesian classifier (Tou and Gonzalez 1974, pp. 124-30). There is an argument from game theory that asserts that supervised classification is a zero-sum game. A consequence of that assertion is that the Bayesian classifier leads to the lowest probability of classification error. While it is computationally costly to implement, it represents a reasonable limit of good performance. Once a feature space is found and verified by a Bayesian classifier, the remaining engineering task is to devise a classifier that is almost as good as Bayesian, but cheaper to implement. Comparison with a Bayesian classifier constitutes a fair test of a neural net.

A supervised classifier that is known *a priori* to require four classes and that has a correct classification rate of 80% when trained with hundreds of samples from each class is not a very good classifier. With a proper feature space and a classification algorithm that operates efficiently in the feature space, four classes and hundreds of training samples per class should lead consistently to classifications that are greater than 99% correct.

The wavelet transform is an excellent technique for extracting features from vehicle signatures. The work of Karlsen et al. (1995) provides a technique for selecting a basis from the physical attributes of a ground-vehicle acoustic signature. While their work is preliminary, it provides encouragement that distinctive ground-vehicle features can be reliably extracted from the acoustic signature. The time localization of the wavelet enables it to extract discontinuous features from noisy data (Szu et al. 1995).

Some vehicle classification development has been done for infrastructure-mounted

sensors. Duckworth et al. (1994) report that gross classifications of vehicles (e.g., big truck vs little car) can be made from short-term Fourier transforms (STFTs) of the returns from infrastructure-mounted Doppler radar. Microwave profiling with frequency-modulated continuous wave (FMCW) radar is reported to have a 75% correct classification rate for five classes of vehicles (Roe and Hobson 1992). A highly reliable vehicle classification scheme based on laser profiling is reported by Schwartz (1994). It makes no claim as to correct classification rate. The foregoing examples are mentioned for completeness only. They depend on a global view of the target that can only be obtained with an infrastructure-mounted sensor. It is not practical to adapt them to on-vehicle use.

Classification of vehicles based on interpretation of radar returns from on-vehicle sensors has not been widely studied. Some work has been done in Germany, using two-dimensional scanning multibeam radar (or lidar) (Najm 1995). Given the present state of the art, this scheme is very expensive. Classification of vehicles based on returns of collision-avoidance type radars has evidently never been attempted.

It is possible, but not certain, that wavelet processing of on-vehicle radar returns might lead to a robust vehicle classifier. Since the wavelet retains features discarded by STFT, it may be possible to improve the classification rate of the system of Duckworth and associates with wavelet processing. Wavelet techniques have been highly effective in military-radar classification schemes (Topiwala and Teng 1995; Baras and Wolk 1995).

Perhaps an anticipatory firing system requires identification of target class, but this is not a foregone conclusion. There is general agreement that the detector must produce an estimate of the severity of an impending crash and that the estimated severity is directly proportional to the energy that the target will transfer to the victim. However, the energy of a target is proportional to its

mass (presumably implied by its classification) and, more importantly, proportional to the square of the target's velocity (easily extracted from the dynamics). Thus, a 3000-lb compact car closing at 100 mph relative speed will produce a crash 2.5 times more severe than a 30,000-lb truck closing at 20 mph relative speed. It is not clearly established whether or not identification of target class is worthwhile even for a precrash firing sensor.

Can signal processing extract good dynamical data about targets?

Many different sensor technologies have been explored for crash-avoidance sensing. A systematic comparative survey of these technologies was performed by Najm (1994). His emphasis was on the sensor technology rather than on the signal processing details. He notes that filtering or gating of radar returns is used to provide an estimate of target dynamics. Useful dynamical information does appear to be present in practical radars; and good, fast, cheap signal processing should do a better job of extracting it than is being done by present commercial devices.

While the recent literature on crash warning provides insights into the capabilities and limitations of radar, it does not provide (nor was it intended to provide) much guidance of the processing side of the precrash restraint problem (Takimoto and Kotaki 1992). For example, a vehicle-control system investigated by Ozguner et al. (1995) uses a good radar-reflecting target as an integral element of the overall system. Helpful exceptions include the work of Fujita, Akuzawa, and Sato (1995) and Najm, Mironer, and Fraser (1995); both of these studies provide insight into the extraction of target dynamics from sensor data.

Can a vehicle sense its own dynamical state?

As with most sensing devices, accelerometers have a limited dynamic range.

The proper use of an accelerometer depends on making the limited dynamic range compatible with the range of values that it is expected to sense in its intended application. Thus, conventional airbag accelerometers must not saturate, or "top out," at the high end of the range, but have no need to provide high-resolution data about small accelerations. At the other end of the range, accelerometers used in noncrash situations, such as intelligent suspension, are not expected to operate properly when exposed to 50 g acceleration during a crash but are expected to provide high-resolution information during noncrash operation.

Conventional airbag actuation uses on-vehicle sensing to detect whether or not a crash is currently in progress. The technology has about two decades of commercial practice behind it and in a sense is quite mature. A fairly extensive selection of silicon micromachined automotive sensors with physical sensing and electronic preprocessing fabricated on a single substrate are presently available (Grace 1991; Bryzek 1992).

Conventional airbag actuation is based on the very large accelerations of an ongoing crash and therefore is mostly concerned with accelerations at the high end of the range between 0 and 50 g. Conventional airbag accelerometers lack the resolution at low accelerations to provide good precrash dynamical data (Najm 1995).

What mostly appears in the recent literature is research in the extraction of ongoing-crash information. A major issue is whether single-point or multiple-point sensing is necessary to adequately describe the vehicle dynamics. Kelley (1993) suggests that single-point monitoring will suffice but requires rather a good deal of signal processing. Alrabady and Mahmud (1993) have developed an actuation algorithm based on estimating the severity of an ongoing crash from acceleration and jerk information extracted from accelerometer data. Allen (1992) has developed an algorithm based on the first derivative of the power output by the accelerometer to estimate the severity of the

ongoing crash. These provide little guidance in sensing the precrash dynamical state of the vehicle.

In more recent applications, the limited dynamic range of the accelerometer is placed at the more sensitive end of the scale. The devices used in intelligent suspension systems have a useful dynamic range from just above zero to 5 g. They provide the highly resolved details about acceleration that are required by intelligent suspension systems. These devices are already appearing on Lincoln Town Cars and are about to be introduced on less expensive vehicles. The price of this technology is falling rapidly, and in a few years it should be widely used. By the time an anticipatory precrash restraint sensor is ready for market, it should be a straightforward task to extract all the necessary data about the dynamical state of the vehicle from the data stream being provided by existing on-vehicle sensors primarily intended for other applications (Najm 1995).

The unresolved engineering problem in having the vehicle sense its own dynamical state is the determination of the minimal set of precrash information actually required by the system model, and the signal-processing task of extracting it from the output of existing on-vehicle sensors. This problem appears to be straightforward, and its solution does not seem to require any new conceptual breakthroughs. It does not need to be addressed in the next phase in the development of the anticipatory precrash sensor.

Can precrash restraints leverage from related technologies?

The problem to be addressed for anticipatory precrash-restraint actuation sensing is fundamentally different from problems of crash-warning systems and intelligent cruise control sensing. The anticipatory precrash-restraint sensor must compute fast enough to provide an accurate prediction of an event less than 100 ms into the future involving nearby targets, as

opposed to 2 or 3 s, and quite lengthy headways, for the other two systems (Takubo 1995; Kikuchi, Ishiyama, and Nakajima 1995). Thus, the processing algorithms must be extremely fast compared to those needed for the other technologies. On the other hand, the only targets of interest in precrash-restraint actuation sensor are those that can cause a crash in the next 100 ms, and the field of view and the overall amount of information to be processed is much less than that for the other two systems.

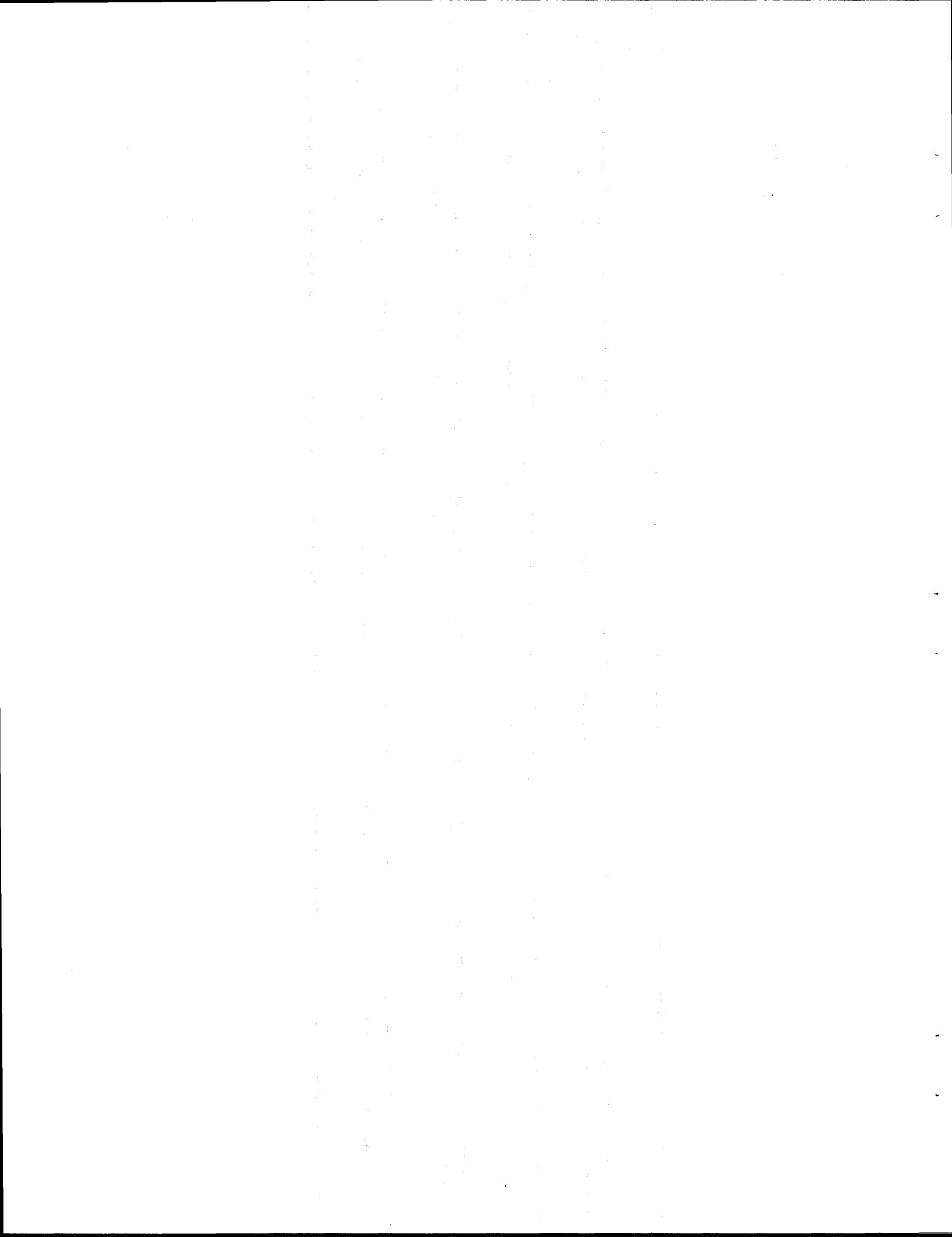
The element that the precrash-restraint actuation sensor has in common with the other crash-predicting technologies is the sensor element. Almost all propose using microwave radar. Infrared proximity detection works for robot guidance but only in low-speed indoor situations (Malik and Yu 1992). The same has been demonstrated for ultrasonic sonar (Kweon et al. 1993).

The precrash restraint system differs from the others in how it processes the radar return signals. In fact, it can be reasonably envisioned that a single radar might be connected to three independent processors, one to feed a driver crash-warning system,

one for an intelligent cruise control system, and one for precrash restraint actuation.

In systems designed to provide crash warning to the driver, there are two problems. The first is to accurately predict a crash. A more difficult problem is to provide warning to the driver in such a way as to elicit the proper evasive action from the driver (Hashimoto, Sasaki, and Kawai 1995; Chakroborty and Smedley 1995). This human-factors issue is the topic of interest in much of the crash-warning research and is particularly of interest to NHTSA's Office of Crash Avoidance Research (Leasure and Burgett 1994). Human factors are not an issue in a precrash-restraint sensory system, since its output goes directly to a machine.

NHTSA shows a similar priority in its obstacle-detection research (NHTSA 1994). Ultrasonic-based rear and side object detectors for use with big trucks were systematically compared. These systems all used ultrasonic detectors with a range of perhaps 3 m. However, the big issue was human factors. The report is primarily concerned with how the driver reacted to the warning.



4. PRACTICAL CONSTRAINTS

The work of Holstrom and associates (1973) suggests that measuring target dynamics is not especially difficult. However, what really matters is target class, and this is very difficult to determine. It should be noted that the signatures collected in their research were photographs of oscilloscope traces, and that these were visually inspected only. It might be profitable to collect the same type of Doppler data from various target classes in digital form and try to find a good feature space.

Holstrom et al. also discovered that ultrasonic sonar was not very good at extracting target dynamics but seemed to be better than radar at distinguishing target classes. It may be fruitful to consider a hybrid system in which the radar detects the target at longer range, and when it comes into sonar range, actuates a sonar device to get a classification trace.

It is constructive to recognize that extraction of target dynamics and target class are fundamentally different problems. The same sensor may or may not be the best device for both tasks. Various types of radar technologies, when correctly processed, appear to be feasible for sensing target dynamics. It is also possible, but not yet practically demonstrated, that target class might be extractable from radar returns.

Acoustic signature analysis is unlikely to be practical for extracting dynamical information about oncoming targets but might be used for target classification (Dress and Kercel 1994). In this application, acoustic signatures might come in two flavors. Active

acoustic signatures would emanate from moving targets such as other vehicles. Passive acoustic signatures emanating from the victim vehicle and reflected by fixed targets might also be detectable. With some knowledge of the victim's normal acoustic signature, it might be possible to correlate the return with the known signal and have a quick determination of whether the signature is active or passive. Experimentation with acoustic signature analysis is recommended to determine its usefulness.

The conventional wisdom is that false positives are less dangerous than false negatives. In a false negative, the detector fails to predict a crash that is about to occur. Nevertheless, a few tens of milliseconds later, an accelerometer fires the airbag anyway. More injury occurs than if the noncontact detector had worked, but the overall restraint system did provide some protection. In the vast majority of cases in which the detector does predict the crash (estimated in the literature to be 60–80%), the system provides the maximum protection. It is noteworthy that any assessment of the consequences of error must take into account the intended application of the sensor, whether to arm or to fire.

Suppose that the arming function is considered to the exclusion of the firing function. The consequences of error may be tolerable. A false positive means that the system arms, but no crash occurs. If a crash does not occur within some prescribed amount of time after arming, the system can disarm, probably with no harm done. A false

negative means that a crash occurs, but the system does not arm until forced to do so by the crash-detection accelerometers, and as noted above, the system provides some protection, but not the best protection.

The firing function is much riskier. Since it requires identification of target mass and/or class (which the arming function does not) as well as dynamics, and since mass/class identification techniques are much more error-prone than dynamics extraction techniques, it is reasonable to expect that an anticipatory firing system will produce significantly more wrong predictions than an anticipatory arming system. The consequences of error are quite severe for an anticipatory firing system. A false positive is very likely to cause a crash where none would have occurred otherwise. A false negative might cause the restraint to fail to actuate during a crash, causing more severe injury (or loss of life) than would have otherwise occurred.

One area that has not been explored is the assessment of the confidence level of predictions. Using the techniques of fuzzy logic, it may be possible not only to predict an impending crash, but to make a prediction along the lines of "this vehicle presently has X.XX possibility of being involved in a crash in the next 100 ms," where X.XX goes from 0.00 to 1.00. This should be explored in a future phase of this work.

For practical deployment in a vehicle, the anticipatory system must meet several constraints. Since the predictions pertain to the state of affairs approximately 100 ms in the future, the process of converting sensor data into a prediction has to happen in much less than 100 ms. The error rate (false positive and false negative) must be extremely low. Finally, the whole system must have a reasonable potential to be producible for a few tens of dollars per copy when produced in quantities of several million.

5. EXPERIMENT

We did a few experiments with a Polaroid ultrasonic ranging system developer's kit. We set up a real-time ultrasonic sonar with the Polaroid sensor, a Keithley A-to-D converter, and a personal computer. In addition, several wheel-actuated microswitch strips were placed in the road to provide reference speed and displacement data as the vehicle passed over them. The schematic arrangement is shown in Fig. 5.1. We used the detector with a Ford Econoline van and found that the range was about 10 to 20 ft.

We do not rule out the use of ultrasonics, particularly in a hybrid system with other sensors. However, we had an extremely limited budget for experimentation during this phase, and we considered that the exploration of RF radar would yield the greatest amount of insight in the least amount of time. Several manufacturers claim that they will soon be able to provide radar systems for a few dollars per copy when produced in volumes of millions. At present, however, a single unit radar set is expensive. We investigated the possibility of using an Amerigon pulsed broadband radar. Amerigon offered the loan of a unit but required a payment of \$14,000 for custom engineering. Our budget did not provide for such a large expense.

Consequently, most of our experiments were done with a surplus burglar alarm radar unit purchased from an amateur radio supplier for \$20 (Figs. 5.2 and 5.3). This was a Doppler radar that puts out 10 MW at 10 GHz. It uses a Gunn diode both as the transmitter oscillator and as the receiving local oscillator, and a hot-carrier diode as a

mixer. The supplier did not furnish a manufacturer's specification sheet. A horn antenna is used with the radar; its beam width is 23° in azimuth and 21° in elevation, and it has a gain of 16.6 dB. The Doppler output was connected to the same A-to-D converter as was used for the sonar experiments.

We did Doppler radar experiments with three vehicles—the Ford van; a small cart with a flat, adjustable reflecting screen; and a passenger sedan—in different geometries as listed in Table 5.1. One set of experiments involved a Ford Econoline van in the head-on geometry shown in Fig. 5.4. A similar set of experiments involved a passenger sedan in three geometries: the pass-by configuration shown in Fig. 5.5, with the radar antenna pointing normal to the direction of motion of the target vehicle; and two side-swipe configurations, shown in Fig. 5.6, with the antenna cocked 45° off normal pointing toward the front of the oncoming target, and with the antenna cocked 45° off normal pointing toward the back of the receding target.

We mounted a flat reflecting screen on a small cart as shown in Fig. 5.7. The angle of

Table 5.1. Data collection geometries

Configuration	Van	Cart	Sedan
Head-on	x	x	
Side-swipe		x	x
Pass-by	x		x

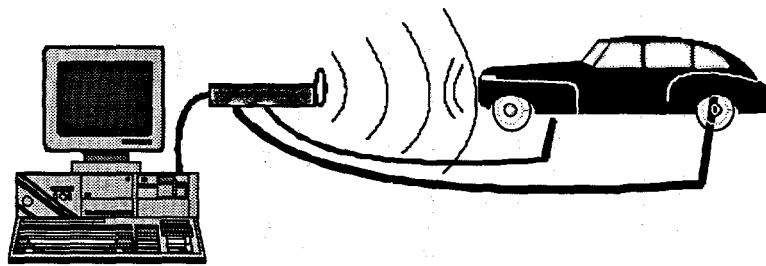


Fig. 5.1. Ultrasonic ranging system.

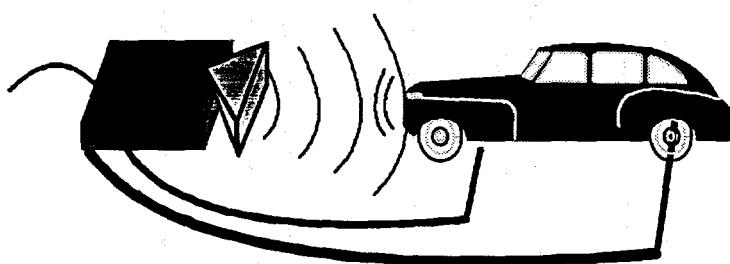


Fig. 5.2. Doppler radar system.

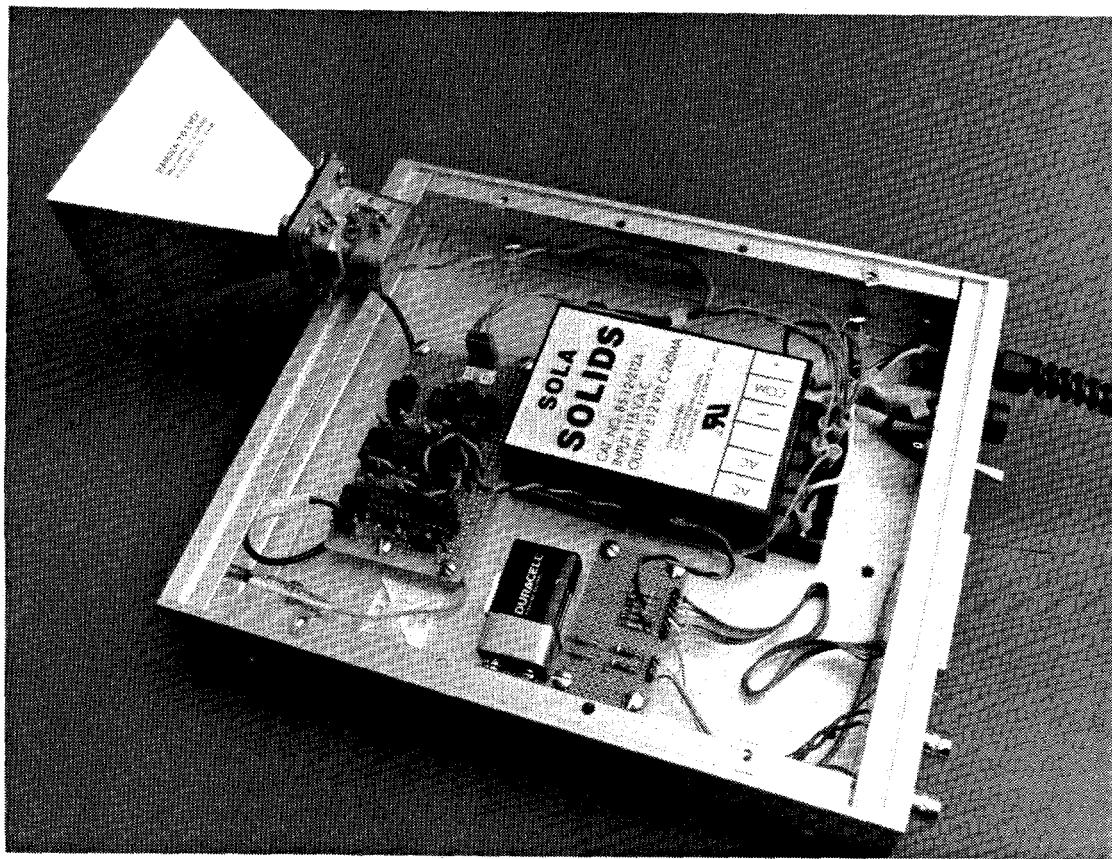


Fig. 5.3. Doppler radar unit used in experiments.

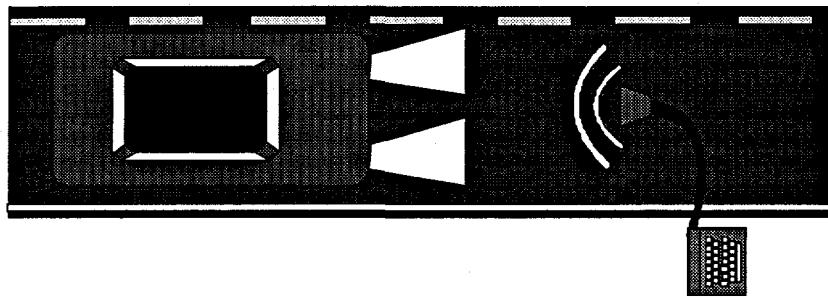


Fig. 5.4. Head-on geometry.

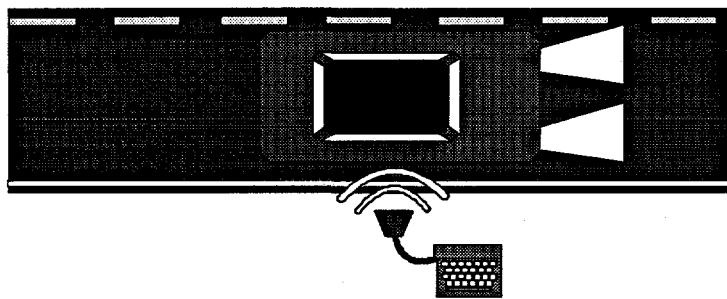


Fig. 5.5. Pass-by geometry.

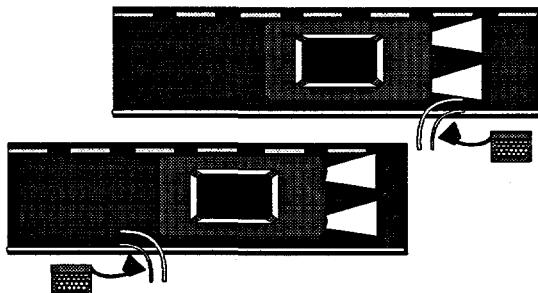


Fig. 5.6. Side-swipe geometry.

the screen with respect to the axis of the cart was adjustable, as shown in Figs. 5.8 and 5.9. The idea was to observe the effect of the angle of reflection on the reflected signal. We ran the cart directly toward the Doppler radar sensor with the screen in both normal-to-axis and 45° -off-axis positions in the head-on geometry. We also ran the cart by the sensor in the other two geometries.

The experiments were conducted with the radar antenna 3 ft above the ground. In the head-on experiments, the vehicle was run directly toward the antenna. The vehicle was 50 ft away at the start of the run and 5 ft away at the end of the run. In the side-swipe and pass-by geometries, the radar was positioned back from the edge of the roadway, and the perpendicular distance from



Fig. 5.7. Cart with reflecting screen.

the antenna to the side of the vehicle was 12 ft at nearest approach.

Some of the tests were conducted in light rain and snow. Doppler radar returns were not affected by mildly adverse weather.

The objective of these experiments was proof of principle. Can we establish that the radar returns really contain the information needed to feed the environment model? The experiments were not conducted with the exhaustiveness or the precision needed to establish engineering specifications.

The only independent check of target dynamics was provided by a set of two wheel-actuated switches. These provide an accurate measure of average target velocity in the time interval during which the front wheel of the target passes between the two switches. Data describing the profile of velocity as a function of time were not collected. In a course of experiments to establish engineering

specifications for a practical sensor (as opposed to proof of principle, as was done here), the velocity profile should be collected.

The duration of each signature was 2 to 4 s. A sampling rate of 1000 samples per second was used for digitizing. A typical Doppler signature has 2000 to 4000 samples.

What the Doppler experiments yielded were sets of time-series data corresponding to the instantaneous velocity of real-world targets, collected under real-world conditions with cheap hardware. The major result of these experiments is that a good estimate of real-time acceleration can be obtained. The implication is that with a system (such as pulsed radar) that also provides displacement, it should be possible to obtain good estimates of acceleration, displacement, and velocity. Our experiments did not yield information on angle of approach.

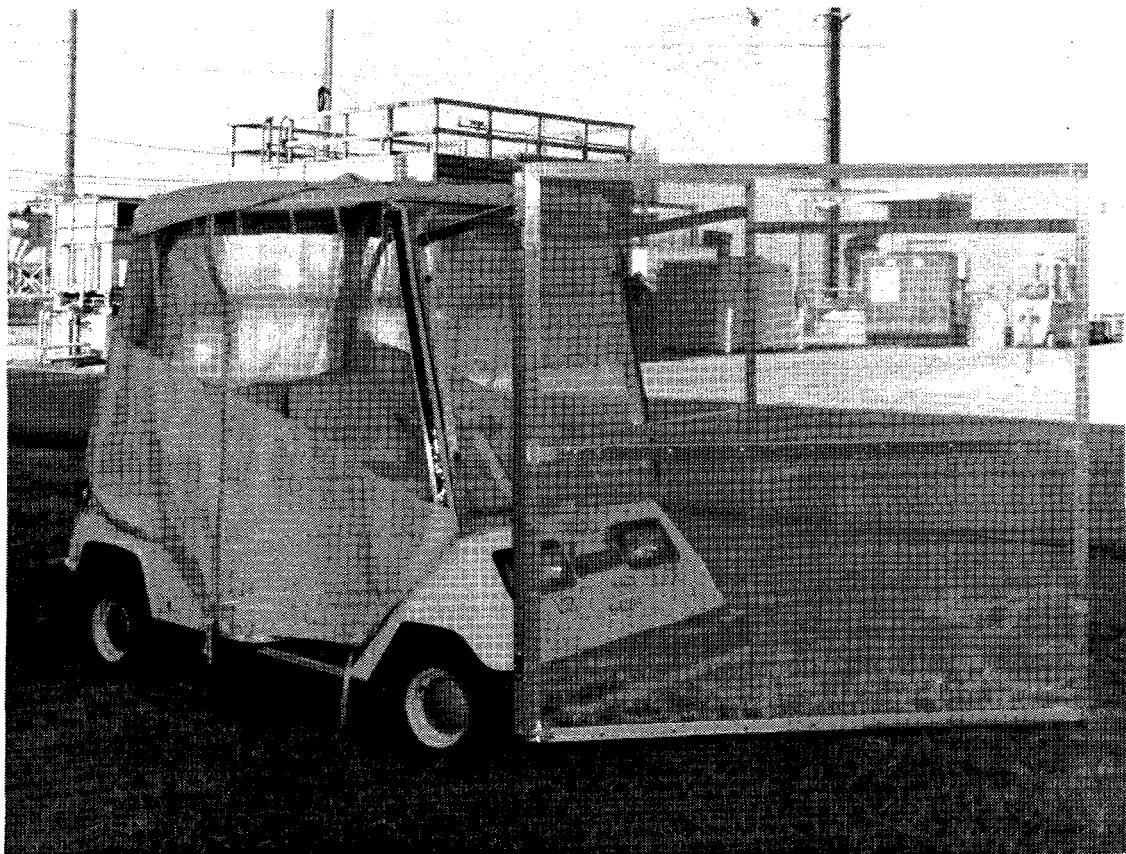


Fig. 5.8. Reflecting screen at 45° angle.

VAN DATA: COLLISION COURSE

The event shown in Fig. 5.10 is the signature of a large van traveling directly toward the transceiver. This signal was digitized at 1000 Hz, so the ordinate corresponds to time, in milliseconds, for the event. The slowing down is evident around time 3500 ms.

A number of other configurations were used to gather a variety of data under varying geometries.

MODIFIED CART DATA: COLLISION AND NONCOLLISION COURSES

A small, two-person electric cart was modified to carry a large planar radar reflector in an attempt to present a uniform radar cross section to the beam. Runs were made with the cart on a collision course (cart towards the detector) and with the cart

passing by the detector. Both sets of runs had the reflector configured in two orientations: normal to the direction of travel and at 45° to the travel direction. The goal in the latter case was to observe the cosine effect of the Doppler signal. This can arise when the return signal is directly reflected from an object moving at an angle with respect to its travel. Such a situation could produce a speed determination that is incorrect by the cosine of the angle (i.e., as much as 0.707).

For the two collision course plots in Fig. 5.11, the markers were well in front of the receiver, and most action took place just before the cart was brought to a halt directly in front of the receiver. The first plot (Fig. 5.11) shows the case in which the radar reflector ("Screen") is at 0°—directly facing the receiver, which was pointed at the cart along the path of motion. Slowing down is evident between points 5000 and 6000 in



Fig. 5.9. Screen supporting structure.

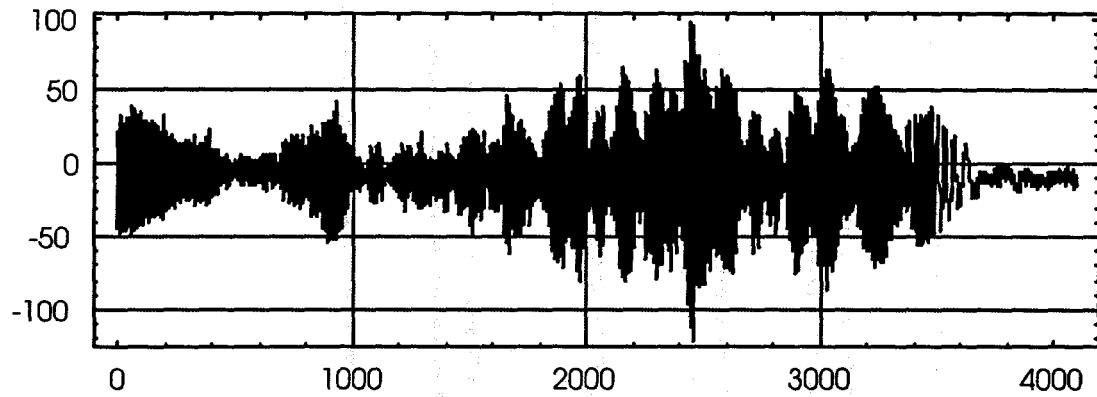


Fig. 5.10. Doppler signature of van.

both cases (the Doppler frequency decreases rapidly). The major quantitative difference in the top two plots of Fig. 5.11 (the second one with the reflector at 45° to the cart motion) is

the intensity. In the second case, decidedly less energy—some 20 times less just after the second marker—is being reflected into the receiver; however, enough is getting in to

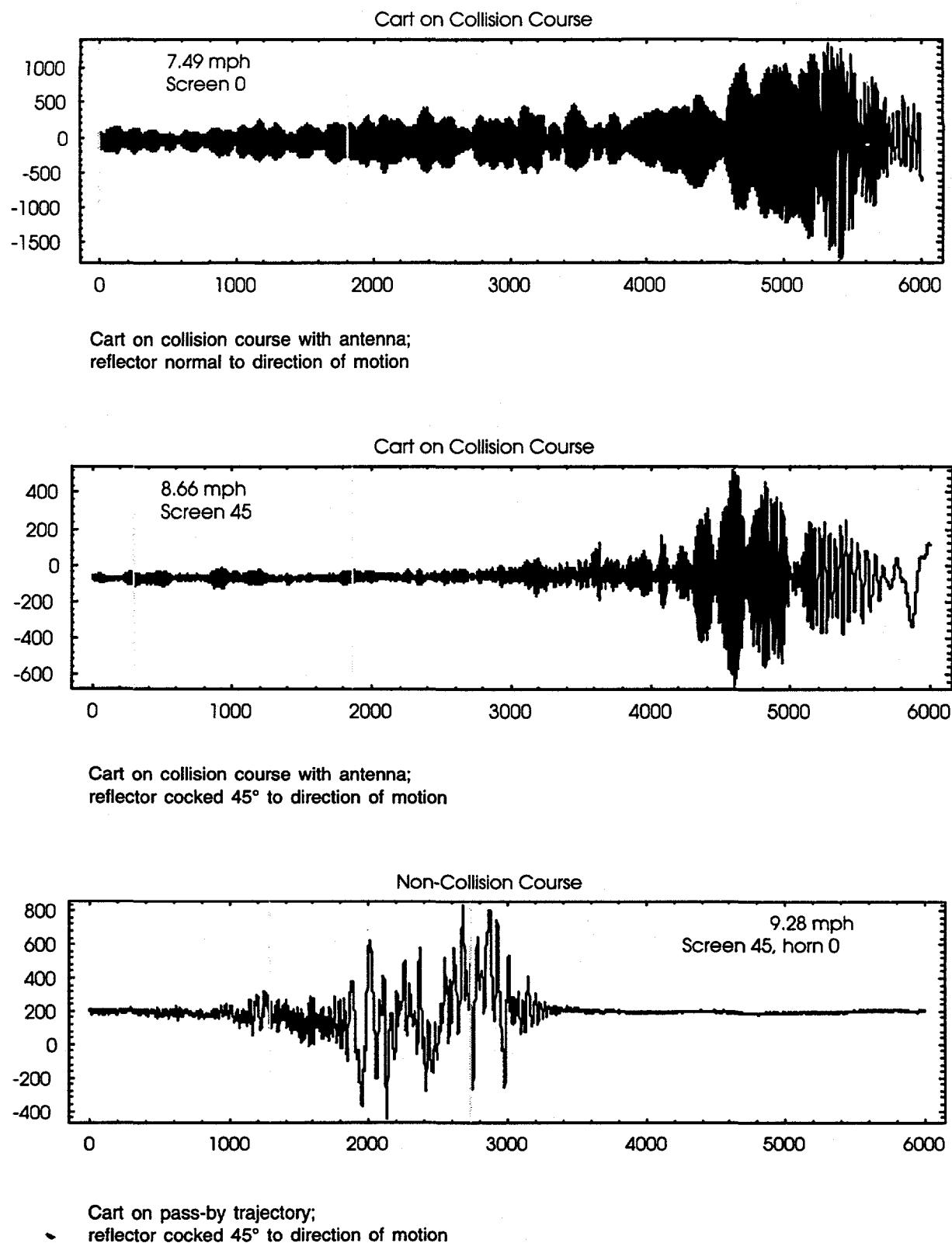


Fig. 5.11. Doppler signatures of cart with screen.

detect a Doppler signal consistent with the linear motion of the cart. The qualitative difference is that a substantial and measurable signal is reaching the receiver throughout the first event, whereas the second event shows a reasonable amount of signal only after about point 3500. Careful analysis of the last region of slowing down in the two cases indicates the presence of the cosine effect. However, the conclusion is that it may safely be ignored, since there are enough portions of any real vehicle that will specularly reflect energy directly into the receiver and thus give a measure of the velocity (i.e., the cosine of the effective angle is nearly 1.00 to within the solid angle of acceptance of the receiver horn).

CAR DATA: NONCOLLISION COURSE

A passenger vehicle was used as the target in the following tests. All tests were done with the vehicle passing by the stationary detector. In one case, the detector was directed normally to the vehicle motion; in the others it was either pointing ahead by -45° (looking at the receding vehicle) or behind by $+45^\circ$ (looking at the approaching vehicle).

The markers shown in both plots in Fig. 5.12 were spaced 3 m apart, with the receiver horn placed in between; thus, the point of closest approach of the vehicle is approximately in the center of the two vertical markers.

Fig. 5.12 shows the Doppler signal for the receiver horn pointed normal to the vehicle trajectory. The front tires triggered the first marker at about point 1800 and triggered the second marker at about point 2700. The shape and quality of the signal suggests that most of the information present indicates the varying radar reflectivities of the various portions of the automobile as it passes the receiver. This reflectivity signal is somewhat modulated by certain corners and curved portions that are reflecting energy directly into the receiver. This is therefore a mixed low-frequency amplitude signal with some Doppler components. Such a signal seems to be characteristic of a "drive-by" event having no significant component of velocity directed toward the receiver. The task of the analysis system must be to assign a "no problem" label to such events.

This type of event is probably the most frequent event to occur and is typical of passing multiple target vehicles on the left in many familiar traffic situations. Such an event is easily recognized by its lack of structure and by the fact that it is dominated by low-frequency features.

Fig. 5.13 shows the Doppler signal with the receiver horn viewing the receding vehicle. The signal does not grow appreciably until the vehicle's front tires are well past the first marker (the vertical line at about point 1400).

Most of the signal, in duration and intensity, is present after the vehicle is past

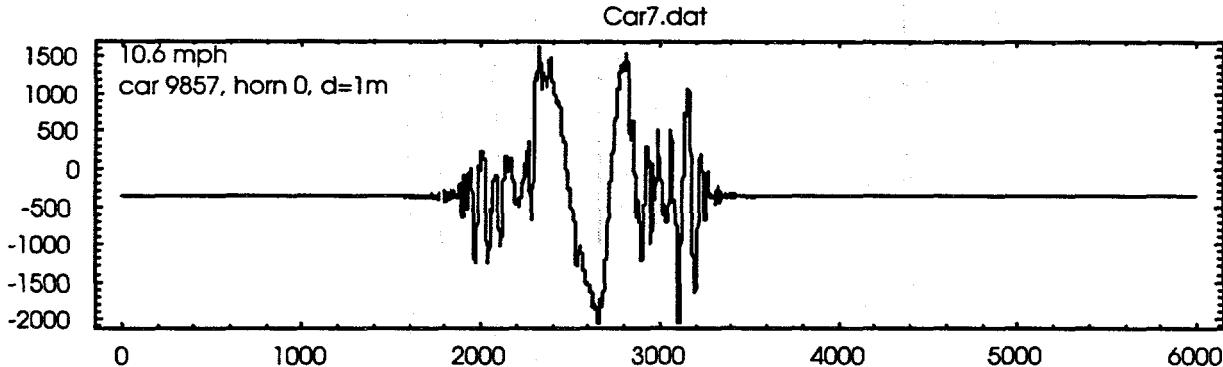


Fig. 5.12. Passenger car drive-by Doppler signature.

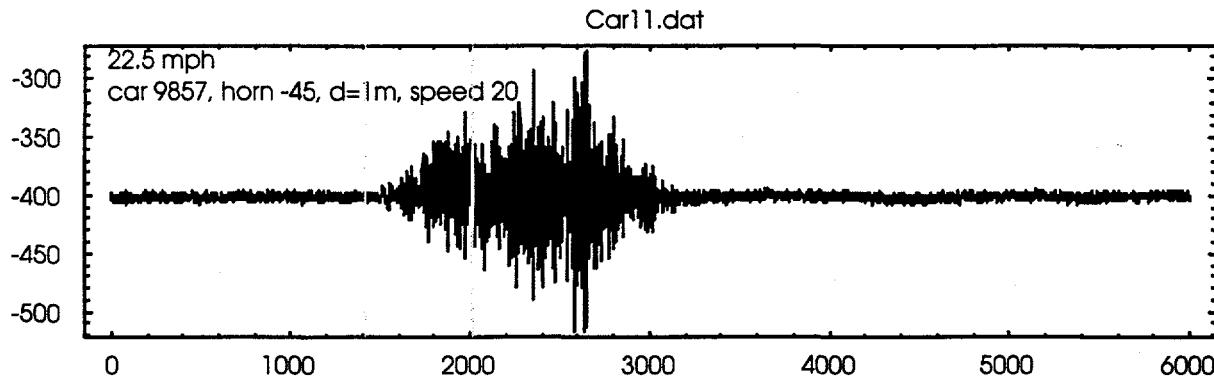


Fig. 5.13. Receding passenger car Doppler signature.

the receiver. There is no evidence of slowing down, as the oscillations are fairly uniform in frequency during the entire event. The event-analysis system must also characterize such events as "no problem." This type of event can easily arise at intersections where roads are at less than 90° angles.

The interpretation of these data depends on the relationship between velocity and Doppler frequency. The frequency of an electromagnetic wave when reflected off an object moving with respect to the transmitter depends strongly on the angle of incidence (for velocities much less than light the angle of reflection is nearly equal to the angle of incidence for a plane reflecting surface). When the angle is small, we can ignore its effect, and we obtain the expression for the ratio of the received frequency, ω_r , and the transmitted frequency, ω_s :

$$\frac{\omega_r}{\omega_s} = \frac{c - v}{c + v},$$

where c is the speed of light and v is the speed of the reflector (positive when moving toward the source). For ordinary vehicular

traffic $|v| \ll c$, we expand the fraction on the right and neglect terms in v/c to the second or higher power. The result is

$$\omega_r = \omega_s (1 - 2v/c).$$

Rearrangement results in the expression

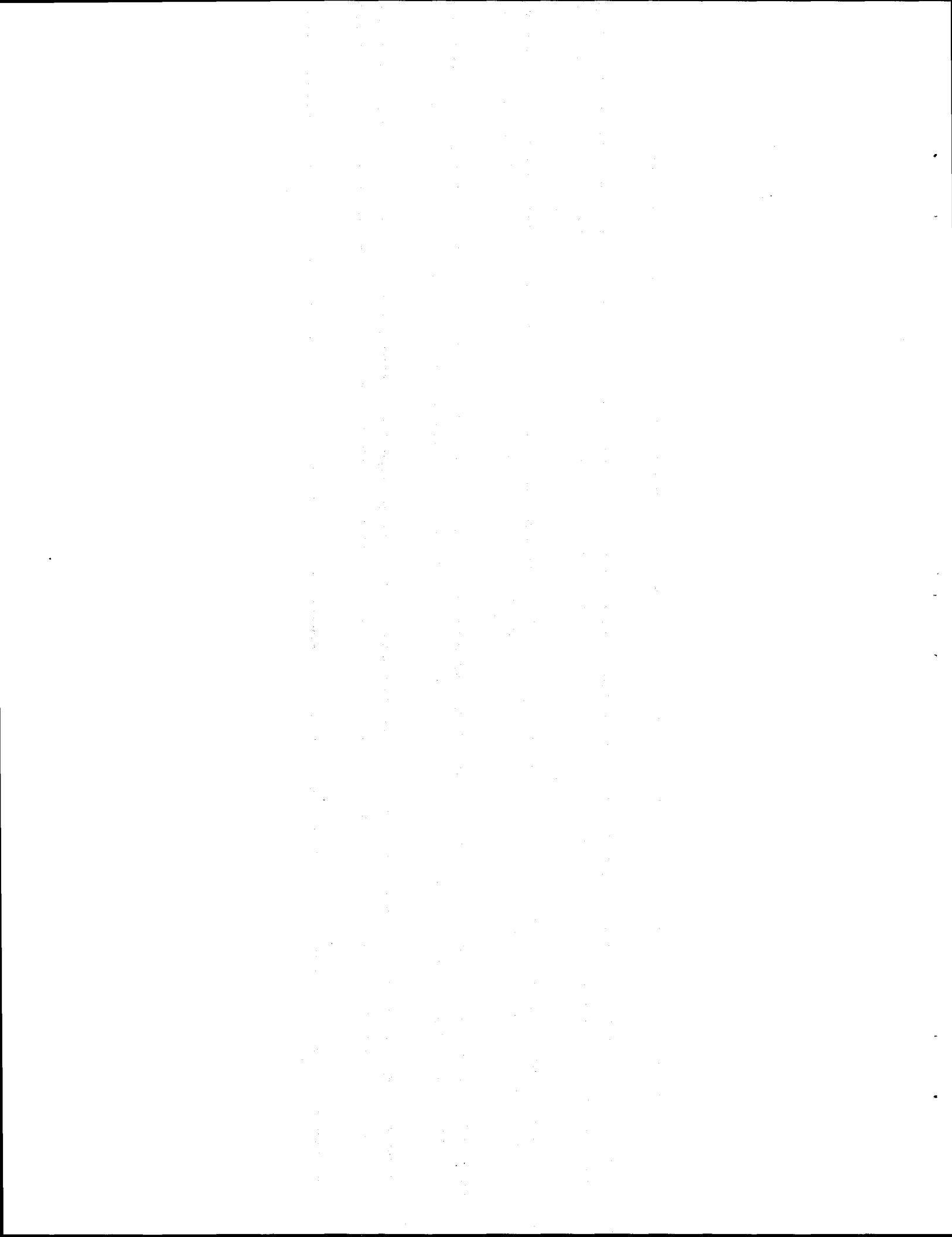
$$v = \frac{c}{2} \frac{\Delta\omega}{\omega},$$

where $\Delta\omega/\omega$ is the fractional Doppler shift due to the moving reflector.

When c is measured in miles per hour (mph) we obtain the expression

$$v = \frac{\Delta\omega}{\omega} \times 3.353 \times 10^8 \text{ mph}$$

for the speed of the moving reflector. Since ω_s is 10 GHz, and the measured shifts range from 0 Hz up to about 330 Hz, the range of speeds measured during the various experiments ranged from 0 mph up to about 11 mph.



6. MATHEMATICAL ANALYSIS

At the conclusion of the experiment, extensive analysis was performed to determine whether or not the signatures obtained contain useful information.

Simulations included analog filtering, extraction of "instantaneous" velocity values by STFT, extraction of velocity and acceleration by discrete wavelet analysis, correlation of the signal with chirped wavelets, parameter extraction by nonlinear curve fitting, extraction of "state-space" parameters by the Karhunen-Loeve transform, singular-value decomposition, direct auditory analysis, and a conceptual look at beam forming.

DATA ANALYSIS: GENERAL COMMENTS

Fourier Analysis

The simplest way to look at the Doppler radar data is to partition the sample stream into small chunks (windowing), apply the fast Fourier transform (FFT) to each chunk, and identify the dominant peak, if any. This works nicely if there is enough signal strength in the given window; otherwise, this technique yields garbage. A problem with this method is extracting acceleration. The Doppler frequency is changing appreciably over the width of a window (typically ~100 ms). By taking the difference in frequencies (peak position in the Fourier spectrum), one obtains a measure of the acceleration. However, the result is an average over several hundred milliseconds rather than an instantaneous value, and it

takes a long time to carry out the measurement.

Discrete Wavelet Analysis

The advantage of using wavelets is purely computational. The signal, again in a window of ~100 ms, is analyzed by a bank of filters. The bank having the highest energy (above a threshold) is identified. The input to that bank is transformed to the frequency domain, and the position of largest peak (as above) determines the Doppler shift. Since the filter bank with the greatest energy is usually not the first bank, the Fourier transform is carried out on a data set that is smaller by some power of 2 than the original. This means faster analysis, but with the linear overhead of the filter-bank calculations. A desirable by-product of the wavelet approach is better noise immunity, leading to cleaner signals for shorter windows.

"VAN SLOWING DOWN"

The data file "Van Slowing Down" shows an event in which the Doppler signal was recorded with a sample rate of 1000 per second while the van accelerated from about 6 mph to 12 mph, and then slowed down to about 2 mph over about 2 s. The velocity behavior is quite clear even when the raw data are plotted, as Fig. 6.1 shows.

Fig. 6.1 shows the Doppler signal over about the final 0.8 s of the event. Note that the higher speeds produce about 10 samples over a single cycle. At 12 mph, there are three samples per cycle; we are getting close

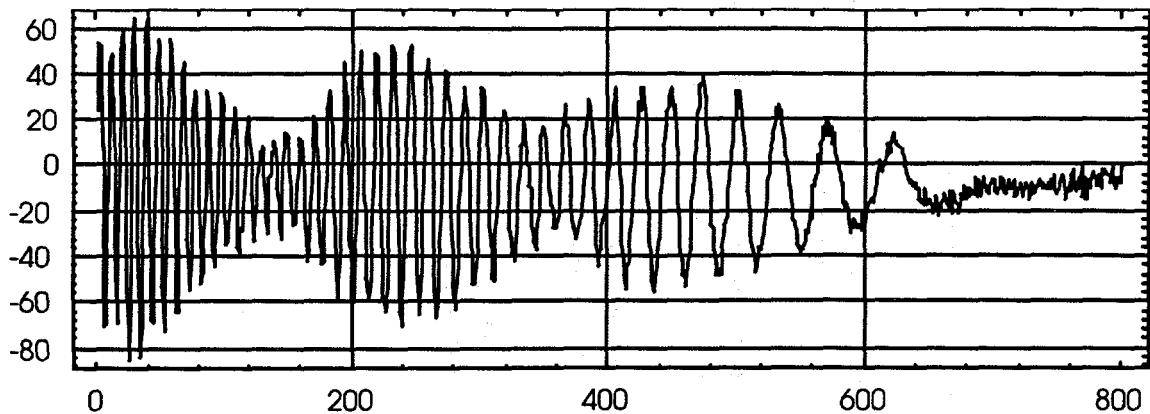


Fig. 6.1. Data plots for van slowing down.

to an aliasing problem. At a 36-mph speed differential between the target and the detector, a 1000-Hz sample rate would be at the Nyquist rate, which is not allowed. If we took the sample frequency up to 1500 Hz, we could handle a 50-mph differential with some margin of error. This would increase our data volume by 50% and hence our time to achieve an answer. It would also push out the "danger," or blind, zone by 50%, to perhaps 10 m (see the discussion below). Such a limitation might be unacceptable.

DATA ANALYSIS: SOME RESULTS

The purpose of the various methods discussed below is to demonstrate that there are many different ways to extract the required information from the Doppler signal. Some of these methods are quite simple—they require very little computational effort—but they return only the barest amount

of information. Others are quite sophisticated; accordingly, they require more computation but yield an abundance of information.

Simple Analog Filter Model

Suppose we wish to create the simplest possible system for analyzing the Doppler signal. If we simply filter the analog signal after setting a threshold on the energy content, we obtain the picture shown in Fig. 6.2, where the analog filter has been simulated with an AR(1) digital filter.

The simulation involved amplifying and clipping the Doppler signal to obtain a square-wave output. This also provided the thresholding, since only a signal with an amplitude of 1.00 was allowed to proceed to the filter.

The several large features in Fig. 6.2 represent regions where the energy in the Doppler signal was sufficient to trigger the

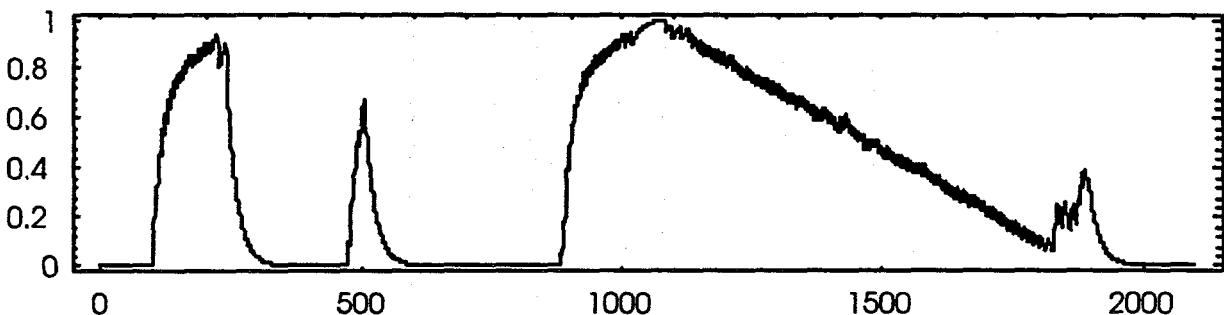


Fig. 6.2. Analog filtered output.

next stage of capacitance filtering. The behavior of the filter in such a case is a rapid charging to its full value (near 1.0). If the Doppler signal showed constant velocity, the response of the filter would be a constant value, whose level depends on the absolute velocity or frequency of the Doppler signal. As the target accelerates, the level rises. As the target slows, the level drops as shown. The instantaneous velocity is proportional to the voltage level, and the acceleration is given by the slope. For example, at about time step 880 (corresponding to about 1.76 s into the event), the signal was large enough to pass the threshold test, so the capacitor started charging. Some 400 ms later, about three time constants of the simulated RC network, the output level is seen to level off. At this point, the level drops linearly through time 3.6 s, at which point the signal disappears into a high-frequency noise. Since there is no simulated low-pass filter, this noise is amplified and shows up as the peak at time step 1890.

Extraction of the acceleration by the same approach would require a high-pass filter to smooth the velocity filter so that a differentiator of discrete components could be made.

The virtue of this method is its simplicity and low cost; it is the least expensive method for determining velocity and acceleration from the Doppler signal. On the other hand, it has no flexibility, and its effectiveness depends entirely on the initial design and execution. There is also a lag of several time constants before valid inferences can be made (at least 0.3 s in this case); depending on the constraints imposed by vehicle motion, this lag may prove fatal for the analog method.

FFT Extraction of Velocity

The next method studied was standard Fourier analysis based on the fast Fourier transform (FFT). This has the advantages of being quite well understood from both theoretical and implementational standpoints, and being reasonably effective at producing quality results.

As an example, suppose we take 128-ms, non-overlapping samples and perform a FFT on each of them over the entire event. The results are shown below in Fig. 6.3.

The position of the sharp peak in each of these plots is a measure of the frequency of the Doppler signal and hence of the average speed during that 128-ms period. We could now write a peak-determining algorithm to extract the position. However, the positions were determined by inspection, and the results of the entire event, analyzed by FFT, are shown in Fig. 6.4.

Discrete Wavelet Analysis

Multirate filter, or discrete wavelet, analysis has certain advantages over the FFT method discussed above, the most important one being computational efficiency: wavelet filters are generally faster than the FFT, yet provide as good or better results for the purpose at hand. As a first-pass analysis, the Doppler signal was passed through five levels of a discrete wavelet transform using the Daubechies extremal-phase wavelet with eight coefficients, otherwise known as D4.

The data stream was segmented into 128-ms samples overlapping 50%. The goal was to produce a velocity (frequency) measurement every 64 ms. Each segment was analyzed into six wavelet levels. The level with the most energy (determined by the sum of the squares of the value) was presumed to have the most capability of providing an accurate velocity determination. Again, this produces an average of the velocity over the 128-ms interval. The method gave results almost identical to those shown above and was about 10% faster.

The plots in Fig. 6.5 show the first three levels of the wavelet transform of 128 samples taken at time step 2400 (2.4 s into the event). It is evident that most of the signal is in the second level (there is a bit of high-frequency noise in the first, and practically none in the third).

An FFT of the resulting 64 samples in level 2, shown in Fig. 6.6, gives a quick determination of the frequency as 281 Hz,

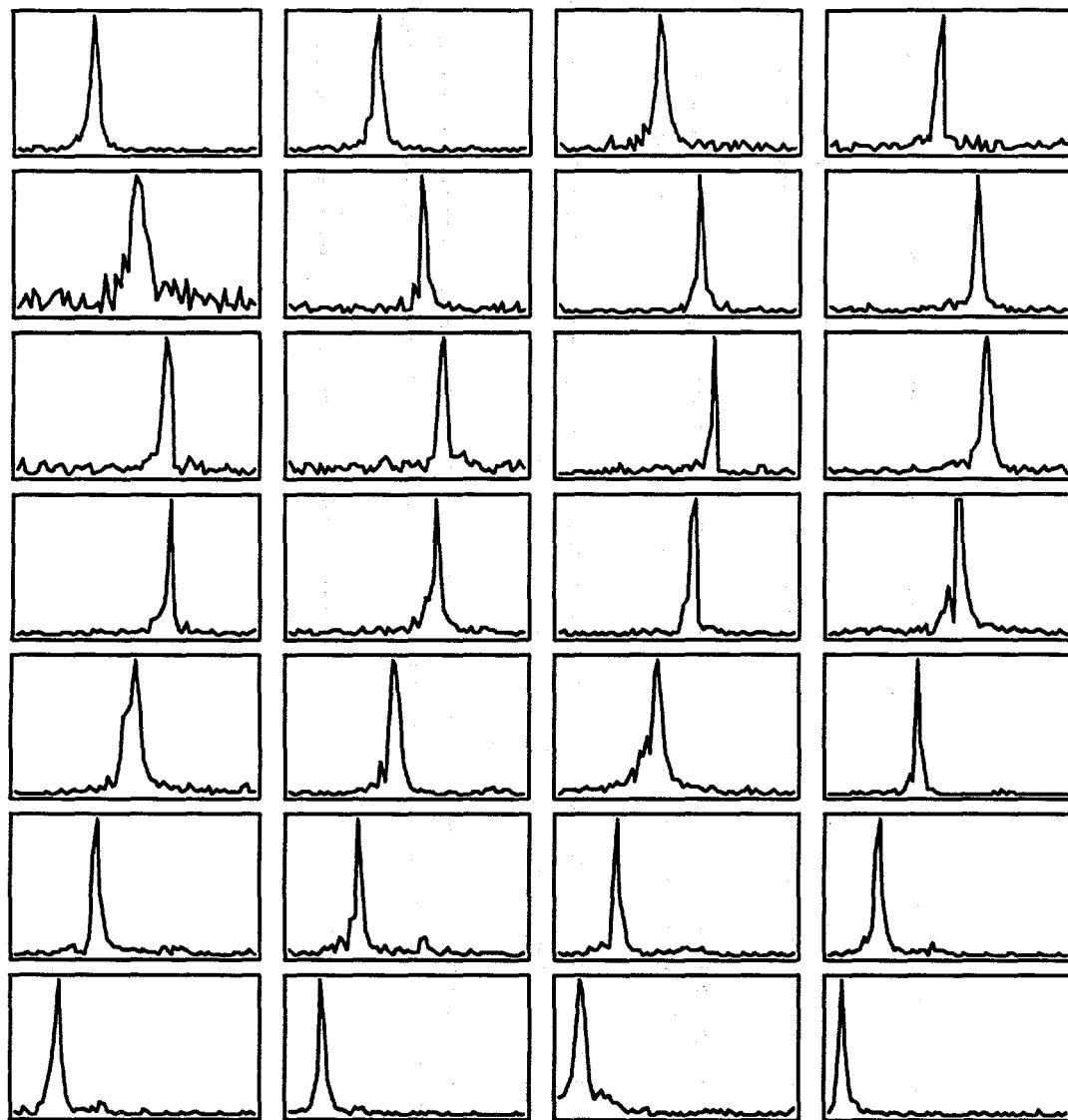


Fig. 6.3. Fourier analysis of van slowing down.

resulting in a speed of 7.0 mph. This value may be as much as 10% in error due to the averaging effect and taking the peak value of the curve rather than the mean position.

The basic idea looks reasonable, but the dyadic scheme may be too coarse. ("Dyadic scheme" refers to the fact in discrete wavelet analysis, the signal is resolved into octave-wide components, as described in the Appendix.) For a more detailed look, it is possible to use the same multirate techniques at other bandwidths, such as critically decimated third-octave analysis.

Chirplet Analysis

Chirplet analysis uses a number of chirp wavelets, each with a different deceleration (linear frequency behavior). The Fourier transform (FT) of such a set convolved with the FT of data segments picks out the appropriate components, identifying regions as to both acceleration and velocity.

Each data section, each perhaps 100 ms in duration, will require an FFT (of the data) and a convolution with each of several mother chirplets (of various linear frequency variations) at each of several scales (corresponding to basic frequency or

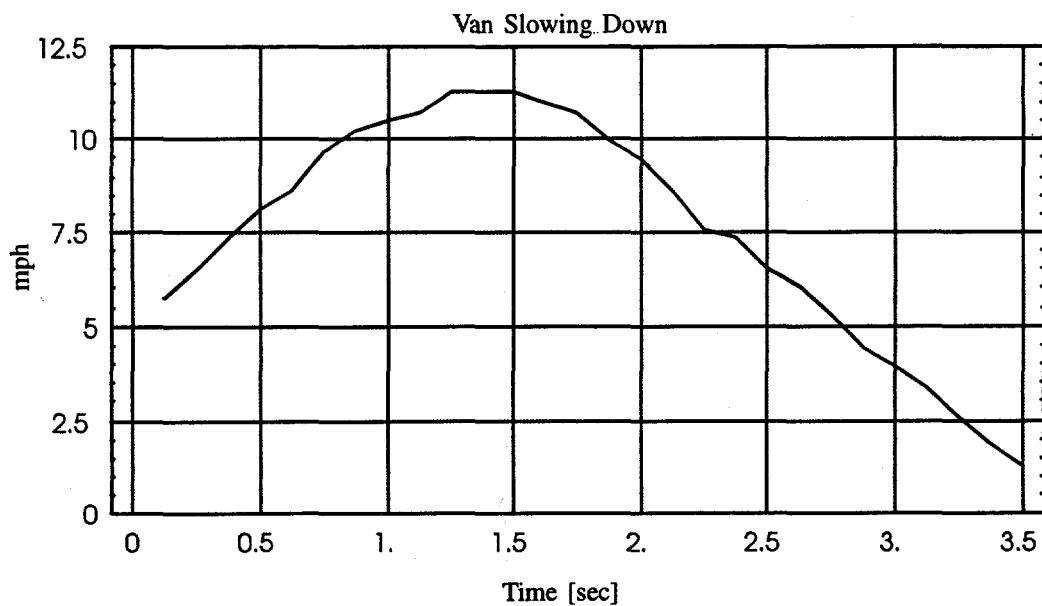


Fig. 6.4. Extraction of velocity from Fourier analysis.

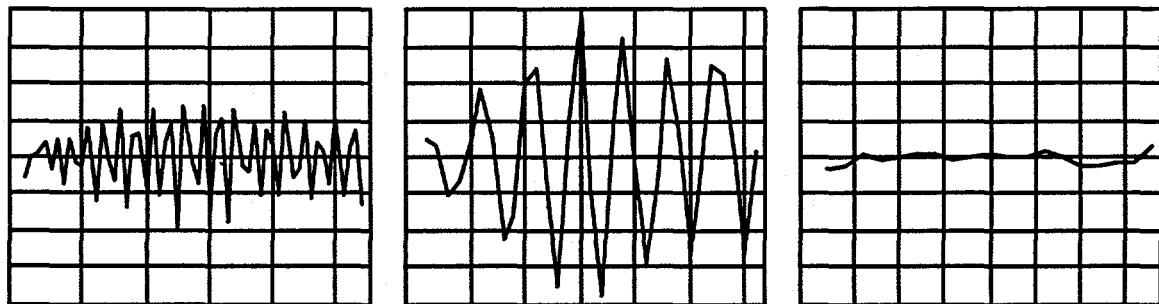


Fig. 6.5. Wavelet analysis of van slowing down.

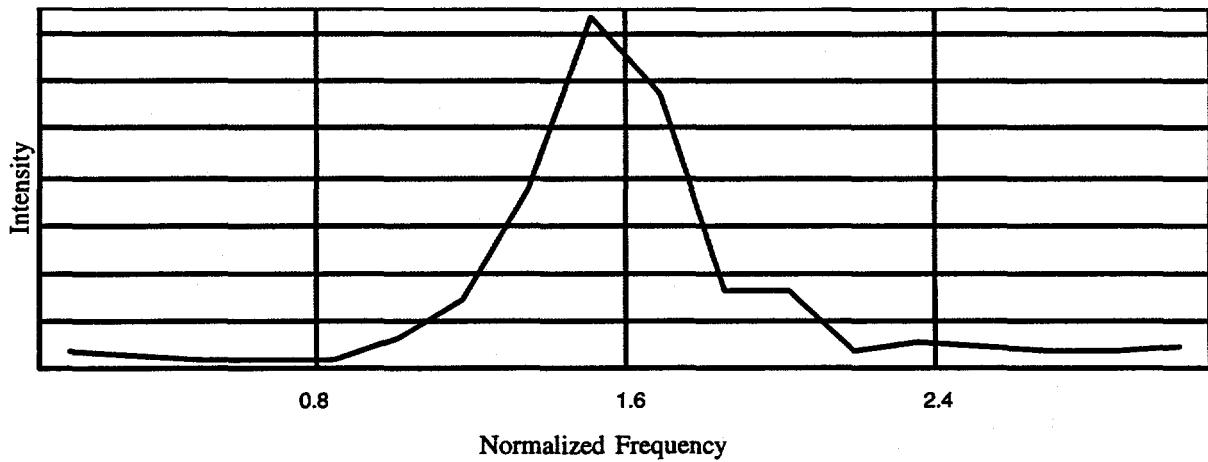


Fig. 6.6. Fourier analysis of wavelet level 2.

velocity). The wavelet with the largest correlation is presumed to be the best description of the data. An alternative is to convolve segmented portions of the Doppler signal with a set of filters, each representing a particular velocity and acceleration. If this set is arranged in an hierarchical manner, processing time can be greatly reduced.

The concept of chirplet analysis is that of a bank of matched wavelet-packet filters chosen to span the Doppler signals anticipated. To illustrate the concept, we constructed a set of chirp functions in the frequency range corresponding to five velocities between 2 and 12 mph, and five accelerations for each velocity ranging from -5 mph/s to $+5 \text{ mph/s}$. Only one generic set is shown in Fig. 6.7 as the actual velocity.

The first example, shown in Fig. 6.8, is from the middle of the event, where velocity is high and is changing only slightly. Only the central filter shows a large response to this data when it is sampled at about 18 points, giving a vehicle speed of around 12 mph. None of the other filters, for this sample spacing, show as high a response, so we conclude that the acceleration is small.

The second example, shown in Fig. 6.9, is near the end of the event, where the velocity is low and the acceleration is quite noticeable. Here, the central filter shows the highest response, for a sampling of about 150 over the length of the filter, corresponding to about 2 mph. The filter on the far right has about ten times the strength of response of the filter on the right for this sampling; this corresponds to an acceleration of -5 mph/s , a value consistent with those obtained by the other methods.

Nonlinear Model of a Single Cycle

The nonlinear curve fit is perhaps the fastest of the digital methods in that we can obtain a value for the velocity and acceleration within the time frame of a single oscillation of the Doppler signal (assuming adequate processing power). The idea is quite simple: the Doppler signal consists of a frequency-modulated sine wave. The obvious amplitude modulation due to the varying configuration of the reflecting surfaces is not important for extracting the motion dynamics.

A typical one-cycle waveform lasts anywhere from 3 ms at about 11 mph to about 33 ms at 1 mph. If we can fit the sampled waveform during a cycle or a half cycle, we can obtain an estimate of both the velocity and acceleration using the model

$$f(t) = a_0 \cos(a_1 + a_2 t + a_3 t^2),$$

where a_0 is the amplitude of the Doppler signal during the particular cycle of interest; a_1 is the phase, allowing for the starting value to be nonzero; a_2 is related to the velocity in the usual Doppler analysis; and a_3 accounts for the acceleration.

For example, the slowing down is quite evident during the cycle shown by the plot in Fig. 6.10.

Here, the dots show the actual sampled waveform (noise reduced by a single Haar wavelet filtering). The smooth curve is a nonlinear least-squares fit to the ten data points using the Levenberg-Marquardt algorithm. The entire event lasts 20 ms (corresponding to about 2 mph). The cost of executing the Levenberg-Marquardt algorithm is about 500 ms in high-level Mathematica.

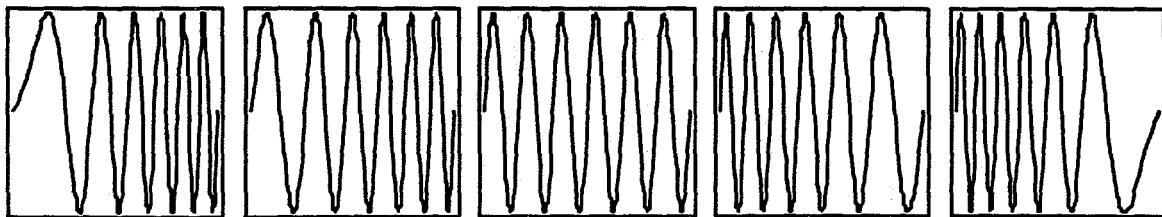


Fig. 6.7. One row of the chirplet sets.

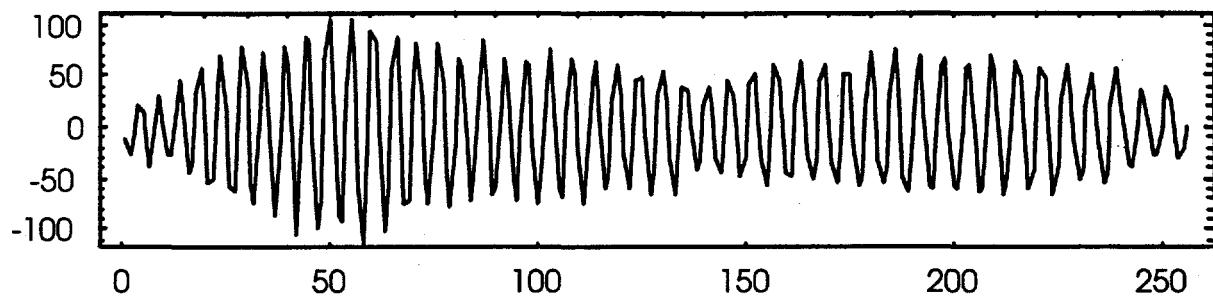


Fig. 6.8. Fixed-speed signal correlation with fixed-speed chirplet.

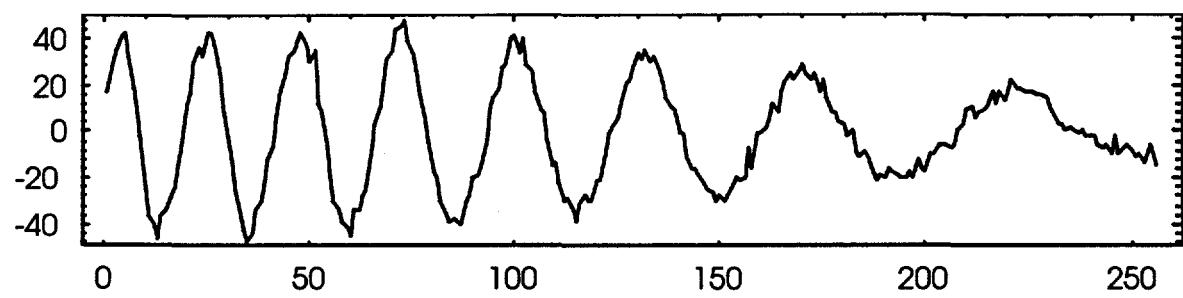


Fig. 6.9. Deceleration signal correlation with deceleration chirplet.

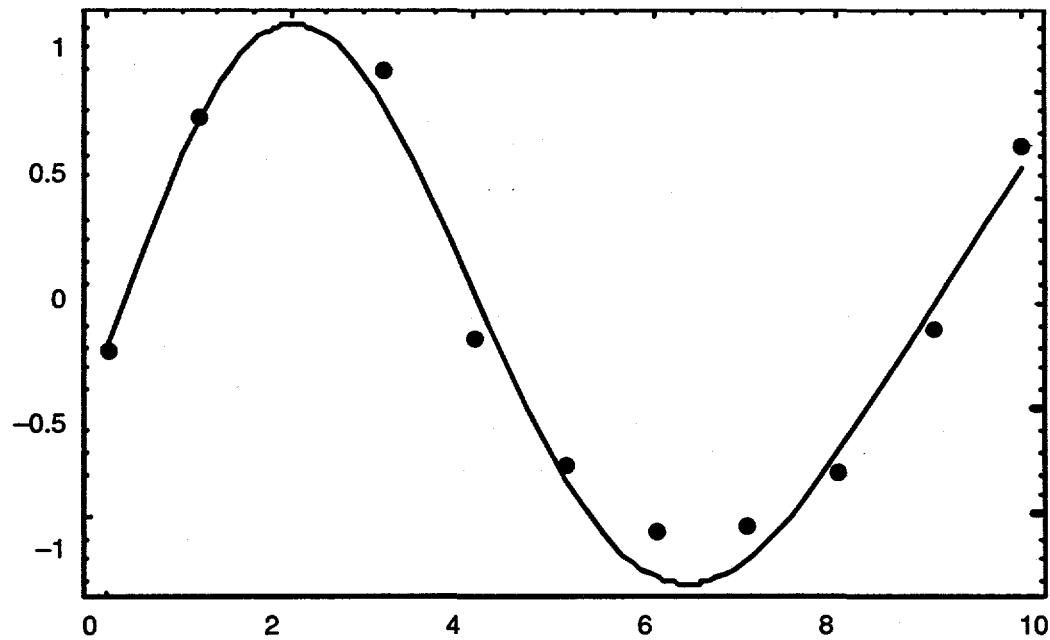


Fig. 6.10. Nonlinear curve fitting.

This can be reduced by a factor of 20 without much effort, for a computation time of 25 ms, comparable to the time of the event itself. A reduction of 50 to 100 times is possible using a higher-speed digital signal processing device; this would allow a determination of velocity and acceleration to be made once per cycle even at the higher vehicle speeds.

The method of obtaining the sections over which to fit the model is simply that of observing zero crossings; each zero crossing of the Doppler signal indicates that the data set is ready to be fit.

A drawback of this method is its susceptibility to noise. Since from 6 to 30 samples (at a sampling rate of 1000 Hz) are used to make the determination, noise in as little as one sample would affect the accuracy of the calculation. For example, consider the final 600 ms of the event. The results for the cycle-by-cycle determination of velocity and acceleration are shown in Fig. 6.11. The variation in the velocity measures are quite reasonable; however, the acceleration parameter shows much more noise (Fig. 6.12), and the mean acceleration during this period is about -1.0 mph/s, although the actual value can be read off the plot as $(2.5 \text{ mph} - 8.3 \text{ mph}) / 0.600 \text{ s}$ or about -9.7 mph/s . Thus, the noise destroys the acceleration readings.

However, these readings can easily be inferred from the difference of the velocity results. For example, a smoothed difference, equivalent to a linear regression on the velocity results, gives a value of -10.4 mph/s for the acceleration, quite in line with the estimate taken from the velocity plot.

An advantage of this method is that it can be used to discriminate against such "nonevents" as a target moving past, not toward, the detector. These events, as shown elsewhere in this report, result in large chi-squared values for the presumed Doppler cycles. The chi-squared values for the true Doppler signals are significantly less than 1 but on the order of several hundred for the non-Doppler events.

Karhunen-Loeve Transform

KLT amounts to using the principal eigenvectors of the covariance matrix as a basis to represent the process. There are (physically) three dimensions to the state space of a Doppler chirp.

Consider the last 750 ms of the slowing-down event for the van. There are three separate regions identifiable as to amplitude. A likely reason for the amplitude variation is that different portions of the van, having different reflectivities, are being viewed by the receiver as the van approaches. These regions are quite evident in the plot in Fig. 6.13.

If we now form a time-delay matrix consisting of

$$A = \begin{bmatrix} a_1 & a_2 & \dots & a_m \\ a_2 & a_3 & \dots & a_{m+1} \\ \dots & \dots & \dots & \dots \end{bmatrix},$$

we create a square, positive definite matrix of dimensions m by m from A multiplying it by its transpose. The eigenvalues are found, and the first three are used to reconstruct the signal in phase space. The optimum value for m is determined by taking the ratio of the first two eigenvalues; when they are approximately equal, we have spread out the signal to a maximum extent. (We assume that the dominant process is a simple sine oscillation.)

The result of analyzing the above signal in this fashion is shown in Fig. 6.14. First we show the eigenvalues for $m = 9$. This value of m spreads out the fundamental oscillation equally along the first two directions.

If we use the first three eigenvectors as a basis for the phase space, we obtain the plot shown in Fig. 6.15 using all 350 data points. Note that there are three disks, each tilted with respect to each other. The three almost-planar trajectories roughly correspond to the three regions in the data set.

We can now look more closely at the event in the same phase space by considering only 50 events at a time, resulting in six regions over the 350-point data set. These

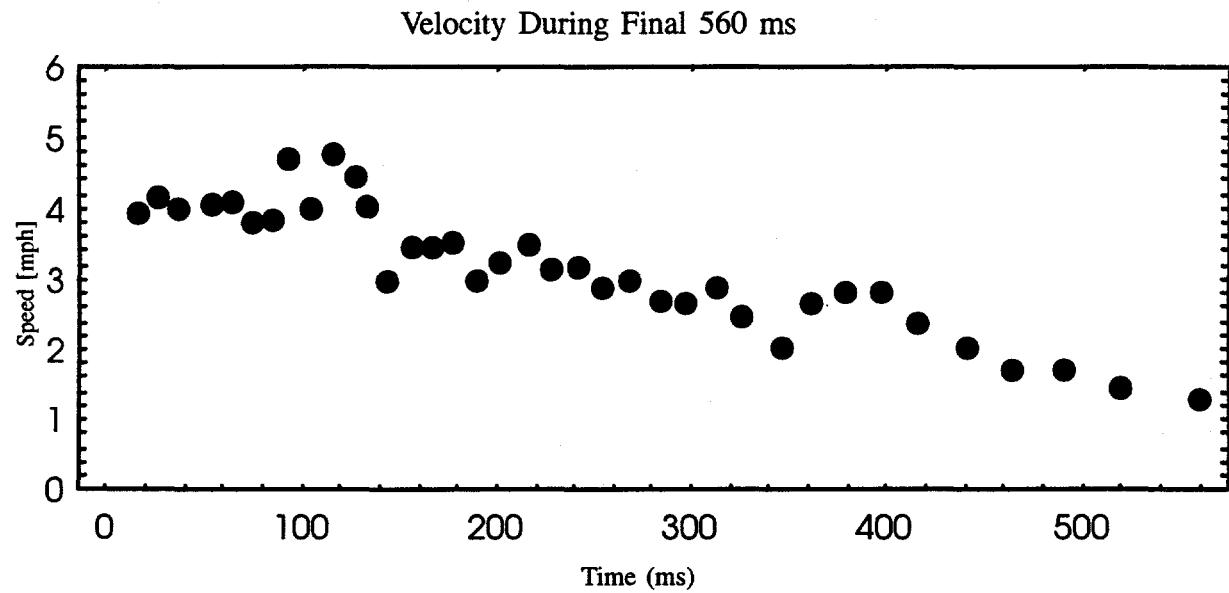


Fig. 6.11. Velocity obtained by nonlinear curve fitting.

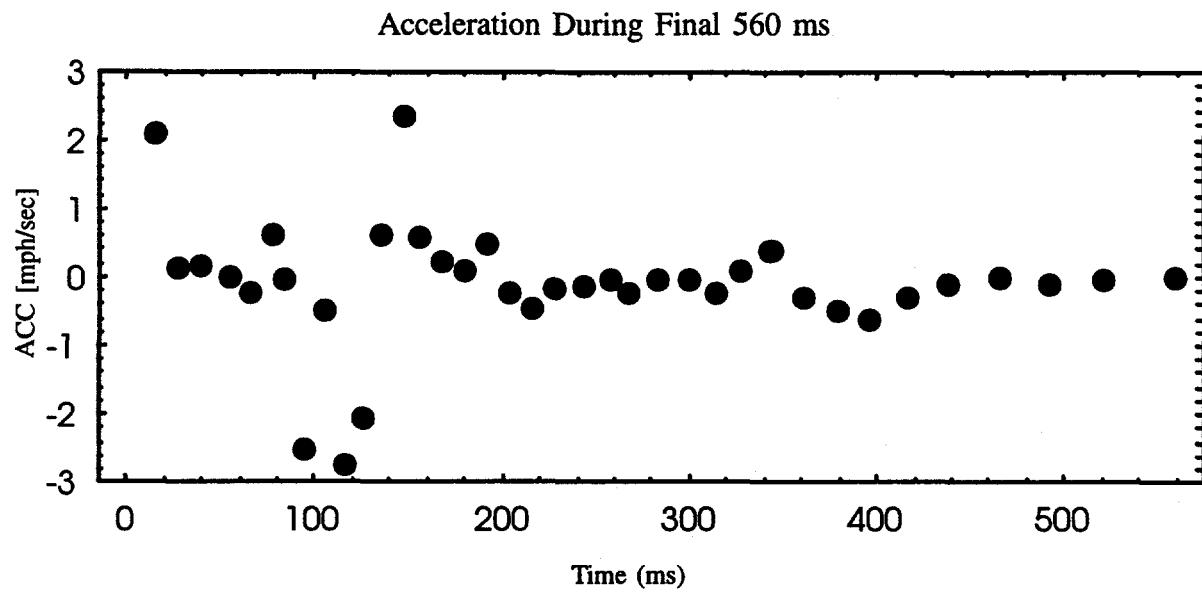


Fig. 6.12. Acceleration obtained by nonlinear curve fitting.

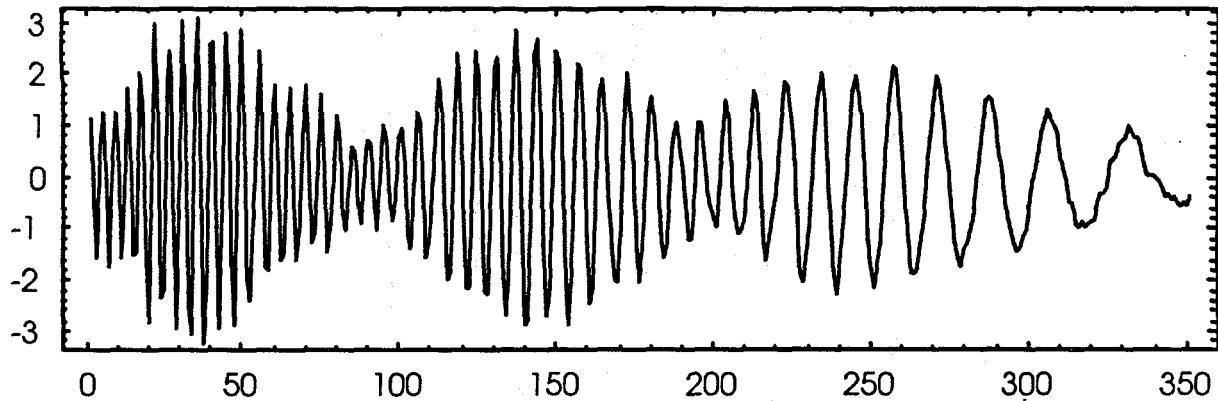


Fig. 6.13. Signature of possible multiple reflections.

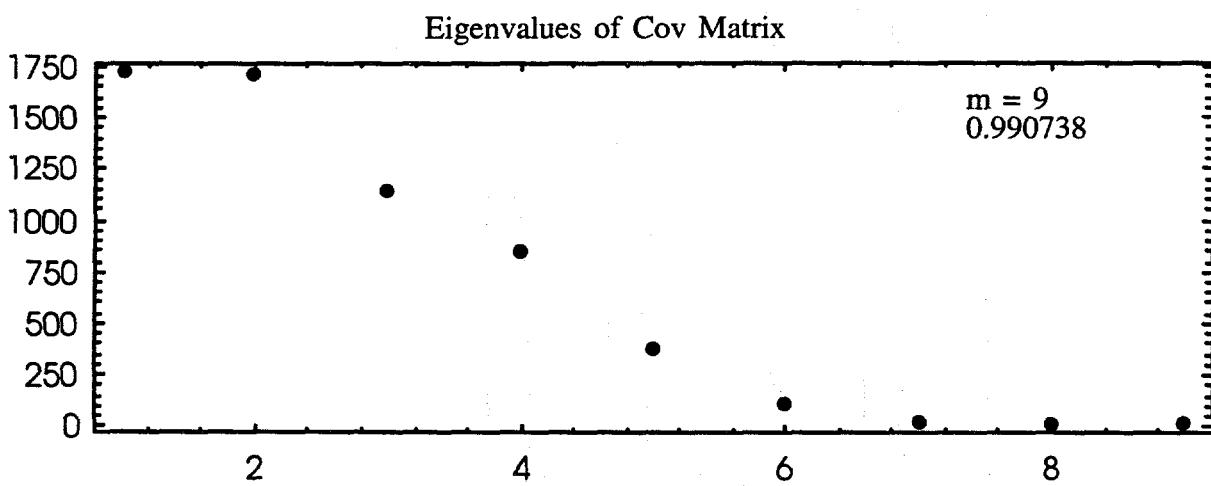


Fig. 6.14. Eigenvalues of covariance matrix.

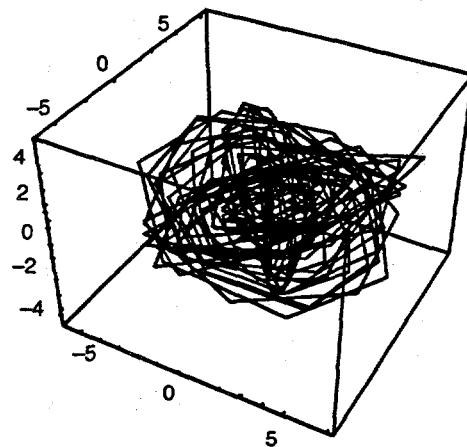


Fig. 6.15. Projection onto the first three eigenvectors.

plots are shown in Fig. 6.16. The first 50-point region in shown in the upper left, and the last in the lower right. Note that all but the first are nearly two-dimensional trajectories. The first one clearly shows the presence of acceleration during the entire 50 time steps by the helical structure. The last trajectory has a decidedly different orientation than the previous four.

Singular Value Decomposition

The idea behind using a singular value decomposition (SVD) for the covariance matrix of the delayed data values is the same as that for KLT, discussed above. The purpose is to achieve a phase-space embedding of a geometrical figure: the trajectory. This is similar to the state-space embedding approach used in analysis of chaotic systems. Our system is definitely not chaotic, as the ultimate attractor is a point in phase space, not a pleasing geometrical attractor. However, much the same methods can be used and provide insight as to the details of the Doppler data.

With this method, we again form the covariance matrix, as shown above. The difference is that we look for singular values. The question of how many dimensions to take for the phase space is more cleanly resolved using SVD. With KLT, we saw that there was not a large difference between the third and fourth eigenvalues. A similar picture

for the SVD case is shown in Fig. 6.17, where we see that the third value is relatively higher and the fourth relatively lower. This may be a slight advantage for representing the event in three rather than four dimensions.

The embeddings look similar to those obtained using the eigen decomposition. The advantage to this method is that there are fast algorithms for the SVD; the drawback is that the methods to obtain the eigenvalues and vectors are fairly computationally intensive.

The plot in Fig. 6.18 shows the three-dimensional character of the last 200 ms of the van's deceleration. All three dimensions are clearly needed to display the trajectory, showing that the acceleration is quite important. The next plot (Fig. 6.19) shows the event at a much earlier stage. Here, the trajectory looks more like a pancake, showing little acceleration relative to the dominant behavior of the velocity component.

The SVD scheme depends on the fact that the signature breaks up into a relatively few distinct components that occupy distinct regions in some mathematical space. To use the method in an automated system, it would be necessary to identify which parts occupy which spaces.

This is the same problem that pattern recognition schemes must overcome. Each point in the data stream can be treated as a sample to be classified. We have *a priori*

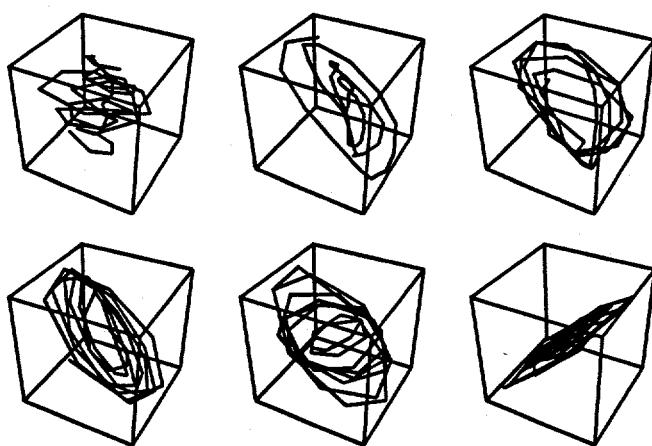


Fig. 6.16. Resolution of projection into distinct surfaces.

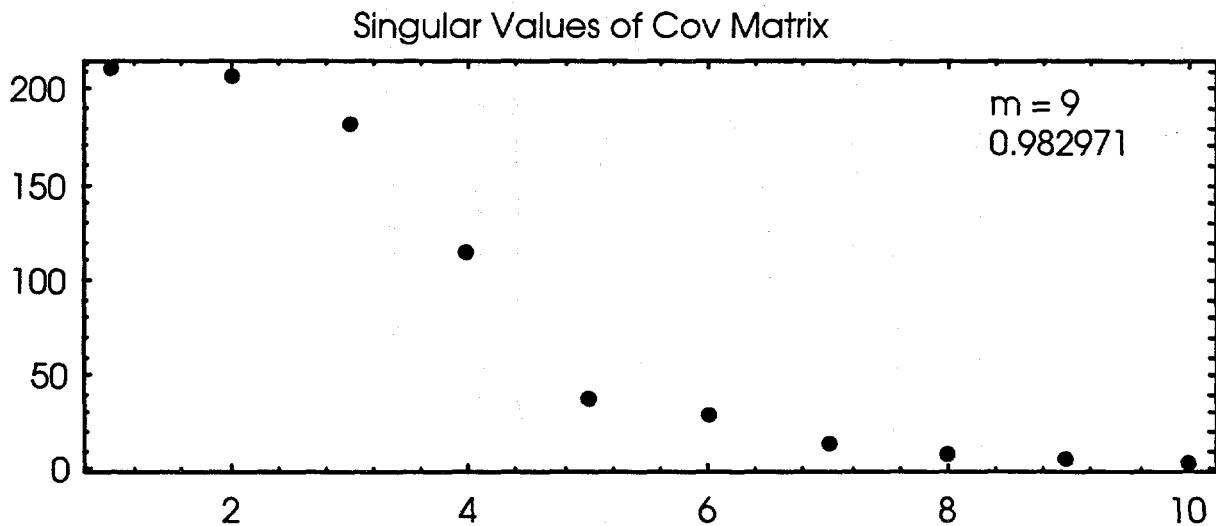


Fig. 6.17. Singular values of the covariance matrix.

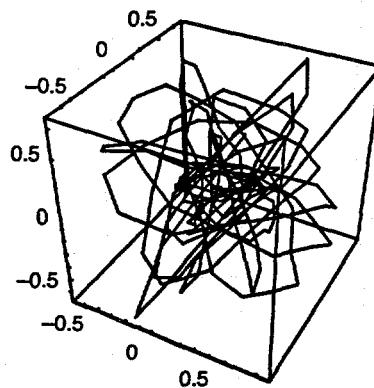


Fig. 6.18. Projection of acceleration signature onto SVD basis.

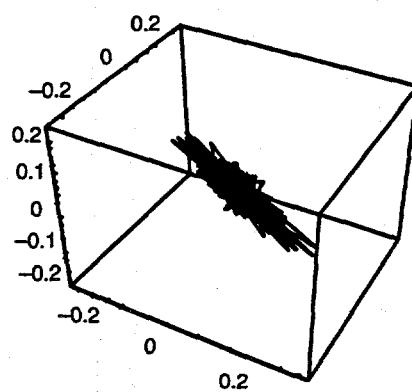


Fig. 6.19. Projection of constant speed signature onto SVD basis.

knowledge that the samples should form several distinct clusters. They should be classifiable with a cluster-seeking algorithm. Once the elements of each cluster are known, it is a simple matter to compute geometric properties of the clusters, and also a simple matter to deduce target dynamics from the cluster's geometric properties.

The problem with cluster-seeking algorithms is that they are typically much too slow to operate in real time. It may be possible to implement a cluster seeker with a massively parallel array of simple processors, similar to what might be used to implement a cellular automaton.

Auditory Analysis

Another simple approach is to simply amplify the Doppler signal and transduce it into sound pressure waves via a loudspeaker. The resulting tone is quite audible and distinctive with no additional processing. Such an auditory signal could serve as a warning to the driver to be aware of an impending collision. We will explore this avenue no further, other than to mention its potential use and a need for a human-factors study should it prove interesting.

Discussion: Danger Zones and Safety Zones

Earlier in this chapter, the concept of a "blind" zone was introduced to indicate that a certain rapidity both of measuring a response and of making the correct determination could be a problem. There is another limit or horizon to the system, best referred to as a "safety" zone.

A 15-mph differential is 22 ft/s, or 6.7 m/s. In 100 ms, the relative distance between vehicle and target will change by 0.67 m. At a 60-mph differential, this distance will be 2.7 m, or about one car length. Given the time a sensor would require to make the decision—perhaps another 100 ms and 30 ms of deployment actuation—the 15-mph differential amounts to a danger zone distance of 1.5 m. We would like to reduce this

danger zone in the future. At a 60-mph differential, this danger zone extends out to 6 m—that is, anything inside the 6-m zone will not be accounted for in a timely fashion.

A safety zone can be defined by the range of the radar device—i.e., probably a few hundred feet.

Fortunately, it is not necessary to wait 100 ms before computing a new data segment. We may want to compute a new segment every 10 or so samples (i.e., 10 ms), assuming that we have the computational throughput to do the required number of computations in 10 ms. How many computations would be involved? The FFT of 256 points should take <2 ms, and a convolution with a 256-point vector should take much less than 1 ms, perhaps 100 μ s. Thus, 100 convolutions plus the FFT will take on the order of 12 ms. A few hundred microseconds would then be required to pipe the data about, find maxima, and so forth. This ~12-ms information delay would then be piped to the anticipatory control model, where a decision would be made. A very rough guesstimate of the time required to extract the target dynamics is ~5–10 ms, for a total time lag of perhaps 20 ms. Thus, the danger zone would be reduced to about 1 m at 60 mph—not bad!

Of course, these numbers may not hold up to a reality check. The data pipeline has a definite lag to a step function equal to about half its length. A step function is quite unrealistic, though. Any fast object entering the radar's receptive field will probably do so from the side (a head-on approach allows plenty of systolic activity in the data pipe). Further study is needed in order to obtain firmer numbers.

In dealing with the desired range, it is important to keep the intended use in mind. NHTSA's Office of Crashworthiness Research is primarily interested in two possibilities. If the target is very close (perhaps 1–2 m), an anticipatory crash detector might be used to provide arming, or perhaps even noncontact firing. If a target is distant (10–20 m) and the dynamics indicate an impending crash, the

detector could be used for arming. The middle of the range is of much less interest.

Beam-Forming Approach

Beam forming is a multireceiver variation of the time-domain pulse method that builds up a picture in distance-angle-velocity space. This method has several advantages over the simple Doppler and time-of-flight methods. None of the methods discussed thus far can provide an indication of bearing—i.e., the angle of the target with respect to the base vehicle. Thus, we are able to get velocity and acceleration from the Doppler radar, and if we time the pulses, we also get the critical distance parameter. But with a single transceiver, there is no sure way to identify the bearing of a potential troublemaker.

The advantage of knowing the bearings of a collection of targets is that we may safely ignore any target whose angle changes from one pulse to the next: the only possible collision candidates are those targets whose bearing remains constant over several measurements. This ability to prefilter the space of all possible targets will greatly aid the next processing stages and reduce the computational burden at all later stages. Note that if we are moving, all stationary objects are removed from consideration by the constant-bearing criterion. Considering only the constant-bearing targets, we next ask which of those are moving toward our vehicle at a speed greater than a predetermined (damage-capable) amount. These are the only targets that the anticipatory subsystem need consider in its predictive model.

We will give priority to those targets whose bearing is stationary $\pm\alpha$ degrees, whose velocity towards the side of the vehicle is greater than a certain value, and whose distance is less than a certain value.

Radar pulse method

The problem with beam forming with electromagnetic pulses (e.g., millimeter waves) is the high signal velocity.

Electromagnetic radiation travels at light speeds (approximately 300 mm/ns); thus, a 1-m resolution in distance entails a 3.3-ns resolution in time. Translated into the velocity-bearing space, this limits the possible angular resolution. The problems with achieving this goal must be analyzed before a decision to employ accurate timing methods are made.

Ultrasonic pulse method

For sound, we have more than enough time resolution because the speed of sound is about a million times slower than that of light. Thus, we can process pulse times on a millisecond time scale instead of having to consider the nanosecond time scale. However, the range of the ultrasonic pulse is much less than that of the radar pulse. The solution may be to have two systems—ultrasound for close range and radar for more distant objects.

Threat identification by beam forming

One beam-forming scenario would work as follows. Consider a pulse radar with one transmitter and three receivers at a known geometric spacing. The trigger lines are arranged such that all three receivers receive the trigger pulse at approximately the same instant; thus, all three receivers "know" when the transmitted pulse started its flight. A comparison of the times of arrival of the signal at each of the three receivers allows a straightforward computation of the bearing to the target. One of the inherent advantages of a pulsed type of system is that the differential in times of arrival can be determined at a much higher resolution than the resolution of the absolute time of flight. It is reasonable to expect that the bearing to the target can be found with high accuracy and repeatability.

It then becomes easy to distinguish between threatening and nonthreatening targets in the environment. If on successive transmitting pulses the bearing of the target remains fixed, the target is either moving directly away from or directly toward the victim; the change in displacement indicates

direction toward or away from the victim. Any target whose bearing varies as a function of time is not a threat. Only threats need to be analyzed further.

Thus, the first useful piece of information that comes out of the process is the angle of approach. The displacement follows immediately. From successive pulses, the velocity can be obtained by numerical differentiation. Acceleration can be obtained by a second differentiation. All four necessary pieces of information about the target dynamics thus become available.

There might be some concern about noise corrupting an estimate based on differentiation. A system such as the Amerigon pulsed radar operates by taking a reading as an average of many (perhaps a thousand) pulses, thereby averaging out much of the noise. It is reasonable to expect the output of such a unit to have sufficiently low noise that useful estimates of acceleration can be computed.

DATA ANALYSIS: COMPARISON OF TECHNIQUES

A comparison of the nine different analysis techniques is shown in Table 6.1. The first eight methods presume the use of a single sensor. The eighth item is not really an analysis method. Rather, it reflects the fact that when the Doppler modulation is played over a speaker, the sound is quite distinct for different types of signatures, and it might have possibilities as a driver-warning system. The ninth item, beam forming, might use one or more of the first seven analysis methods in extracting the target dynamics.

The speed of the chirplet is extremely high if the filters are implemented on an application-specific integrated circuit and if massive parallel processing is used.

Table 6.1 lists relative degrees of accuracy. To make numerical predictions about reasonable expectations of error would require the massive data collection of a detailed engineering study. The proof-of-principle experiments of this study simply are not sufficient to make this kind of prediction.

What dynamical data are extractable depends more on the sensor than on the analytical technique. A broadband pulse radar using a single receiver and any of the seven techniques shown in the table should give displacement, velocity, and acceleration, but not angle of approach. A Doppler radar would give only velocity and acceleration. Using a pulse or FMCW radar with several receivers, chirplets, and beam-forming algorithms should produce all four relevant quantities. This should be explored in more detail in the next phase of this research.

TECHNIQUES NOT EXPLORED IN THIS STUDY

Time did not allow for the exploration of two other promising analysis methods, fuzzy logic and cellular automata. Both are capable of fast implementation on dedicated parallel hardware, and both provide robust results in the face of real-world disruptions. Both are worthy of future exploration.

Remarkably, fuzzy logic not really "fuzzy." It is a set of rules whose validity is justified by an appeal to traditional crisp logic. What the rules provide are mathematical measures of the possibility that a given element is a member of any one of a number of sets. Starting from that foundation, fuzzy-logic techniques provide the solution to problems intractable to conventional techniques.

Fuzzy logic has at least two applications to the anticipation engine. First, it is a pattern recognizer, and it might be used to classify the features of radar signatures into "impending crash" or "no crash" classes. It has already been shown to be effective in distinguishing between classes of airplanes from the features of their acoustic signatures (Dress and Kercel 1994), and might be equally effective in classifying the acoustic signatures of ground vehicles.

However, the more crucial application is in the area of decision making. A fuzzy-logic decision engine is eminently well suited to the task of taking data in the form of high-dimensional vectors and producing a

qualitative evaluation. It is quite reasonable to expect a fuzzy-logic system to take lists of numbers from the environment and system models of an anticipation engine and convert them into a prediction like "little bump," "big bump," "moderate crash," "bad crash," or "really bad crash."

A cellular automaton divides a multidimensional space into adjoining cells, assigns a numerical value to each cell, and provides a set of rules for updating the values

of each cell depending on its present state and the present state of its nearest neighbors. When implemented on simple dedicated parallel hardware, a cellular automaton can provide astonishingly fast solutions to differential equations with complex boundary and initial conditions. A cellular automaton is a strong candidate for the algorithm to be used at the heart of the environment model in the anticipation engine.

Table 6.1. Comparison of analytical methods

Computational				Extractable parameters			
Method	Speed	Simplicity	Accuracy (rank) ^a	Angle	Displacement	Velocity	Acceleration
Analog	0.1–0.5 s	Simple	1	No	Yes	Yes	Yes
FFT	10–30 ms	Well-known	2	No	Yes	Yes	Yes
Wavelet	5–20 ms	Simple	3	No	Yes	Yes	Yes
Chirplet	0.1–1.0 ms	Simple	4	No	Yes	Yes	Yes
Curve fit	20–40 ms	Complex	2	No	Yes	Yes	Yes
KLT	50–100 ms	Complex	2	No	Yes	Yes	Yes
SVD	25–100 ms	Complex	2	No	Yes	Yes	Yes
Audio				Audible warning only			
Beam forming	<i>b</i>	<i>c</i>	<i>b</i>	Yes	Yes	Yes	Yes

^aAccuracy rank is a relative ascending scale, with 1 representing the lowest rank and 4 the highest in this particular rating. Therefore, the analog method is not very good. FFT, curve fit, KLT, and SVD all have about the same degree of accuracy, although KLT and SVD may reveal more information about multiple surfaces. The wavelet and chirplet methods have finer resolution than the others.

^bThe speed and accuracy of beam forming depends on the analytical techniques used. If we assume dedicated chirplet processing on each receiver plus a comparable amount of time to combine the results, the beam former should make its prediction in approximately 2 ms.

^cThe complexity of a beam former depends on the analytical techniques used. If dedicated chirplet chips are used, all the processing for three receivers to extract all four quantities should be less complex than KLT extraction of three quantities from a single receiver using pulsed radar.

7. PROSPECTS FOR THE FUTURE

In an optimistic vision of the future development of this system, we might expect the following scenario. NHTSA supports the collection of enough experimental data to allow for realistic specification of the sensor element needed for the environment model and then supports development of the anticipation engine (particularly of the software), with the task being completed within the next 5 years. (Private industry is unlikely to do this because the technique is so obscure.) Because this information is in the public domain, there is virtually no incremental cost for commercial developers to copy it for use in their systems.

In the meantime, on-vehicle sensors become so inexpensive that they become standard equipment on virtually all new cars. Consequently, the data stream describing the state of the vehicle is there for the taking. The data can be made available to the anticipatory system by the addition of a single multi-cable jack somewhere in the onboard data system. The incremental cost might be \$2 per vehicle.

In the meantime, at least one manufacturer markets a practical microwave radar priced at \$10 per unit when purchased in lots of

100,000 or more. Assume that beam forming is required and that two such radars are required. Assume pessimistically that this system requires dedicated radar. The incremental cost is \$20 per vehicle.

Assume that the major algorithms can be implemented on application-specific integrated circuits (ASICs) that can be produced for \$5 each in large quantities. Assume conservatively that one ASIC is needed to process the environment data, one is needed to process the on-vehicle data, and one is needed to implement a set of cellular automata to implement the anticipation engine. Also assume that the cost of the ORNL fuzzy logic chip (used to actually make the prediction) can be reduced to \$2.

The total incremental hardware cost is \$39 per vehicle, exclusive of packaging, overheads, profit margins, etc. Assume that all these can be met at 100% markup. This suggests that in 5 or 10 years, the system could be available at a price of a little less than \$80, with the prospect of the price falling in future years as the hardware technology advances.

The possible costs per vehicle are summarized in Table 7.1.

Table 7.1. Best-case costs for a future precrash restraint sensor system

Component	Estimated cost per vehicle
On-vehicle sensor with multi-cable jack	\$ 2
Microwave radar, 2 units	20
Application-specific integrated circuit, 3 units	15
Fuzzy logic chip	<u>2</u>
Total incremental hardware cost	\$39
Total with packaging, overhead, and profit margins (100% markup)	\$80

8. CONCLUSIONS

An anticipatory-precrash restraint sensor is feasible but not yet developed. A system based on formal anticipatory principles should significantly outperform conventional technologies. Development can be expected to be a long-term effort with the prospect of high payoff in prevention of death and injury. That such a system is feasible, however, does not guarantee its development. The level of technical and financial risk is sufficiently high that American industry is unlikely to develop the technology on its own initiative or with its own money. Therefore, the development of such a technology is probably dependent upon R&D support from NHTSA.

A major priority that remains to be defined by NHTSA is whether the ultimate objective of this research is to use an anticipatory system to arm a restraint device or to deploy a restraint device. There is a risk-reward tradeoff to be considered in this decision. The arming function is less expensive and less error-prone and the consequences of error are less dire than the firing function, but the arming function has a lower potential payoff.

The existing literature makes clear that two issues have been resolved. First, the formal anticipation engine is realizable. However, the only past development has been for large-scale nuclear-power-plant control systems. Such a system has never been attempted for a small instrument. None of the previous research in predictive crash detection has used the formal anticipation paradigm. Second, sensors and processors are available to provide a good, fast, and inexpensive

description of the present dynamical state of the vehicle to the embedded system model in the anticipation engine.

There is a fair consensus in the literature that a prediction of the severity of the impending crash is needed. It is generally believed that severity depends on the energy transferred in the collision, and energy is dominated by relative velocity rather than by mass or vehicle classification. The literature suggests that it is possible, though not assured, that targets can be classified on the basis of acoustic signatures or RF radar returns. This is a much more difficult problem than extracting target dynamics, and it is not certain that the benefits will justify the effort.

What is not discussed in the literature is whether or not sensors and processors are available to provide a good, fast, and inexpensive description of the present dynamical state of the *targets* to the embedded environment model in the anticipation engine. Unavailability of target dynamical information would make an anticipatory system infeasible. The object of the experimental and analytical part of this study was to determine whether or not this enabling technology is available, and if so, what form it might take.

This study found that inexpensive radar in a "real-world" setting does return useful data on target velocity and acceleration. The principle has been proven with a \$20 Doppler radar. The velocity and acceleration of the target can easily be extracted from the Doppler radar signal by any of several

methods that are implementable in real time. In addition, Doppler radar shows distinct signatures for distinct approach geometries.

The data produced by a radar system can be converted to target dynamical information by good, fast, and inexpensive signal-processing techniques. While the Karhunen-Loeve transform (KLT) is computationally too costly for use in a small real-time instrument, it is the best available algorithm for revealing the information buried in a signal. The KLT of the Doppler radar time series appears to simultaneously resolve velocity and acceleration of multiple surfaces. It demonstrates that, in principle, all this information is extractable. What remains is the engineering task of finding a computationally cheaper algorithm, such as the wavelet transform, that is almost as good as KLT and can be implemented on inexpensive real-time hardware.

While Doppler radar proves the principle that dynamical information is buried in the returns of cheap radar, it does not provide

displacement information. Displacement information is necessary to feed the environment model of an anticipation engine. FMCW or pulsed broadband radar both appear to be workable alternatives. Each is able to provide displacement, velocity, and acceleration data for multiple targets. The makers of both claim that their technology can be mass-produced for a few tens of dollars per copy.

Pulsed broadband radar may or may not be the panacea that its makers claim it to be. For example, it might be vulnerable to receiving jitter. In addition, since it involves taking multiple numerical derivatives, it could be extremely vulnerable to noise. Both the capabilities and the pitfalls of this method need to be investigated. There is a reasonable possibility that either broadband pulse or FMCW, its less heralded competitor, is capable of providing the target dynamics needed for the environment model of the anticipation engine.

9. THE NEXT PHASE: A TASK LIST

The following task list does not constitute a statement of work for a follow-on phase. Instead, it is a menu of choices of reasonable lines of investigation, along with the expected costs of each. The idea is that these possibilities, and NHTSA's research priorities and budget limitations, would form the basis for developing the scope of the statement of work for the next phase.

Task 1: General Data Collection

The initial phase makes clear the distinction between an "anticipatory crash predictor" and a "radar sensor." The latter is only one of several key components of the former, and probably not the trickiest. Nevertheless, the sparse set of samples of radar-return data and other signatures presently available allow only for claims of demonstration of principle. Before serious engineering specification of the anticipatory crash predictor can occur, it is necessary to know what dynamical information can be repeatably extracted from the available sensor types. This is likely to be the biggest effort in the front-end development of the anticipatory crash predictor.

The kind and quality of experimental data sought in this task would be such that engineering specifications for practical crash warning systems could be deduced from them. In addition, the data should be captured and reported in such a manner that practically any reasonably capable hardware developer should be able to use them.

The problem is that a massive amount of data must be collected. There will be between

four and six vehicle types. There are perhaps six different trajectories of interest. Several different speeds, perhaps four, should be investigated. For each of these 144 possibilities, at least 100 signatures need to be acquired. Each of these 14,400 signatures would be of perhaps 3 s duration and sampled at 40,000 samples per second. Each sample is a 2-byte word, leading to 240 kilobytes of data per signature. Assume that signatures for four out of five possible device types (Doppler radar, pulse radar, FMCW radar, ultrasonic sonar, and microphone) are collected simultaneously for each pass, using parallel data acquisition channels. This results in a little less than 14-gigabytes just for the signatures.

Add in vehicle type, device type, trajectory identifiers, and data from contact switches for a independent indication of the dynamics for each signature, and the data set rapidly approaches 15 gigabytes. It is worth noting that such an annotated data set, collected under controlled conditions, and stored on a CD-ROM, would be a valuable consideration in its own right. In addition to serving as a foundation for the subsequent tasks of this project, it could be used by other experimenters in crash detection.

The type and ranges of data to be collected depend on the functionality of the ultimate system. If it is an arming system and not a firing system, then vehicle mass and class signatures are not needed. Also, the firing system would emphasize dynamical data at relatively short ranges collected at short durations, while an arming system

might look at a longer range and longer collecting times.

Since angle-of-approach information is required, some kind of beam forming will ultimately be needed. At least some of the data should be collected with a single-transmitter, multiple-receiver configuration.

There are four active sensing technologies for which signatures might be collected—ultrasonic sonar, Doppler radar, FMCW radar, and pulse radar. Doppler radar and ultrasonic are on hand. FMCW radar and technical support for its use can be acquired through a cooperative arrangement with Hittite Microwave, a company already working with NHTSA. Pulse radar and technical support might be acquired through a cooperative arrangement with Amerigon if Amerigon is interested. Millitec already has a cooperative agreement with NHTSA's Office of Crash Avoidance Research and may be easy to include in the next phase of this project.

It is not recommended that either FMCW or pulse radar be procured as a simple purchase of an end item for this project. Such a purchase typically does not include detailed engineering support or disclosure of the internal workings of the device. A cooperative agreement in which the supplier furnishes the temporary use of one or more specimens of the device, engineering support, and disclosure of internal operational details, as needed, is more in keeping with the goals of this project. ORNL can enter into nondisclosure agreements if necessary to protect the suppliers' trade secrets.

It is reasonable to assume 1 person-month of direct labor and \$5000 of travel expenses for the activity involved in setting up each of the cooperative agreements.

For acoustic signature collection, a 20-kHz scientific microphone and processing electronics are required.

For data acquisition from the active devices, one channel of A/D conversion is needed for each device. The sampling rate should be at least 40 kHz. The resolution should be at least 12 bits.

To provide independent verification of the dynamics of each signature, a precision velocity profile recorder should be mounted on the vehicle. Since this is a specialty item for on-vehicle research, it is assumed that the sponsor will arrange for the loan of the device for the duration of the research rather than provide funding for the researchers to purchase the device.

Data recording would be performed by an on-hand personal computer.

For data acquisition, there is a fixed cost independent of the amount of data acquired. Hardware to perform the data collection would cost \$10,000. This amount does not include items already on hand. It also does not include payments by NHTSA to radar suppliers for access to pulse and/or FMCW radar. Setting up and taking down the data-collecting apparatus would require 2 person-months of direct labor.

There is a variable cost in proportion to the amount of data acquired. An average day's data collection should result in 100 signatures, or a complete data set for one of the 144 possible combinations of trajectory, vehicle type, and speed. The data collection would require a full day's work by two employees; another person-day would be required for support activities (picking up the car from the rental agency, transcription of data to CD-ROM, etc.). In addition, it is assumed that the target would be a rented vehicle, at an average rental of \$75 per day.

Since the data would be reusable, and a valuable asset in its own right, it should be documented separately from the overall project. One person-month is assumed for preparation of the report.

Summary of subtasks for data collection

1. Development of cooperative agreements with radar suppliers

(Comments: Expenses are for one agreement; two may be required. Fees to the radar supplier not included.)

Direct labor	1 person-month
Other expenses	\$5000

2. Data collection, fixed costs		
Direct labor	2 person-months	
Other expenses	\$10,000	
3. Data collection, variable costs		
<i>(Comments: Expenses are for 1 av. day of data collection, resulting in a complete data set for one combination. Six vehicle types, 6 trajectories, and 4 speeds result in 144 possible combinations.)</i>		
Direct labor	3 person-days	
Rental of target vehicle	\$75	
4. Documentation		
Direct labor	1 person-month	

devices simultaneously, would generate 300 signatures.)

 Direct labor 0.25 person-month

Task 3: Beam-Forming Proof of Principle

One datum that is required for crash prediction is the angle of approach. It has not yet been proven that this can be obtained inexpensively in real time. As discussed in the section on beam forming, a pulsed radar using multiple receivers and beam-forming analysis should lead quickly (and potentially inexpensively) to five data for a given target: class (threat or nonthreat), bearing, displacement, speed, and acceleration. A reasonable next step in this research would be to investigate whether or not obtaining this data is practical.

The investigation would be both theoretical and experimental. It would be necessary to simulate the one-transmitter, multiple-receiver configuration for many hypothetical targets mathematically in the time domain. This would help to establish sensitivity to noise and other anomalies in the signal, establish the design requirements for the radar hardware, establish systems definition for the real-time implementation of the beam-forming software, and define what experiments need to be done to prove principle.

The experiment would require fairly significant custom engineering by an enthusiastic radar manufacturer. The development of the radar hardware would also require significant interaction between the researchers performing the investigation and the radar developer. Also, until the simulation is done, detailed specifications and requirements for the radar would not be available.

Since the experiment is for proof of principle rather than for detailed engineering specifications, the data collection would not be as extensive as in some of the other data-collection tasks. It would probably

Task 2: Dynamical Information Extraction

It must not be forgotten that the real objective of this task is to determine how to acquire a description of the dynamics of the target to feed to the environment model of a formal anticipatory system. Incidental to this goal, it is necessary to determine which sensor or suite of sensors most inexpensively produces a usable description. Incidental to this goal is an analysis of the return signatures collected under the data-collection task. This analysis must be much more exhaustive than the proof of principle analysis performed during the feasibility study phase.

The task has both fixed and variable costs. The fixed cost, independent of the number of signatures analyzed, is for setting up the analysis algorithms and reporting the results. This can reasonably be expected to be 2 person-months. The average variable cost is 1 person-week per 100 signatures analyzed.

Summary of subtasks for dynamical data analysis

1. Data analysis, fixed costs		
Direct labor	2 person-months	
2. Data analysis, variable costs		
<i>(Comments: Expenses are for analysis of 100 signatures. One day's data collection, resulting in a complete data set for one combination of vehicle type, trajectory, and speed, and using three different active</i>		

involve only two vehicles and the collection of a few hundred signatures from each one.

One of the objectives of the experiment would be to demonstrate whether or not the information can be inexpensively extracted in real time from the radar signature. For this reason, one of the subtasks in this experiment would be to program a bank of DSP chips to extract the data in real time.

This would leverage off our previous experience in implementing real-time wavelet applications on DSP hardware. A working DSP implementation would prove principle. If the information-extraction algorithm works on a DSP chip, it can always be mass-produced better, faster, and less expensively on an application-specific integrated circuit.

The simulation would require 6 person-months of direct labor. It is reasonable to assume 2 person-months of direct labor, and \$5000 of travel expenses for the ongoing interactions with the radar supplier.

To provide independent verification of the dynamics of each signature, a set of five contact switches, activated by the vehicle wheel, should be used. This would provide displacement at five points, four velocity readings, and three acceleration readings. In addition, data acquisition would require five channels of digital I/O to record the contact switch operations. This should be a sufficient check for proof of principle.

For data acquisition, eight channels (three receivers and five wheel-actuated switches) of A/D conversion are needed. The sampling rate should be at least 40 kHz. The resolution should be at least 12 bits.

Data recording would be performed by an on-hand personal computer.

For data acquisition, there is a fixed cost independent of the amount of data acquired. Hardware to perform the data collection would cost \$5,000. This cost does not include items already on hand. It also does not include payments by NHTSA to the radar supplier for access to the radar and custom engineering. Setting up and taking down the data-collecting apparatus would require 1 person-month of direct labor.

There is a variable cost in proportion to the amount of data acquired. An average day's data collection should result in 100 signatures, or a complete data set for one of the possible combinations of trajectory, vehicle type, and speed. The data collection would require a full day's work by two employees; another person-day would be required for support activities. It is assumed that the targets would be on-hand DOE vehicles, furnished at no direct cost to NHTSA.

The development of the DSP circuitry would require 6 person-months of direct labor and \$5000 for board fabrication.

Assume 2 person-months to prepare the report plus the expense for two people to travel to NHTSA to present a technical briefing.

Summary of subtasks for beam-forming proof of principle

1. Simulation study		
Direct labor		6 person-months
2. Interaction with radar supplier		
<i>(Comment: Fees paid to the radar supplier not included.)</i>		
Direct labor	2 person-months	
Other expenses	\$5000	
3. Data collection, fixed costs		
Direct labor	1 person-month	
Other expenses	\$5000	
4. Data collection, variable costs		
<i>(Comment: Expenses assume 3 person-days per labor to collect data resulting in a complete data set for one combination. Two vehicle types, 4 trajectories, and 2 speeds result in 16 possible combinations.)</i>		
Direct labor	48 person-days	
5. DSP development		
Direct labor	6 person-months	
Other expenses	\$5000	
6. Documentation		
Direct labor	2 person-months	
Travel	\$4000	

Cost summary for beam-forming proof of principle

(Does not include fees to the radar supplier)	
Direct labor	\$294,000
Other expenses	19,000
Contingency	<u>57,000</u>
Total	\$370,000

As noted earlier, the simulation study would require 6 calendar months. Discussions with the radar supplier would start at the onset of the project, with a goal of furnishing a working radar 9 months after the inception of the project. DSP development would start 4 months into the project and be completed 10 months into the project. The technical briefing at NHTSA would occur when the hardware is ready, but before serious data collection begins. Data collection would require 1 month. Production and review of the final report would require 4 months, and would not start until after the data collection is finished. Total duration of this task is 15 months.

(Note: Target bearing would be required for either functionality, arming or firing. Thus, NHTSA would not need to decide which function is its ultimate goal before this work could begin. This task is especially necessary. If it is impossible to provide good, fast, and inexpensive bearing information, then the anticipatory crash detector will not work. If the near-term funding and priorities of NHTSA are limited and it can support only one task, proof of principle with beam forming would be our first choice.)

Task 4: Identification of a Sparse On-Vehicle Sensor Array

In addition to the environment model, the anticipatory system uses a system model. This is a description of the present dynamical state of the vehicle itself. It is possible, but probably undesirable, to heavily instrument the vehicle. Sensors could determine the translational and angular velocity and acceleration of the car body relative to the road. Other sensors [ideally, the preexisting

antilock braking system (ABS) sensors] could determine the angular dynamics of each wheel. By comparing discrepancies it could be determined whether or not the vehicle is slipping or spinning. Strain sensors in the steering, brakes, and accelerator could indicate the forces acting to change the dynamics of the vehicle.

What is unclear is how much (if any) of these data are actually needed. The knowledge of whether the preexisting dynamical state of the vehicle is carrying it into or out of harm's way is probably necessary. However, 100 ms is not especially far into the future when describing the movements of something as massive as a car, and an extremely crude model, requiring a sparse set of initial conditions, is likely to be sufficient. An adequate description of initial conditions can probably be derived from existing sensors (speedometer, ABS, etc.)

The objective of this task is to determine how sparse the model and its data sources can reasonably be. It is not resolved into subtasks. Direct labor would be 3 person-months.

Task 5: System Definition of the Anticipation Engine

The novelty of this research, and the practicality and robustness of the sensory system expected to ultimately emerge from it, presume the use of a formal anticipation system. The bulk of the effort of the next phase is the development of a way to inexpensively provide a reliable description of the current state to the environment model. Nevertheless, the ultimate goal is the development of the anticipation engine itself.

With this in mind, it is reasonable for the next phase not to include development of the anticipatory system, but rather, to lay the groundwork for its development in the phase after that. The other tasks of the next phase should lead to a detailed knowledge of the nature of the current information available to the nested models. The operational requirements for an intelligent precrash-restraint system should be available from

NHTSA. Given these two items of information, it should be possible to develop a comprehensive system-definition document for the anticipation engine.

The objective of this task is to produce a system-definition document. It is not resolved into subtasks. Direct labor would be 3 person-months.

Task 6: Kinematic simulation

It would be desirable to look at the kinematics of crashes from a number of different orientations and configurations. The objective would be to run kinematic simulations that predict the occurrence of crashes and then perturb or remove selected elements of the input data. The goal is to obtain an envelope of performance for simple kinematic models. This kinematic model might also become the nucleus of the environment model of an anticipatory system in a future phase of the research. Direct labor would involve 6 person-months.

Task 7: Survey of Existing Crash Data

NHTSA's observation is that 20–30% of crashes are collinear and overlapping. All the others occur at a glancing geometry. It would be desirable to investigate the existing crash data and obtain a distribution of the probability of a collision as a function of incidence angle. Obtaining these data for analysis would probably involve more than the usual amount of travel. Direct labor would involve 4 person-months, and travel would cost \$20,000.

Task 8: Target Classification

As noted earlier, a major priority that remains to be defined by NHTSA is whether the anticipatory system being investigated in this research will be used to arm a restraint device or to fire a restraint device. A firing system requires target mass and class data; an arming system does not. It is reasonable to suppose that NHTSA does not consider that it has enough information to make this decision. Is target classification hard or easy? What is

the actual rate of misclassifications in a real-world system? How difficult is it to implement a classifier in an on-vehicle sensory system? In other words, before NHTSA tries to make a decision about whether or not an anticipatory firing system is worth doing, it might be worthwhile to determine just how much trouble it really is. That would involve study of the classification problem described in this section.

In the crash-warning literature, both past and recent, identification of the class of the target is an urgently wanted datum. The reason for this is to get some indication of the severity of the impending crash. If the type of target is not available, an indication of relative mass might be as useful.

Investigation of signatures for pattern recognition features was not part of the work for the first phase of this project. However, since target class is a desirable output, investigation of target signatures might be an objective of the next phase.

This is a fundamentally different task from the task of extracting vehicle dynamical data. Four possible approaches are discussed below: (1) identification of acoustic signatures collected by others, (2) identification of acoustic signatures collected by ORNL, (3) interpretation of active device returns for classification features, and (4) assessment of transponder returns.

Identification of acoustic signatures collected by others

Passive signature recognition would involve receiving a signature emanating from the target and identifying the target from the distinguishing features of the signature. Acoustic signatures appear to be the only emanations from a vehicle that might include enough distinguishing features to make a quick classification possible.

The work of James and Sampan (1995) is instructive. They acquired a systematic, statistically significant, and tightly controlled collection of acoustic signatures for four classes of vehicles. They did not seek a feature space; instead, they compressed the

data in an arbitrary manner. Using these compressed data sets, they explored a wide range of neural networks and typically got 80–90% correct classifications. Had their data been projected onto a good feature space, the correct classification rate would presumably have been much higher.

A reasonable line of investigation would be to acquire a copy of their uncompressed data, project it onto several different spaces, and find a good feature space. A Bayesian classifier could then be used to predict the limiting performance of a classifier based on acoustic signature identification. Then deliberate degradations could be imposed on the data to assess the reliability of an acoustic signature classifier under nonideal conditions.

Summary of subtasks for classifying acoustic signatures collected by others

1. Acquisition of data and performance of format conversion		
Direct labor	0.5 person-month	
2. Identification of a feature space		
Direct labor	3.0 person-months	
3. Training of a Bayesian classifier and assessment of results		
Direct labor	2.0 person-months	
4. Performance of a degradation study		
Direct labor	2.0 person-months	
5. Documentation		
Direct labor	0.5 person-months	
Total	8.0 person-months	

The cost (if any) of purchasing the raw data from the original experimenters would be in addition to these costs.

Identification of acoustic signatures collected by ORNL

As noted under the data acquisition task, adding a scientific microphone to the sensor array and collecting acoustic signatures simultaneously with active-device return signatures could be accomplished at relatively small incremental cost.

Thus, another reasonable line of acoustic signature investigation would be to record acoustic signatures in parallel with the active-device data collection, project it onto several different spaces, and find a good feature space. A Bayesian classifier could then be used to predict the limiting performance of a classifier based on acoustic signature identification. Then deliberate degradations could be imposed on the data to assess the reliability of an acoustic-signature classifier under nonideal conditions, such as the use of a cheap microphone in the end item sensor.

Summary of subtasks for classifying acoustic signatures collected by ORNL

1. Identification of a feature space		
Direct labor	3.0 person-months	
2. Training of a Bayesian classifier and assessment of results		
Direct labor	2.0 person-months	
3. Performance a degradation study		
Direct labor	2.0 person-months	
4. Documentation		
Direct labor	0.5 person-months	
Total	7.5 person-months	

Interpretation of active device returns for classification features

The basic shortcoming of passive methods is that no classifiable signatures emanate from some targets of interest, such as trees and guardrails. For active methods, the impending crash sensor sends out a signal and analyzes its returns for distinguishing features by class.

As described elsewhere, a key element of the next phase will be the collection of a statistically significant set of return signatures from a representative set of targets for several different active device types, possibly including Doppler, FMCW, and broadband pulse radar or ultrasonic sonar. A distinct subsequent task might be the analysis of these signatures by target class for distinguishing

features. It should be noted that there is a substantial risk that radar returns will not contain the necessary distinguishing features and that such an investigation may not lead to a useful classifier.

The following costs presume that a statistically significant set of active device return signatures already exists.

Summary of subtasks for classifying active device signatures collected by ORNL

(Comment: Expenses are shown for one active device type; there may be as many as four.)

1. Identification of a feature space	
Direct labor	3.0 person-months
2. Training of a Bayesian classifier and assessment of results	
Direct labor	2.0 person-months
3. Documentation	
Direct labor	0.5 person-months
Total	5.5 person-months

Assessment of transponder outputs

It may be that neither passive nor active signatures contain enough distinguishing features to identify the class of target. An active transponder on the target vehicle could transmit various kinds of information, including a code for vehicle type. This would lead to unambiguous identification of the target.

The major drawback is the high percentage of targets that do not have transponders. This includes fixed targets such as trees and guardrails. In addition, for the

next several decades, it is unlikely that a substantial percentage of vehicles on the road will have active transponders.

On the other hand, many of the most dangerous targets—large commercial vehicles—do have transponders for use in various intelligent transportation system (ITS) functions. It may be unwise to ignore this valuable information in those cases for which it is present.

An assessment of the value of transponders in crash warning systems might consist of two subtasks, involving 1 person-month each. The first subtask would be to acquire statistical data from NHTSA on the incidence of crashes, the distribution of vehicle classes involved, and the proliferation of transponders and projected rates of growth. This would allow preliminary estimates of the probability of collisions with a target equipped with a transponder, both now and in the foreseeable future. The second subtask, based on these data, would be to estimate savings in injury, loss of life, and property damage due to using transponder data, when available, as part of a crash prediction. One could then consider whether or not the use of available transponder data is worthwhile.

Summary of subtasks for assessing the use of active transponders

1. Acquisition of statistical data from NHTSA	
Direct labor	1.0 person-month
2. Estimates of benefit of using transponder data	
Direct labor	1.0 person-month
Total	2.0 person-months

10. RECOMMENDATIONS

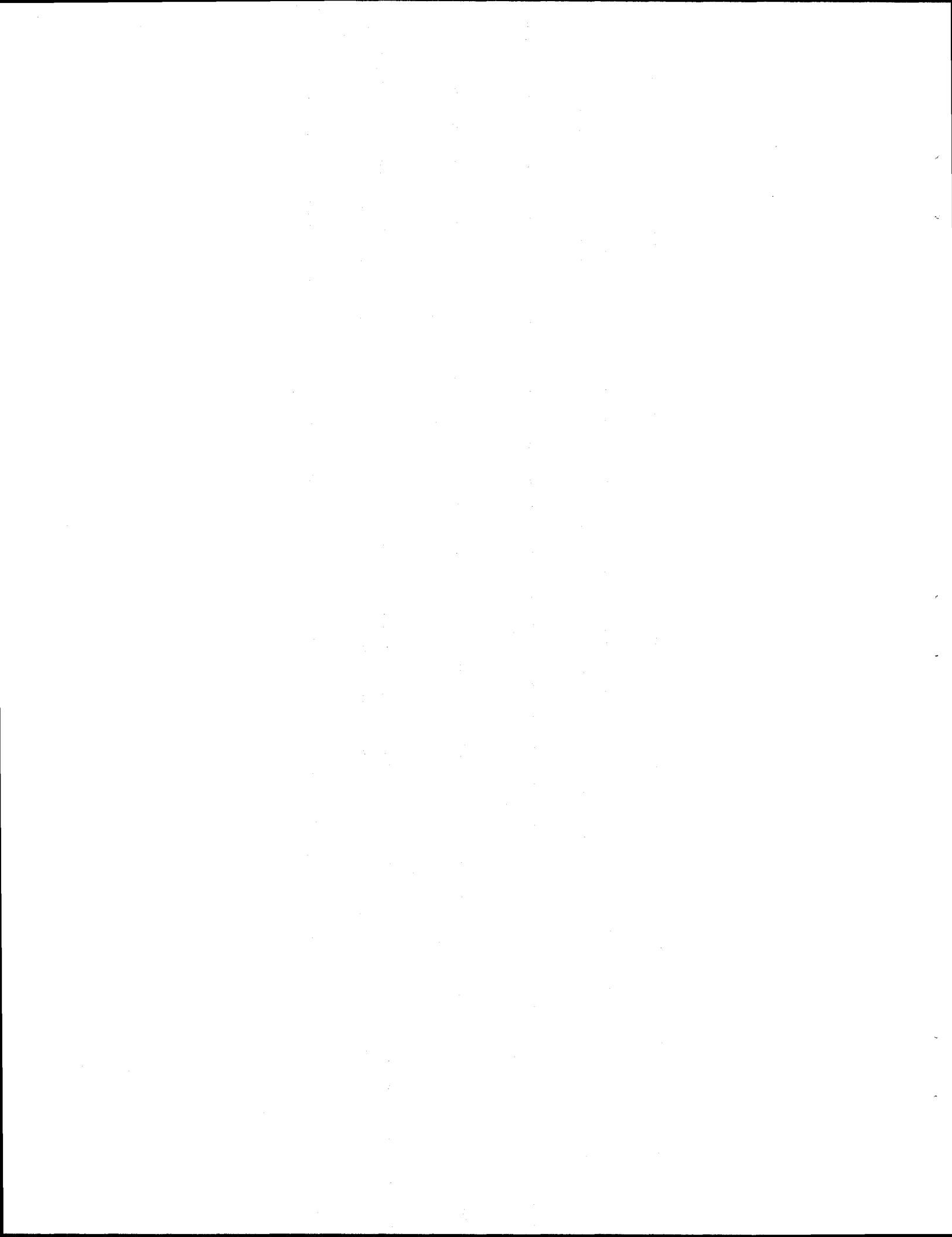
We recommend that NHTSA support the beam-forming proof-of-principle task as a near-term priority. Our investigation has shown that it is practical to extract acceleration from radar data in real time with inexpensive equipment. Velocity and displacement are less difficult to extract. What is presently missing is the bearing of the target. This datum is required for both arming and firing systems. Beam forming with pulsed radar has the potential of providing an inexpensive real-time method of simultaneously producing bearing and a threat/no threat target classification. Data on acceleration, velocity, and displacement can be obtained at virtually no incremental cost.

Before research priorities other than the investigation of beam forming are established, NHTSA must decide whether its ultimate goal is an anticipatory firing system or an anticipatory arming system. The tradeoff is a classic exercise in the judging of risk versus reward. An arming system requires only dynamical data, the acquisition of which is not especially error-prone. In addition, an arming system is tolerant of a reasonable

incidence of false predictions. The drawback is that an anticipatory arming system provides only about 10 ms of improvement over a conventional airbag actuator. An anticipatory firing system requires mass and class data, which are much more error-prone than dynamical data. The consequences of an incorrect prediction from an anticipatory firing system can be quite severe. The great advantage of this system is that it might deploy the crash restraints several tens of milliseconds faster than a contact-based actuator.

Depending on NHTSA's funding and research priorities, NHTSA may wish to sponsor several of the other tasks suggested in Section 9 in addition to the beam-forming proof-of-principle.

The ultimate goal of this research is a working prototype of an anticipatory crash detector (whether firing or arming) 5 to 10 years in the future. Development of such a prototype is unlikely to be undertaken by private industry. The objective of interim sensor development is ultimately to feed the environment model in such a detector.



REFERENCES

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

Allen, J. L. 1992. "Power Rate Crash-Sensing Method for Safety Device Actuation," SAE Technical Paper Series, presented at the International Congress and Exposition, Detroit, Feb. 24-28.

Alrabady, A. I., and Mahmud, S. M. 1993. "Development of a Decision-Making Algorithm for Airbag Control," in *Conference Record: IEEE Instrument and Measurement Technology Conference*, IEEE Catalog No. 93CH3292-0, pp. 81-84.

Baras, J. S., and Wolk, S. I. 1995. "Wavelet-Based Progressive Classification with Learning," in *Wavelet Applications II*, ed. H. H. Szu, Proceedings of SPIE 2491, pp. 339-50.

Bryzek, J., et al. 1992. "New Technologies for Silicon Accelerometers Enable Automotive Applications," SAE Technical Paper Series, Paper 920474, presented at the International Congress and Exposition, Detroit, Feb. 24-28.

Chakraborty, S., and Smedley, D. G. 1995. "Adaptive Cruise Control for Heavy-Duty Vehicles," ITS America Paper 95-040, presented at the 1995 ITS America Annual Conference, Washington, D.C., March 15-17.

Daubechies, I. 1992. *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia.

Dress, W. B., and Kercel, S. W. 1994. "Wavelet-Based Acoustic Recognition of Aircraft," in *Wavelet Applications*, ed. H. H. Szu, Proceedings of SPIE 2242, pp. 778-91.

Drumheller, D. M. 1991. *Theory and Application of the Wavelet Transform to Signal Processing*, Naval Research Laboratory Report 9316, July 31; available through NTIS (AD-A239 533).

Duckworth, G. L. 1994. "Comparative Study of Non-Intrusive Traffic Monitoring Sensors," in *Intelligent Vehicle Highway Systems*, ed. R. J. Becherer, Proceedings of SPIE 2344, pp. 16-29.

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

Fujita, Y., Akuzawa, K., and Sato, M. 1995. "Radar Brake System," ITS America Paper 95-056, presented at the 1995 ITS America Annual Conference, Washington, D.C., March 15-17.

Goodman, J. W. 1985. *Statistical Optics*, Wiley, New York.

Grace, R. H. 1991. "Semiconductor Sensors and Microstructures in Automotive Applications," SAE Technical Paper Series, Paper 910495, presented at the International Congress and Exposition, Detroit, Feb. 25–March 1.

Hashimoto, Y., Sasaki, K., and Kawai, M. 1995. "A Study of Brake Control for Collision Avoidance System," ITS America Paper 95-211, presented at the 1995 ITS America Annual Conference, Washington, D.C., March 15–17.

Holstrom, F. R., et al. 1973. "A Microwave Anticipatory Crash Sensor for Activation of Automobile Passive Restraints," *IEEE Transactions on Vehicular Technology* VT-22(2), 46–54.

Hopkins, J., et al. 1974. *Development and Evaluation of Anticipatory Crash Sensors for Automobiles*, DOT-TSC-NHTSA-73-6, Feb.; available through NTIS.

James, R. D., and Sampan, S. 1995. "Vehicle Classification of Acoustic Signals Using Neural Networks," ITS America Paper 95-131, presented at the 1995 ITS America Annual Conference, Washington, D.C., March 15–17.

Karlsen, R. E., et al. 1995. "Wavelet Analysis of Ground Vehicle Acoustic Signatures," in *Wavelet Applications II*, ed. H. H. Szu, Proceedings of SPIE 2491, pp. 560–67.

Kelley, J. P. 1993. "Sensing Considerations and Tradeoffs for Single-Point Sensing," SAE Technical Paper Series, Paper 932916, presented at the Worldwide Passenger Car Conference and Exposition, Dearborn, Mich., Oct. 25–27.

Kikuchi, H., Ishiyama, M., and Nakajima, T. 1995. "Development of Laser Radar for Radar Brake System," ITS America Paper 95-099, presented at the 1995 ITS America Annual Conference, Washington, D.C., March 15–17.

Knipling, R. 1995. Note from Ron Knipling, NHTSA Office of Crash Avoidance Research, to S. W. Kercel, ORNL, June 14.

Kweon, I. S., et al. 1993. "Sonar-Based Behaviors for a Behavior-Based Mobile Robot," *IEICE Transactions on Information and Systems* E76D(4), 479–85.

Leasure, W. A., and Burgett, A. L. 1994. "NHTSA's IVHS Collision Avoidance Research Program: Strategic Plan and Status Update," Paper 94 S3 O 01, NHTSA Office of Crash Avoidance Research.

Malik, R., and Yu, H. 1992. "The Infrared Detector Ring: Obstacle Detection for an Autonomous Mobile Robot," in *Proceedings of the 35th Midwest Symposium on Circuits and Systems*, IEEE, New York, pp. 76–79.

Najm, W. G. 1994. "Comparison of Alternative Crash Avoidance Sensor Technologies," in *Intelligent Vehicle Highway Systems*, ed. R. J. Becherer, Proceedings of SPIE 2344, pp. 62–72.

Najm, W. G. 1995. Discussion between Dr. Wassim Najm, Volpe National Transportation Systems Center, and S. W. Kercel, Oak Ridge National Laboratory, May 26.

Najm, W. G., Mironer, M. S., and Fraser, L. C. 1995. "Analysis of Target Crashes and ITS/Countermeasure Actions," ITS America Paper 95-118, presented at the 1995 ITS America Annual Conference, Washington, D.C., March 15–17.

NHTSA (National Highway Traffic Safety Administration) 1994. *A Study of Commercial Motor Vehicle Electronics-Based Rear and Side Object Detection Systems*, Final Report, DOT HS 808 080, U.S. Department of Transportation, GPO, Washington, D.C., January.

Ozguner, U., et al. 1995. "Radar-Based Integrated Vehicle Control," ITS America Paper 95-227, presented at the 1995 ITS America Annual Conference, Washington, D.C., March 15-17.

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

Rosen, R. 1985. *Anticipatory Systems: Philosophical, Mathematical, and Methodological Foundations*, Pergamon Press, Oxford.

Schwartz, W. C., 1994. "Laser Vehicle Detector/Classifier," in *Intelligent Vehicle Highway Systems*, ed. R. J. Becherer, Proceedings of SPIE 2344, pp. 81-87.

Swihart, W. R., and Lawrence, A. F. 1995. "Investigation of Sensor Requirements and Expected Benefits of Predictive Crash Sensing," SAE Paper 950347, preprinted from *Issues in Automotive Safety Technology: Offset Frontal Crashes, Airbag, and Belt Restraint Effectiveness*, presented at the SAE International Congress and Exposition, Detroit, Feb. 27-March 2.

Szu, H. 1995. Private communication between Harold Szu, NSWC, and Steve Kercel, ORNL, April 18.

Szu, H., et al. 1995. "Multiresolution Wavelet Techniques for Noisy Inverse Sensing Problems," in *Wavelet Applications II*, ed. H. H. Szu, Proceedings of SPIE 2491, pp. 481-89.

Takimoto, Y., and Kotaki, M. 1992. "Automotive Anticollision Radar," *Applied Microwave*, Fall, 70-82.

Takubo, N. 1995. "Influence of Adaptive Cruise Control on Rear-End Collision," ITS America Paper 95-014, presented at the 1995 ITS America Annual Conference, Washington, D.C., March 15-17.

Topiwala, P., and Teng, C. 1995. "Wavelet Radar Target Classification," in *Wavelet Applications II*, ed. H. H. Szu, Proceedings of SPIE 2491, pp. 420-28.

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

Tsoukalas, L. H., 1989. "Anticipatory Systems Using a Probabilistic-Possibilistic Formalism," Ph.D. Dissertation, University of Illinois.



APPENDIX: TUTORIAL ON ORTHONORMAL WAVELETS

Nonstationary Signals

The wavelet literature refers fairly casually, and without formal definition, to nonstationary signals. At a superficial level, a nonstationary signal appears to be one whose Fourier spectrum varies as a function of time. This is not a logically consistent definition; a Fourier spectrum is the projection of a signal, possibly having nonzero values of finite duration, onto a set of infinitely long basis functions. The Fourier spectrum cannot vary with time.

The distinction between stationary and nonstationary processes is established in random process theory. Goodman defines a random process as a set of functions of time, each representing a possible outcome of different instances of the process. He offers a formal mathematical definition of stationary processes and degrees of stationarity, but at a conceptual level, he says that the relevant statistical properties of a stationary process are independent of the location of the time origin (Goodman 1985, pp. 60–67). Presumably, a nonstationary process is one whose statistical properties do depend on the location of the time origin.

Wavelet theory concerns itself with processes that produce signals of finite duration. Different instances of the same process can be expected to have different Fourier spectra, even if the instances are adjacent in time. It is not too much of a conceptual stretch to suppose that if an ensemble of such signals were represented as a random process, its statistical properties would depend on the location of the time origin. Thus, an ensemble of different instances of a process, each with a different Fourier spectrum, might reasonably be said to constitute a nonstationary process, as the term appears in the wavelet literature.

Sets and Spaces

Wavelet theory makes many appeals to set theory, and the following sets are especially relevant (Akansu and Haddad 1992, p. 292). \mathbb{Z} , \mathbb{R} , and \mathbb{R}^+ are the sets of integers, real numbers, and positive real numbers, respectively. $L^2(\mathbb{R})$ is the Hilbert space of measurable, square integrable functions. In practical terms, $L^2(\mathbb{R})$ is the set of all functions, $f(t)$, that satisfy the proposition

$$\int_{-\infty}^{\infty} |f(t)|^2 dt < \infty . \quad (A.1)$$

The inner product of two functions is

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(t)g^*(t) dt . \quad (A.2)$$

Continuous Wavelet Transform

The wavelet transform of a function $f(t) \in L^2(R)$ is a measure of the correlation of the function with scaled and shifted versions of the mother wavelet, or wavelet kernel function, $\psi(t)$ (Akansu and Haddad 1992, pp. 296–300). The Morlet wavelet is an example of such a continuous wavelet kernel function; it is a Gaussian windowed complex sinusoid. If the scale parameter is a and the shift parameter is b , then the scaled and shifted wavelet is

$$\psi_{ab}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) , \quad (A.3)$$

and the continuous wavelet transform is the inner product

$$W(a, b) = \langle \psi_{ab}, f \rangle = \int_{-\infty}^{\infty} \psi_{ab}(t) f^*(t) dt , \quad (A.4)$$

where $a \in R^+$, and $b \in R$.

The frequency domain behavior of the wavelet kernel function can be observed by taking its Fourier transform (FT) (Akansu and Haddad 1992, pp. 304–10). It is a consequence of the usual properties of the FT that if $\psi_{ab}(t)$ and $\Psi_{ab}(\Omega)$ constitute an FT pair, then

$$\psi_{ab}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \Rightarrow \Psi_{ab}(\Omega) = \sqrt{a} \Psi(a\Omega) e^{-jb\Omega} . \quad (A.5)$$

There is no specific function, $\psi(t)$, that is the one and only wavelet kernel. However, for the wavelet transform to be useful, there are constraints on $\psi(t)$. Most importantly, for the wavelet transform to be invertible, the kernel must be admissible.

The inverse wavelet transform is

$$f(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_0^{\infty} \frac{dadb}{a^2} W(a, b) \psi_{ab}(t) , \quad (A.6)$$

when

$$C_\psi = \int_0^{\infty} \frac{|\Psi(\Omega)|^2}{\Omega} d\Omega . \quad (A.7)$$

The wavelet $\psi(t)$ is admissible if C_ψ is finite. This requires that $\Psi(0) = 0$. At $\Omega = 0$ the FT degenerates to

$$\Psi(0) = \int_{-\infty}^{\infty} \psi(t) dt . \quad (A.8)$$

In other words, to lead to an invertible transform, the wavelet must be admissible. To be admissible, it must have zero average value.

There is a wavelet version of the Parseval relation:

$$C_\psi \int_{-\infty}^{\infty} |f(t)|^2 dt = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |W(a, b)|^2 \frac{dadb}{a^2} . \quad (A.9)$$

Discrete Wavelet Transform

The continuous wavelet transform, $a \in R^+$ and $b \in R$, is typically impractical to implement. It is overdetermined and redundant (Akansu and Haddad 1992, pp. 310–15). A more tractable version, the discrete wavelet transform, can be created by sampling the scale and shift parameters, a and b . To create a wavelet basis set of practical usefulness, it must be convenient to reconstruct functions in terms of the basis. The process leads to several questions:

1. Is the basis set obtained by sampling a and b complete in $L^2(R)$? The basis set is the set of all the functions, $\psi_{ab}(t)$, obtained by a particular set of samples of a and b . “Complete” means that within the basis set, $\psi_{ab}(t)$, there are enough different members so that any function $f(t) \in L^2(R)$ can be expressed exactly as a weighted sum of members of the basis set.
2. Is the basis set obtained by sampling a and b redundant in $L^2(R)$? A redundant basis set may have more members than a complete set. From a redundant basis set, not only may any function $f(t) \in L^2(R)$ be exactly reconstructed as a weighted sum of basis functions, but there are also several different possible weighted sums that will reconstruct $f(t)$.
3. If the basis set obtained by sampling a and b is complete, how coarse can the sampling be made to keep the basis set minimal? If the basis set has too few elements, it will not be possible to reconstruct every $f(t) \in L^2(R)$. If the basis set has too many elements, there are multiple ways to reconstruct $f(t) \in L^2(R)$. If the basis set is minimal, it has just the right collection of elements; there will be one weighted sum to reconstruct every $f(t) \in L^2(R)$, but only one possible sum for each $f(t) \in L^2(R)$.

A reasonable sampling scheme is as follows:

$$a = a_0^m, \quad b = nb_0 a_0^m, \quad m, n \in Z \Rightarrow$$

$$\psi_{mn}(t) = a_0^{-m/2} \psi(a_0^{-m}t - nb_0) . \quad (A.10)$$

If the basis set $\psi_{mn}(t)$ is complete for some choice of $\psi(t)$, a_0 , and b_0 , then any $f(t) \in L^2(R)$ can be reconstructed as follows:

$$f(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} d_{mn} \psi_{mn}(t) , \quad (A.11)$$

where

$$d_{mn} = \langle f(t), \psi_{mn}(t) \rangle = \int_{-\infty}^{\infty} f(t) \psi(a_0^{-m}t - nb_0) dt , \quad (A.12)$$

are the wavelet coefficients.

A complete basis set, $\psi_{mn}(t)$, constitutes a frame. Frames are not in general orthonormal bases. They do not satisfy Parseval's theorem. Expansions based on frames are not unique. Instead, the following energy relations hold:

$$A \int_{-\infty}^{\infty} |f(t)|^2 dt \leq \sum_m \sum_n |\langle f, \psi_{mn} \rangle|^2 \leq B \int_{-\infty}^{\infty} |f(t)|^2 dt . \quad (\text{A.13})$$

If $\psi(t)$ is admissible, and $0 < A \leq B < \infty$, then $\psi_{mn}(t)$ constitutes a frame. For any choice of a_0 and b_0 the frame bounds are constrained:

$$A \leq \frac{\pi}{b_0 \ln a_0} \int_0^{\infty} \frac{|\Psi(\Omega)|^2}{\Omega} d\Omega \leq B . \quad (\text{A.14})$$

If $A = B = 1$, the frame is tight, but this does not guarantee linear independence for $\{\psi_{mn}\}$. A frame is exact if the removal of one element leaves it incomplete. If the frame is tight and exact, then it constitutes an orthonormal basis in $L^2(\mathbb{R})$, and Parseval's relation holds. An orthonormal wavelet is orthonormal to both scaled and shifted versions of itself:

$$\int_{-\infty}^{\infty} \psi_{mn}(t) \psi_{m'n'}(t) dt = \delta_{mn'} \delta_{nn'} . \quad (\text{A.15})$$

Recall that wavelet coefficients are generated by correlation of the signal with m th scale wavelet.

$$\begin{aligned} W(a, b) &= \langle \psi_{ab}, f \rangle = \int_{-\infty}^{\infty} \psi_{ab}(t) f^*(t) dt \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) f^*(t) dt \Rightarrow W(m, n) \\ &= \int_{-\infty}^{\infty} a_0^{-m/2} \psi(a_0^{-m}t - nb_0) f^*(t) dt \\ &= a_0^{m/2} \int_{-\infty}^{\infty} \psi(a_0^{-m}\tau - nb_0) f^*(\tau) d\tau . \end{aligned} \quad (\text{A.16})$$

The last step is simply a change of variable. Subsequently, it will be seen that it is convenient to choose $a_0 = 2$ and $b_0 = 1$. Also, the signals being processed are real, allowing $f(\tau) = f^*(\tau)$. Thus, the discrete wavelet coefficients are expressed as follows:

$$W(m, n) = 2^{m/2} \int_{-\infty}^{\infty} f(\tau) \psi(2^{-m}\tau - n) d\tau . \quad (\text{A.17})$$

This is equivalent to saying that the discrete wavelet coefficients at each scale, m , can be obtained by convolving the input signal, $f(t)$, with a filter whose impulse response is $h(t) = 2^{m/2} \psi(-2^m t)$, and sampling the output at every $t = n2^m$. This is also equivalent to saying

that the wavelet coefficients at all scales, m , can be generated by feeding the input signal into a bank of such filters.

Note that thus far in the discussion, only concepts and goals have been defined. Subsequent discussion will show that the dyadic tree is a practical structure to implement the filter bank that generates the discrete wavelet transform (DWT). Furthermore, if the filters in the structure meet the paraunitary constraint and if the low pass filter has a zero at $z = -1$, the basis function is an orthonormal wavelet.

Multiresolution Analysis

A method of constructing orthonormal compactly supported wavelets and the design of filter banks to produce them emerges from the theory of multiresolution analysis (Akansu and Haddad 1992, pp. 313–34). The idea is that $f(t) \in L^2(R)$ is expressed as the limit of successive approximations at different resolutions. The approximations are done by filtering $f(t)$ through the scaling function $\phi(t)$.

Previously, the space $L^2(R)$ was defined as the set of all square integrable functions. The subspaces $\{V_m \mid m \in \mathbb{Z}\}$ are subsets of $L^2(R)$. Each subspace corresponds to an integer. The greater the integer, the lower the resolution of the subspace, and the coarser the functions contained within it.

Note that, at some cost, coarse functions can be described in a space that allows for fine detail, but not the other way around. Fine functions cannot be described in a space that allows only for coarse resolution. Thus, V_1 contains all the coarse elements in V_2 , plus finer elements not contained in V_2 . This property of containment is described in the language of set theory as

$$\dots V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \dots \quad (\text{A.18})$$

The subspace at the limit of fine resolution contains all possible elements of $L^2(R)$. That is, $V_{-\infty} = L^2(R)$. The property of completeness is formally described as

$$\bigcap_{m \in \mathbb{Z}} V_m = \{\phi\}, \quad \bigcup_{m \in \mathbb{Z}} V_m = L^2(R) . \quad (\text{A.19})$$

Functions can be scaled by compressing or stretching. A function compressed by a factor of 2 is an element of the next finer subspace, and a function stretched by a factor of 2 is an element of the next coarser subspace. The scaling property is formally described as

$$f(x) \in V_m \Leftrightarrow f(2x) \in V_{m-1} \quad \forall f(x) \in L^2(R) . \quad (\text{A.20})$$

Within each subspace there is a basis function that is orthonormal to translation. This is the scaling function, whose prototype is $\phi(t) \in V_0$. The basis set covers the m subspaces and translations by n , and is designated $\{\phi_{mn}(t) = 2^{-m/2}\phi(2^{-m}t - n)\}$. The scaling function is orthonormal to translations within a scale, m , but not to changes in scale.

W_m is the orthogonal complement of V_m in V_{m-1} . The idea is that if $f(t) \in V_m$ is an approximation of $g(t) \in V_{m-1}$, then W_m contains the detailed information about $g(t)$ that is missing from $f(t)$. Furthermore, there is no overlap; V_m has no information in common with W_m . Symbolically,

$$V_{m-1} = V_m \oplus W_m , \quad (A.21)$$

$$V_m \perp W_m . \quad (A.22)$$

At the extreme limit of coarseness, V_∞ would contain an infinitely coarse approximation or, on a conceptual level, no information. W_∞ would contain infinitely coarse details, but on a conceptual level, these details do convey information. In fact, the details in W include all the information needed to describe the next finer level of V_m . Each succeeding finer level of detail, W_m , contains the information which, when combined with V_m , completely describes V_{m-1} . Combining the details from all possible W_m completely describes $L^2(R)$. Symbolically,

$$\dots \oplus W_j \oplus W_{j-1} \dots \oplus W_0 \dots \oplus W_{-j+1} \oplus W_{-j+2} \dots = L^2(R) . \quad (A.23)$$

The scaling function $\phi(t) \in V_0$. The wavelet function $\psi(t) \in W_0$.

Projection operators, P_m and Q_m , project any function $f \in L^2(R)$ onto V_m and W_m , respectively. In practical effect, $P_m f$ is the approximation of f at a level of coarseness m . $Q_m f$ is the detail that must be combined with $P_m f$ to obtain the approximation of f at a level of coarseness $m - 1$:

$$P_{m-1} f = P_m f + Q_m f . \quad (A.24)$$

From this, it follows that $Q_m f = P_{m-1} f - P_m f$, $Q_m f \in W_m$. From this, it follows that the complimentary space is $W_m = V_{m-1} \ominus V_m$. Translates of the scaling function $\phi(t - n) \in V_0$ and span the space V_0 . Any element of V_0 can be constructed as a weighted sum of $\phi(t - n)$ terms if enough different values of n are used. Similarly, $\phi(2t - n) \in V_{-1}$ and spans the space V_{-1} . Given the relationship between the complement and the two subspaces, it is expected that there should be a function $\psi(t - n) \in W_0$ that spans W_0 . Because of the scaling property of V_m , the scaled wavelet $\psi_{mn}(t) = \psi(2^{-m}t - n)$ will span W_m .

The containment and completeness properties of V_m , and the propositions $W_m \perp V_m$ and $V_{m-1} = V_m \oplus W_m$, imply that all the complimentary spaces W_m are mutually orthogonal and that all the information contained in all W_m completely constitutes $L^2(R)$. The expression $\psi_{mn}(t)$ constitutes an orthonormal basis in W_m . Since spaces W_m are mutually orthogonal, it follows that $\psi_{mn}(t)$ constitutes an orthonormal basis in $L^2(R)$, being orthonormal to scales, m , and translates, n , of the mother wavelet, $\psi(t)$.

The point of all this discussion is that if the goal is to generate an orthonormal wavelet, it is convenient to start with a scaling function that meets the conditions described above. If $\phi(t - n) \in V_0$ spans V_0 , then $\phi(2t - n) \in V_{-1}$ spans V_{-1} . As a consequence of the containment property, the coarser version of the scaling function can be generated as the weighted sum of translates of the finer version:

$$\phi(t) = 2 \sum_n h_0(n) \phi(2t - n) . \quad (A.25)$$

Similarly, the wavelet function can be generated as the weighted sum of translates of the next finer version of the scaling function:

$$\psi(t) = 2 \sum_n h_1(n) \phi(2t - n) . \quad (A.26)$$

The process of generating a scaling function or an orthonormal wavelet from a weight list is most easily understood in the Fourier frequency domain:

$$\phi(t) \leftrightarrow \Phi(\Omega) \Rightarrow \phi(2t - nT_0) \leftrightarrow \frac{1}{2} e^{-jn\Omega T_0/2} \Phi\left(\frac{\Omega}{2}\right). \quad (\text{A.27})$$

The FT of the weighted sum that generates the scaling function is

$$\Phi(\Omega) = \sum_n h_0(n) e^{-j(\Omega/2)nT_0} \Phi\left(\frac{\Omega}{2}\right). \quad (\text{A.28})$$

Define $\omega = \Omega T_0$, and recall that h_0 is a discrete list of weights, or a sequence, and that the FT of a sequence is $H_0(e^{j\omega}) = \sum_n h_0(n) e^{-jn\omega}$. The FT of the scaling function then becomes

$$\Phi(\Omega) = H_0(e^{j\omega/2}) \Phi\left(\frac{\Omega}{2}\right). \quad (\text{A.29})$$

By a similar argument

$$\Phi\left(\frac{\Omega}{2}\right) = H_0(e^{j\omega/4}) \Phi\left(\frac{\Omega}{4}\right). \quad (\text{A.30})$$

Hence

$$\Phi(\Omega) = H_0(e^{j\omega/2}) H_0(e^{j\omega/4}) \Phi\left(\frac{\Omega}{4}\right). \quad (\text{A.31})$$

The process can be repeated indefinitely. Applying the observation that $\lim_{n \rightarrow \infty} \Omega/2^n = 0$, the FT of the scaling function is seen to be an infinite product:

$$\Phi(\Omega) = \Phi(0) \prod_{k=1}^{\infty} H_0(e^{j\omega/2^k}). \quad (\text{A.32})$$

As a consequence of the completeness property, and after some mathematical steps, it follows that $|\Phi(0)| = 1$ and that $|H_0(e^{j\omega})|_{\omega=0} = 1$. If the weight list $h_0(n)$ is finite over the range of $0 \leq n \leq N - 1$, then the scaling function has compact support over the same range $[\phi(t) = 0 | 0 \leq t \leq N - 1]$.

There are some constraints on $H_0(e^{j\omega})$ that will assure that $\phi(t)$ is orthonormal to shifts within a given scale, m . If $\phi(t - n)$ spans V_0 and is orthonormal, then the unitary condition must be satisfied (Akansu and Haddad 1992, Appendix B):

$$\sum_k |\Phi(\Omega + 2\pi k)|^2 = 1. \quad (\text{A.33})$$

Into this equation we can substitute

$$\Phi(2\Omega) = H_0(e^{j\omega}) \Phi(\Omega). \quad (\text{A.34})$$

After a good deal of algebra, what remains is

$$|H_0(e^{j\omega})|^2 + |H_0(e^{j\omega+\pi})|^2 = 1 . \quad (\text{A.35})$$

This is within a normalizing constant of the constraint on the low-pass filter in a maximally decimated perfect reconstruction filter bank.

By similar reasoning, constraints on H_1 can be obtained that ensure that the wavelet, $\psi(t)$, is orthonormal. It should be recalled that the wavelet is related to the scaling function at the next finer scale:

$$\psi(t) = 2 \sum_n h_1(n) \phi(2t - n) . \quad (\text{A.36})$$

The FT of a sequence $h_1(n)$ is $H_1(e^{j\omega})$. Hence,

$$\Psi(\Omega) = H_1(e^{j\omega_2}) \Phi\left(\frac{\Omega}{2}\right) . \quad (\text{A.37})$$

The infinite product representation of the scaling function can be substituted into this:

$$\Psi(\Omega) = H_1(e^{j\omega_2}) \prod_{k=2}^{\infty} H_0(e^{j\omega_2^k}) . \quad (\text{A.38})$$

If $\psi(t - n)$ spans W_0 and is orthonormal, then the unitary condition must be satisfied:

$$\sum_k |\Psi(\Omega + 2\pi k)|^2 = 1 . \quad (\text{A.39})$$

Into this equation we can substitute

$$\Psi(2\Omega) = H_1(e^{j\omega}) \Phi(\Omega) . \quad (\text{A.40})$$

After a good deal of algebra, what remains is

$$|H_1(e^{j\omega})|^2 + |H_1(e^{j\omega+\pi})|^2 = 1 . \quad (\text{A.41})$$

This is within a normalizing constant of the constraint on the high-pass filter in a maximally decimated perfect reconstruction filter bank.

In addition, the wavelet function is orthogonal to the scaling function. The consequent unitary condition in the frequency domain is

$$\sum_k \Phi(\Omega - 2k\pi) \Psi^*(\Omega - 2k\pi) = 0 . \quad (\text{A.42})$$

Using the infinite product representations of Φ and ψ in the above and applying even more manipulations lead to

$$H_0(e^{j\omega}) H_1(e^{-j\omega}) + H_0[e^{j(\omega+\pi)}] H_1[e^{-j(\omega+\pi)}] = 0. \quad (\text{A.43})$$

This can be satisfied if $h_1(n) = (-1)^{n+1}h_0(N-1-n)$, or in the z domain, $(H_1(z) = z^{-(N-1)}H_0(-z^{-1})$. Letting $z \rightarrow e^{j\omega}$, and substituting into either orthonormality equation, this results in

$$|H_0(e^{j\omega})|^2 + |H_1(e^{j\omega})|^2 = 1. \quad (\text{A.44})$$

Since it is already known that $|H_0(e^{j\omega})|_{\omega=0} = 1$, the preceding requires that $|H_1(e^{j\omega})|_{\omega=0} = 0 = \sum_n h_1(n)$. In other words, in order to generate invertible, compactly supported orthonormal wavelets, $H_1(z)$ must have a zero at $z = 1$.

To appreciate how the multiresolution analysis operates, start with a function $f \in V_0$. It can be represented as the weighted sum of translates of the scaling function:

$$f(t) = \sum_n c_{0,n} \phi(t-n), \quad (\text{A.45})$$

where

$$c_{0,n} = \langle f, \phi(t-n) \rangle = \int f(t) \phi(t-n) dt. \quad (\text{A.46})$$

It is also possible to express f in terms of the next coarser approximation, and the interscale details:

$$f_v^1(t) = P_1 f = \sum_n c_{1,n} \phi\left(\frac{t}{2} - n\right), \quad (\text{A.47})$$

$$f_w^1(t) = Q_1 f = \sum_n d_{1,n} \psi\left(\frac{t}{2} - n\right), \quad (\text{A.48})$$

$$f(t) = P_1 f + Q_1 f. \quad (\text{A.49})$$

Multiply $f(t) = f_v^1(t) + f_w^1(t)$ by $\phi(t/2 - n)$ and integrate, resulting in $\langle f, \phi(t/2 - n) \rangle = \langle f_v^1(t) \phi(t/2 - n) \rangle + \langle f_w^1(t) \phi(t/2 - n) \rangle$. The second term is the weighted sum of translates of $\psi(t/2 - n)$, which are orthogonal to $\phi(t/2 - n)$. The second term must be zero. The result is that $c_{1,n} = \langle f, \phi(t/2 - n) \rangle = \langle f_v^1(t) \phi(t/2 - n) \rangle$. A similar result obtains for $d_{1,n}$.

$$c_{1,n} = \frac{1}{\sqrt{2}} \int f_v^1(t) \phi\left(\frac{t}{2} - n\right) dt, \quad (\text{A.50})$$

$$d_{1,n} = \frac{1}{\sqrt{2}} \int f_w^1(t) \psi\left(\frac{t}{2} - n\right) dt. \quad (\text{A.51})$$

It has already been established that the scaling function at different scales is related by

$$\phi(t) = 2 \sum_n h_0(n) \phi(2t - n) . \quad (A.52)$$

Therefore,

$$\phi\left(\frac{t}{2} - n\right) = 2 \sum_k h_0(k) \phi(t - 2n - k) . \quad (A.53)$$

This can be substituted into the integral for $c_{1,n}$ and simplified with a little manipulation. A similar operation leads to $d_{1,n}$.

$$c_{1,n} = \sqrt{2} \sum_k h_0(k - 2n) c_{0,k} , \quad (A.54)$$

$$d_{1,n} = \sqrt{2} \sum_k h_1(k - 2n) c_{0,k} . \quad (A.55)$$

In general, a filter by $g(n)$ followed by a downsample by 2 is realized as $y(n) = \sum_k g(2n - k)x(k)$. Hence, the two processes leading to the next coarsest approximation and details consist of filtering the approximation by $2^{1/2}h_0(-n)$ and $2^{1/2}h_1(-n)$, respectively, and downsampling the results of both by 2. The process can be repeated as often as necessary to obtain the next coarsest approximation and details.

Synthesizing the original signal from this decomposition is done as follows. Starting with the interscale equations $\phi(t) = 2 \sum_n h_0(n) \phi(2t - n)$ and $\psi(t) = 2 \sum_n h_1(n) \phi(2t - n)$, it can be shown that

$$h_0(n) = \frac{1}{2} \int \phi\left(\frac{t}{2}\right) \phi(t - n) dt , \quad (A.56)$$

$$h_1(n) = \frac{1}{2} \int \psi\left(\frac{t}{2}\right) \phi(t - n) dt . \quad (A.57)$$

The reconstructed fine estimate has already been established as $c_{0,n} = \langle f_v^1, \phi(t - n) \rangle + \langle f_w^1, \phi(t - n) \rangle$. The first term can be expressed as follows:

$$\begin{aligned} \langle f_v^1, \phi(t - n) \rangle &= \int \phi(t - n) f_v^1 dt \\ &= \int \phi(t - n) \frac{1}{\sqrt{2}} \sum_k c_{1,k} \phi\left(\frac{t}{2} - k\right) dt \\ &= \sum_k c_{1,k} \frac{1}{\sqrt{2}} \int \phi(t - n) \phi\left(\frac{t}{2} - k\right) dt . \end{aligned} \quad (A.58)$$

The integral is seen to be $2h_0(n - 2k)$. Hence, the first term in $c_{0,n}$ is $2^{1/2} \sum_k c_{1,k} h_0(n - 2k)$. By similar reasoning, the second term in $c_{0,n}$ is $2^{1/2} \sum_k d_{1,k} h_1(n - 2k)$. In general, an upsample by 2 followed by a filter by $g(n)$ is realized as $y(n) = \sum_k g(n - 2k)x(k)$. Hence, the two processes of

obtaining the approximation from the next coarsest approximation and details consist of upsampling both input streams by 2 and filtering the upsampled streams by $2^{1/2}h_0(n)$ and $2^{1/2}h_1(n)$.

The wavelet-based decomposition and reconstruction scheme is identical to the dyadic-tree-based scheme developed for perfect reconstruction subband coders.

Cascade Algorithm

Except for trivial cases, it is not possible to express orthonormal wavelet functions or their corresponding scaling functions in closed form. For finite impulse response (FIR) filters, the proposition that $\psi(t) = \sum h_1(n)\phi(2t - n)$ means that the wavelet function is simply the weighted sum of translates of the next finer scaling function. The scaling function is expressed implicitly; $\phi(t)$ is the function that satisfies $\phi(t) = \sum h_0(n)\phi(2t - n)$. What is needed is a convenient method for computing $\phi(t)$.

Daubechies developed a fast method for approximating the value of $\phi(t)$ to any desired accuracy; she styles it the “cascade algorithm” (Daubechies 1992, pp. 202–6). The justification for the algorithm begins with the observation that the scaling function has finite support and $\int \phi(x) dx = 1$. A consequence of this is that $2^j\phi(2^jx)$ also has an area of 1 for all integer j , and as $j \rightarrow \infty$, $2^j\phi(2^jx) \rightarrow \delta(x)$. Due to the sifting property, $f(x) = \int f(x + y) \delta(y) dy$. Hence,

$$f(x) = \lim_{j \rightarrow \infty} 2^j \int f(x + y) \phi^*(2^jy) dy . \quad (A.59)$$

If x is a dyadic rational, $x = 2^{-j}K$, Daubechies shows that the above leads to

$$\phi(w^{-j}K) = \lim_{j \rightarrow \infty} 2^{j/2} \langle \phi, \phi_{-j}, 2^{-j}K \rangle . \quad (A.60)$$

Define an interpolating function $\eta(x)$ as a piecewise constant (on subintervals $[2^{-j}p, 2^{-j}p + 1]$) approximation of $\phi(x)$. From the above, $\eta_j(2^{-j}p) = 2^{j/2} \langle \phi, \phi_{-j}, 2^{-j}p \rangle$. The inverse wavelet transform consists of a cascade of the following step:

$$\langle f, \phi_{j-1, k} \rangle = c_{j-1, m} = \sum_k h_{m-2k} c_{jk} + g_{m-2k} d_{jk} . \quad (A.61)$$

As a consequence of the foregoing and the orthogonalities $\langle \phi, \phi_{0m} \rangle = \delta_{0m}$ and $\langle \phi, \psi_{-jk} \rangle = 0$, the following recursion ensues:

$$\eta_j(w^{-j}p) = \sqrt{2} \sum_q h_{p-2q} \eta_{j-1}(2^{1-j}q) . \quad (A.62)$$

In this formula, h is the low-pass FIR filter in the inverse wavelet transform cascade (or synthesis bank). The filter is normalized such that the sum of its coefficients is $2^{1/2}$. It has length $2n$ and is translated such that it spans from h_{-n} to h_{n-1} . All other values of h are zero. The bottom of the recursion is $\eta_0(p) = \delta_{0, p}$.

Regularity

The fact that $\Phi(\Omega) = \prod_{k=1}^{\infty} H_0(e^{j\omega/(2^k)})$ leads to another constraint on the design of wavelet functions (Akansu and Haddad 1992, pp. 336–41). Depending on the properties of H_0 , the infinite product might converge to a smooth function or to a fractal function. It is desirable that

Φ , and the corresponding functions ϕ , Ψ , and ψ , be smooth, and regularity is a measure of smoothness.

Regularity is defined as the maximum value of r for which the following proposition is true:

$$|\Phi(\Omega)| \leq \frac{c}{(1 + |\Omega|)^{r+1}}, \quad (A.63)$$

where $r \geq m$, $\phi(t)$ is m times continuously differentiable.

To generate a wavelet, $H_0(z)$ must have at least one zero at $z = -1$. Suppose that it has L zeros at $z = -1$ and consists of an N coefficient FIR filter. For $P(z)$ is a polynomial in z^{-1} and of degree $N - 1 - L$, the filter is characterized as follows:

$$H_0(z) = \left(\frac{1+z^{-1}}{2} \right)^L P(z) \Rightarrow$$

$$|H_0(e^{j\omega})| = \left| \cos \frac{\omega}{2} \right|^L |P(e^{j\omega})|. \quad (A.64)$$

This can be substituted into the infinite product form. After some manipulation, the following result ensues:

$$|\Phi(\Omega)| = \prod_{k=1}^{\infty} \left| \cos \frac{\omega}{2^{k+1}} \right|^L \prod_{k=1}^{\infty} |P(e^{j\omega_2^k})|$$

$$= \left| \frac{\sin \omega/2}{\omega/2} \right|^L \prod_{k=1}^{\infty} |P(e^{j\omega_2^k})|. \quad (A.65)$$

If the last term is bounded, then the sinc term will contribute to the decay of Φ . If $l > 1$ and $P(e^{j\omega})$ satisfies the following proposition, then $\phi(t)$ is m times continuously differentiable.

$$\max_{\omega \in R} \left| \prod_{k=0}^1 P(e^{j\omega_2^k}) \right| \leq 2^{1(N-m-1)}. \quad (A.66)$$

The presence of L zeros in $H_0(z)$, and the constraint that orthonormality imposes on the relationship between $H_0(z)$ and $H_1(z)$, imply a degree of flatness on both filters, described as follows:

$$\frac{d^r H_0(e^{j\omega})}{d\omega^r} \Big|_{\omega=\pi} = \frac{d^r H_1(e^{j\omega})}{d\omega^r} \Big|_{\omega=0} = 0, \quad 0 \leq r \leq L - 1. \quad (A.67)$$

The imposition of zeros to ensure regularity has another important consequence. It provides a reasonable foundation for a method to actually design the filters to generate orthonormal wavelets. To reiterate a point already made, forcing the $H_0(z)$ to have L zeros at $z = -1$ requires the following:

$$H_0(z) = \left(\frac{1+z^{-1}}{2} \right)^L P(z) \Rightarrow$$

$$|H_0(e^{j\omega})| = \left| \cos \frac{\omega}{2} \right|^L |P(e^{j\omega})| , \quad (A.68)$$

where $P(z)$ is a polynomial in z^{-1} with real coefficients. Therefore, $Q(z) = P(z)P(1/z)$ is a symmetric polynomial.

$$Q(z) = \sum_{-N-1-L}^{N-1-L} q_n z^{-n}, \quad q_n = q_{-n} \Rightarrow$$

$$Q(e^{j\omega}) = |P(e^{j\omega})|^2 = q_0 + 2 \sum_{n=1}^M q_n \cos(n\omega) . \quad (A.69)$$

Here, $\cos(n\omega)$ is a polynomial in $\cos(\omega)$, which can be expressed in terms of $\sin^2(\omega/2)$. Thus, $|P(e^{j\omega})|^2$ is a polynomial, $f(x)$, of degree $N - 1 - L$, where $x = \sin^2(\omega/2)$. By some manipulation, the term that imposes zeros, $\{(1 + 1/z)/2\}^L$, becomes $(1 - x)^L$. Hence,

$$|H_0(e^{j\omega})|^2 = (1 - x)^L f(x) . \quad (A.70)$$

The power complimentary condition is a consequence of the orthonormality requirement:

$$|H_0(e^{j\omega})|^2 + |H_0[e^{j(\omega+\pi)}]|^2 = 1 . \quad (A.71)$$

Substituting the polynomial expression into power complimentary, and noting that $R(x)$ is an odd polynomial such that $R(x) = -R(1 - x)$, the following result occurs:

$$(1 - x)^L f(x) + x^L f(1 - x) = 1 \Rightarrow$$

$$f(x) = \sum_{k=0}^{L-1} \binom{L-1-k}{k} x^k + x^L R(1 - 2x) . \quad (A.72)$$

By selecting various possibilities for R and L , the power spectrum for the H_0 filter for various wavelet families can be specified in terms of x . Since $x = \sin^2(\omega/2)$, the power spectrum can be expressed in terms of complex exponentials in ω . With the substitution $e^{j\omega} \rightarrow z$, the power spectrum becomes a symmetric polynomial in z , $|H_0(z)|^2 = H_0(z)H_0(1/z)$. With some difficulty, $H_0(z)$ can be recovered from the spectrum by one of several spectral factorization techniques that appear in the literature.

Daubechies Wavelet

One of the more popular wavelet designs is one devised by Daubechies (Daubechies 1992, pp. 194–202). She assumes that $R(x) = 0$ and $L = N/2$, where N is the length of the generating FIR filter. She factorizes the resulting polynomial to obtain the filter design.

The number of derivatives of $|H_0(e^{j\omega})|^2$ at $\omega = 0$ and $\omega = \pi$ are equal, and the Daubechies filter is identical to the maximally flat FIR filter (Akansu and Haddad 1992, pp. 248–50). The constraint of maximal flatness can be imposed by forcing selected derivatives equal to zero at $\omega = 0$ and $\omega = \pi$ at the outset. From this point it is possible, independent of wavelet methods, to derive the same power spectrum polynomial as Daubechies derived from the regularity formula.

In Drumheller's notation, $\psi_n(t)$ is the Daubechies wavelet generated by a FIR filter of length $2n$ (Drumheller 1991). He proves some useful properties of the Daubechies wavelet: $\psi_n(t)$ is bounded for all n , is continuous for all n , is continuously differentiable for $n \geq 4$, and has finite spectral variance, $\int_{-\infty}^{\infty} \Omega^2 |\Psi_n(\Omega)| d\Omega < \infty$, for $n \geq 4$.

INTERNAL DISTRIBUTION

1. G. T. Alley
2. G. O. Allgood
3. B. M. Beets
4. C. R. Brittain
5. C. L. Carnal
- 6-10. W. B. Dress
11. B. G. Eads
12. P. D. Ewing
13. R. K. Ferrell
14. R. A. Hess
15. R. B. Honea
16. J. M. Jansen
- 17-21. S. W. Kercel
22. D. W. McDonald
23. G. N. Miller
24. M. R. Moore
25. A. K. Rathi
26. J. O. Stiegler
- 27-29. R. W. Tucker, Jr.
30. R. E. Uhrig
31. J. D. White
- 32-33. Central Research Library
34. Y-12 Technical Reference Section
- 35-36. Laboratory Records
37. Laboratory Records—Record Copy
38. ORNL Patent Section
39. I&C Division Publications Office

EXTERNAL DISTRIBUTION

40. Assistant Manager for Energy Research and Development, DOE-ORO, P.O. Box 2001, Oak Ridge, TN 37831-8600
41. Ron Moore, The RM Group, 12024 Broadwood Dr., Knoxville, TN 37922
42. David Norton, Houston Advanced Research Center, 4800 Research Forest Dr., The Woodlands, TX 77381
43. Maurice M. Sevik, Carderock Division, Naval Surface Warfare Center, Code 1900, Bethesda, MD 20084-5000
44. Ernesto Suarez, Pratt & Whitney, P.O. Box 109600, MS 716-87, West Palm Beach, FL 33410-9600
45. P. I. Moynihan, Jet Propulsion Laboratory, MS 306-431, 4800 Oak Grove Dr., Pasadena, CA 91109-8099
- 46-48. David Kelley, Radar Systems, Amerigon, 404 East Huntington Dr., Monrovia, CA 91016
- 49-51. Randa Samaha, Office of Crashworthiness Research, National Highway Traffic Safety Administration, Code NRD-12, Washington, DC 20590
- 52-54. Wassim Najm, Volpe National Transportation Systems Center, Kendall Square, Cambridge, MA 02142
- 55-56. Office of Scientific and Technical Information, U.S. Department of Energy, P.O. Box 62, Oak Ridge, TN 37831