Published in
The International Journal of Robotics Research
13(5), Oct. 1994

Application of Change Detection to Dynamic Contact Sensing

Brian Eberman

J. Kenneth Salisbury, Jr.

Abstract

The forces of contact during manipulation convey substantial information about the state of the manipulation. Textures, slip, impacts, grasping, and other contact conditions produce force and position signatures that can be used for identifying the state of contact. This paper address the fundamental problems of interpreting the force signals without any additional context on the state of manipulation. Techniques based on forms of the generalized sequential likelihood ratio test are used to segment individual strain signals into statistically equivalent pieces. The results of the segmentation are designed to be used in a higher level procedure which will interpret the results within a manipulation context. We report on our experimental development of the segmentation algorithm and on its results for detecting and labelling impacts, slip, changes in texture, and condition. The sequential likelihood ratio test is reviewed and some of its special cases and optimal properties are discussed. Finally, we conclude by discussing extensions to the techniques and lessons for sensor design.

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-91-J-4038. The authors were also supported by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contracts N00014-92-J-1814 and N00014-91-J-4038; and in part by the Office of Naval Research University Research Initiative Program under Office of Naval Research contract N00014-86-K-0685.

1 Introduction

A tremendous amount of information is available in the contact forces of manipulation. Figure 1 shows a spectrogram of an impact event. The impact results in an increase in energy at all frequencies locally around the event, and a persistent residual vibration at the sensor's natural frequency. This signal, like all contact force signals, can be broken into regions that are similar. We term the different regions *contact states* and the transitions between regions *contact events*.

During a manipulation task like grasping, pushing, or typing, numerous such events occur. One approach to robot programming, for any of these tasks, is to write a program component that detects and labels each event and a component which generates an appropriate action based on the sensed event. The two components must mesh because each provides a decision or knowledge context for the other. The current action selects possible interpretations of sensor signals, and the changes in sensor signals guide the choice of new actions. In general, most robot programming has used an action centered paradigm. That is the programmer first determines the sequence of actions that should be performed and then determines how to use the available sensors to guide the actions. The guarded move is the typical sensing/action strategy that results. Brock [8] provides a recent study and review of this approach.

This paper is based on a sensing centered programming paradigm. In this approach, the properties characteristic of the sensor signals alone are determined and then an event detector for changes in those properties is designed. Ideally these properties are not biased by any particular task model and therefore are useful for all tasks. Because the approach is sensor based, the types of allowed contexts are controlled by the possible interpretations given the sensor measurements. The actions are then designed based on these possible contexts. This paper discusses our work on identifying characteristic properties of contact force signals.

In manipulation, force and position events guide the task. This paper focuses on force signals since little work has been done on interpreting these signals during common tasks. Earlier investigations of this problem, by other researchers, focused primarily on designing and proving sensor technologies tailored to different contact events. In contrast, this paper presents algorithms, derived from the theory of sequential hypothesis testing, that are designed for detecting contact events and labeling contact states.

Labeling is much more difficult then detecting changes in signal characteristics. In general, labeling requires knowledge about the possible sources of the signals and the source characteristics. This research looked at labeling some simple events without context.

To our knowledge, this is the first investigation of the discrimination of contact events and the first application of sequential decision theory to this problem. Although we applied the techniques to signals from an intrinsic contact sensor [5], the ideas can be applied to any form of tactile sensor to enhance performance. Based on the experimental results in this paper, we are investigating a sensor that is a combination of the intrinsic contact sensor and a piezoelectric film. This sensor, which is similar in design to [12], has better dynamic characteristics and should give better performance.

In the following sections, we first summarize the work that has been done on detecting grasping events and on human models of temporal tactile sensing. We then briefly discuss our experiments and the signal models. Then sequential hypothesis testing is introduced and used to derive the necessary statistical tests. Next our algorithm is presented and related to the general theory. We then present the experimental results of our algorithm, discuss its theoretical characteristics and compare it to other techniques. We conclude by discussing the research issues in contact event perception, and our plans for future development.

2 Previous Work

During the last decade, considerable research has been performed on tactile sensing. Howe [13] provides the most recent comprehensive review of current and past research. Most of this research has focused on designing surface array sensors and using these sensors for obtaining geometric information from static measurements. Some research has looked at the information that can be acquired by actively moving the contact sensor and monitoring both the sensor and joint locations. This is termed haptic sensing. Primarily prior haptic research has focused on actively tracing the contours of objects to determine geometry and critical features [7, 27]. This work assumes that each measurement is taken with the force sensor in a quasi-static state so that normal forces and contact locations can be computed. All of this work essentially treats the tactile array sensor as a primitive form of vision.

In contrast, a more recent type of contact sensor processing is the detection of information characteristic of the dynamic aspects of motion [13]. Mechanical properties of objects like mass, friction, and damping can only be determined by actively probing and manipulating the object. Similarly, the initial contact with an object, and slip between the sensors and environment require detecting relative motion. All of these sensing modalities are unique to contact force sensing.

A few studies have been done on this type of sensing. By monitoring the acoustic emission from a metal gripper, Dornfeld [9] was able to detect the onset of slip for some metallic workpieces. Howe and Cutkowsky [12] constructed an instrumented latex covered finger. Piezoelectric sensors are embedded in the latex cover, and a miniature accelerometer is mounted on the inside surface of the cover. The piezoelectric sensors are very sensitive to strain rate. Because of the small mass of the cover, the accelerometers are sensitive to very small forces normal to the surface of the sensor. They found that the piezoelectric sensor was very sensitive to the changes in tangential strain associated with slip, and that the accelerometer was fairly sensitive to the vibrations normal to the sensor associated with slip.

Bicchi [6] used a six-axis fingertip force-torque sensor to estimate the onset of slip (figure 2). This sensor has a

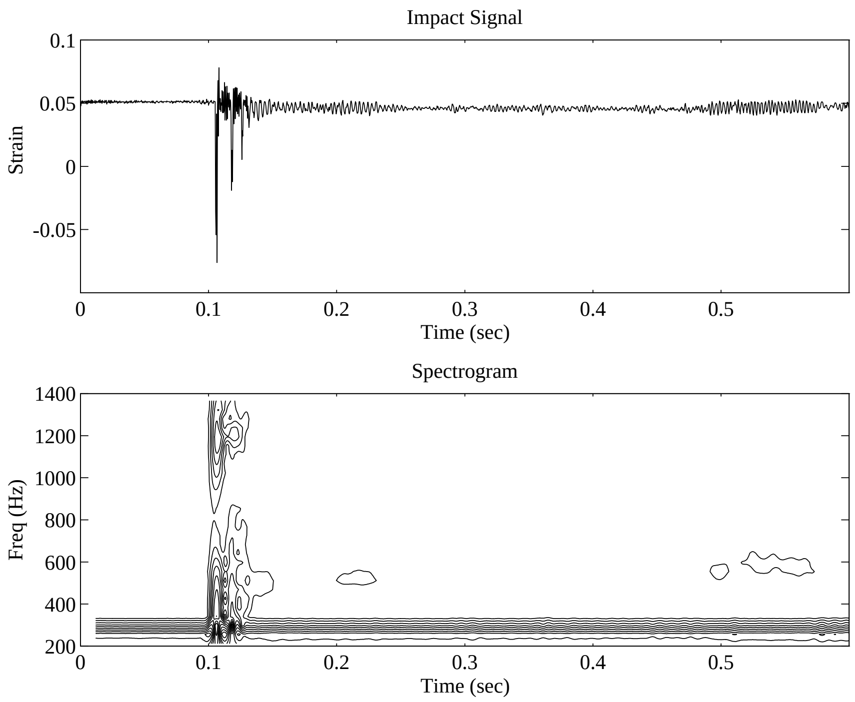

Figure 1: Spectrogram of an impact event. The figure shows a contour plot of the energy in frequencies from 200-1350 Hz as a function of time. The signal was sampled at 2700 Hz. Sixty-four points windowed with a Hamming window were used for each fast fourier transform (fft). The fft was computed for every new data point. Note the broad frequency band that occurs at an impact and the short time scale of this event.

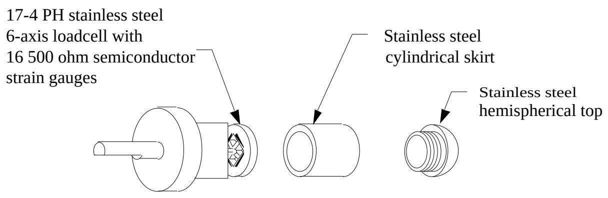

Figure 2: 6-axis fingertip force-torque sensor

Maltese-cross connecting the outer shell to the base. The cross is instrumented with 8 strain-gauge half-bridges. The shell has a lightly damped natural frequency of approximately 690 Hz when the base is fixed and the shell free. In his experiments, Bicchi first determined a coefficient of friction for the object to be grasped. Then, by monitoring the ratio of the tangential force to the normal force, he was able to determine when the contact state was approaching the slip condition determined earlier.

All of these methods generally make decisions about the contact state based on the instantaneous values of the measured signals. In some cases, a lowpass filter may be introduced to reduce the "noise" in the signal. One exception is McCarragher [20]. McCarragher examined the planar assembly process and constructed a discrete event dynamic system controller to make decisions about the current configuration of the parts based on the history of force measurements. He used qualitative reasoning on the assembly dynamics to construct interpretations of the force signal, and a Petri net to provide decision context.

In contrast, the contribution of this paper is to show how the entire history of the signal can be used to make decisions in a statistically robust way using the techniques of sequential decision theory. This technique is applicable to all of the sensors that have been investigated for dynamic contact sensing. In our experiments we have applied it to the 6-axis fingertip force-torque sensor.

In this paper, only the individual strain gauge signals produced by the sensor (figure 2) are sensed. Each signal is treated as independent in order to determine what information can be extracted about the contact process purely from the characteristics of the scalar signal. Without additional measurements or prior knowledge, all event decisions must be based on characteristics of the individual strain time series. We test the signal for whiteness, changes in mean, changes in vibration level, short discontinuities caused by impulses, and spectral structure. The tests are formulated as hypothesis testing problems based on experimental models of the signal characteristics. Section 4 presents our experiments and the signal models. Section 5 discusses the theory of sequential hypothesis testing and develops the form of the test used in our algorithm. Section 6 shows how the algorithm segments some contact signals and discusses performance.

3 Human Capabilities

Some insights on the design of robot contact sensing algorithms may be gained by a study of human temporal contact sensing. A robot sensing through an intrinsic contact sensor or a single piezoelectric sensor is analogous to human exploration using a stick. The stick encodes the distributed contact information into a temporal force signal which is transmitted along the length of the stick. A surprising amount of information can be gained purely through this channel. Our goal in designing intrinsic contact sensors and algorithms is to understand and at least match human performance in this mode of exploration.

Research on human tactile perception has shown that there are four types of mechanoreceptors [14, 16, 15]. A fair amount is known about the frequency response of these receptors to tightly defined inputs. The Merkel disks and the Meissner corpuscles are type I populations and are near the surface and closely spaced. These surface sensors are primarily sensitive to low frequency information. The Merkel disks respond to static and slow changes with a lowest threshold at about 10 Hz. The Meissner corpuscles respond only to changing signals with a lowest threshold at about 30 Hz. Most robot tactile arrays attempt to duplicate the spacing and capabilities of these two sensors.

The deep skin sensors, type II, are more widely spaced with a density of $20/\mathrm{cm}^2$ and a 10 mm receptive field. The Pacianian corpuscles have a sensitive region from 50 to 400 Hz. The Ruffini organs are directionally sensitive and sense skin stretch.

Johnson [17] provides a recent comprehensive review of the tactile sensing system and provides a working hypothesis for the role played by three of the four basic systems: the SAI, RA, and PC systems. The SAI system is the slowly adapting type I population, ending in the Merkel disks, and all the pathways conveying its signals to memory and perception. Similarly, the RA system is the rapidly adapting type I population ending in the Meissner corpuscles, and the PC system is the rapidly adapting type II population ending in the Pacianian corpuscles.

The SAI system is thought to be responsible for encoding low frequency, widely separated, spatial contact signals. Experiments with Braille and Roman letters show that the SAI system provides an isomorphic image of the contact signal. Therefore the SAI system is thought to be responsible for directly imaging the contact shape.

The encoding of texture is still being determined. It is clear that the perception of texture requires relative movement between the skin and the surface. For widely spaced textures, the SAI units are able to determine the spatial distribution. However for very fine textures and slip, vibration sensing seems to be critical. Both the RA and the PC systems appear to be involved.

Recent studies have shown that slip is detected, for textured surfaces, by these rapidly adapting fibers when a vibration is set-up by the relative motion of a textured surface [26]. The spatial distribution of the contact is of only secondary importance. Relative motion of very small raised dots (4 μm high, 550 μm diameter single dots, and 1 μm dots spaced at 100 μm center-to-center) on a smooth plate produced reliable activation of the fibers. The direction of slip is determined by the slowly adapting fibers which measure skin stretch. This suggests that the information about the onset of slip is carried by the high frequency component of the contact force, and that the direction of slip is carried by the direction of skin stretch.

In addition, textured surfaces can be differentiated by scanning a rigid probe across a surface. The probe encodes only temporal information. The induced vibration level seems to be used for discrimination. The PC system is the most likely candidate for the transduction system, since it is the only system excited by the impulses that occur during the initial placement of an object. Neurophysiological study of this mechanism for perception has just begun.

In summary, the SAI, RA, and PC seem to have a separation of function loosely based on frequency. This separation is used in perceiving the contact state while doing manipulation with a tool. The SAI system is tuned to determine the contact distribution between the hand and the tool. The PC system is able to detect the high frequency events that are transmitted down the tool to the hand and is insensitive to the shape of the tool. And the RA system has a lower spatial sensitivity then the SAI system but is more sensitive to vibration and so can determine local movements between the tool and the hand.

This work suggests that it is possible to extract information about texture and impacts with an intrinsic contact sensor. Like the PC system, the algorithms should look for high frequency events. In addition, the low frequency response from the contact sensor should be related to the neural encodings for joint torques. Finally, results developed for the temporal response from a single contact sensor, may be extendable to analyzing the temporal response from a sensor array.

4 Signal Models

In order to detect and label different events, models of the different signals had to be developed. We built two pieces of experimental apparatus in order to characterize the response of the system to impacts, sliding across a surface, no contact, and grasping contacts. Statistical process models were developed based on the experiments. These were captured as a set of signal source hypotheses. The next subsection described the experiments and the signal models, and the last subsection discusses the collection of the models into source hypotheses.

4.1 Signal Examples and Models

We considered four basic contact signals: impacts, slip, no contact, and grasping contacts. For each basic contact, an experiment was developed to isolate that particular event. Based on these experiments, a statistical signal model was developed by testing the description against the data. The assumption that these signals were basic and could be used to span the contact set

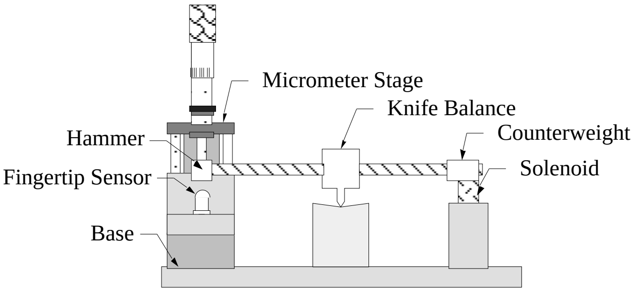

Figure 3: Impact Test Apparatus

was tested by considering some examples of more complex manipulation. This is discussed in the section 6.

4.1.1 Impact Model

To gather impact data, a solenoid released impact hammer was placed on a knife pivot (figure 3). The eight strain gauge signals resulting from the hammer striking the fingertip sensor was recorded at 2.7 kHz. The impact load could be varied by changing the initial height of the force sensor with the micrometer. A representative impact signal from one strain gauge and its corresponding Fast Fourier Transform (FFT) is shown in figure 4. There are a number of important characteristics in this signal. First, three strikes by the hammer are shown by the three sharp rises in the figure. Each of these strikes is separated by approximately 150 ms. Second, each of the sharp impact signals decays very quickly (approximately 1.5 ms). Finally, the sensor continues to ring for an extensive period, over 0.5 sec, after the impacts (not shown). Finally, the rapid vibration that follows each impact events is at approximately 1000 Hz.

As an initial model of the impact process consider the ideal model shown in figure 5. In this model, each impact of M_2 with M_1 is an inelastic collision with coefficient of restitution e. The time and distance of the collision are small compared to the travel time and distance of M_2. Under these assumptions, the process can be modeled as a series of impulses applied to M_1. Let x_i be the normalized height of M_2 after collision i with $x_0 = 1$; $\omega_0^2 = K_1/M_1$ be the natural frequency of the sensor; $\eta = \omega_0\sqrt{2h/g}$ be the natural unit of time, with h being the true initial height of M_2; t_i be the time of the ith collision with $t_1 = e\eta$; z be the normalized displacement of x_1; $\beta = M_2/M_1$; and ζ the damping coefficient which is less than 1. The equations of motion for the model are then:

$$\begin{aligned} x_{i+1} &= e^2 x_i \\ t_{i+1} &= t_i + 2e\eta\sqrt{x_i} \\ \ddot{z} + 2\zeta\dot{z} + z &= \frac{4\beta(1+e)}{\eta}\sum_{j=1}^{\infty}\delta(t-t_j)\sqrt{x_j} \end{aligned}$$

Using linearity, and the convolution properties of δ, the trajectory for M_1 is

$$z(t) = \frac{\beta(1+e)\eta}{\sqrt{1-\zeta^2}}\sum_{t_j \le t}\sqrt{x_j}\exp^{-\zeta(t-t_j)}\sin(\sqrt{1-\zeta^2}(t-t_j))$$

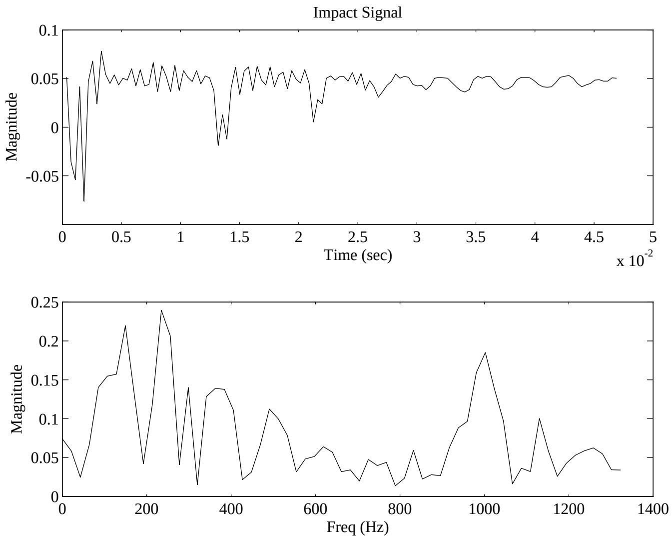

Figure 4: Representative impact signal from one strain gauge bridge. There are 128 samples taken at 2.7 kHz. The FFT was computed using a square window. The signal mean has been removed from the FFT.

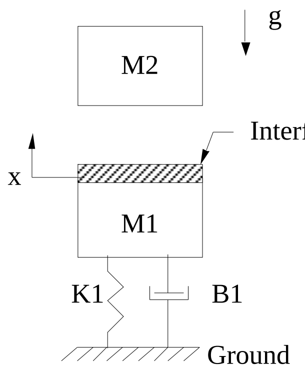

Figure 5: Ideal model of the impact process. An inelastic collision occurs at the boundary over time and distance that are small compared to the travel time and distance for the impacting mass.

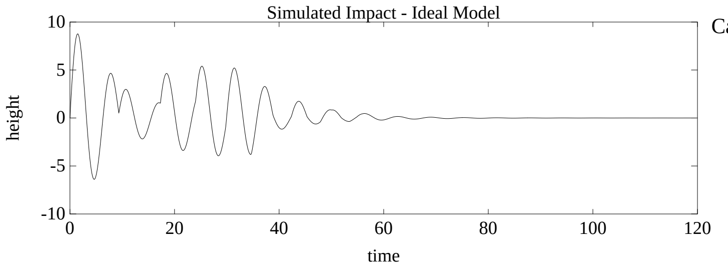

Figure 6: Simulation of model response using first 20 impacts with: $e = 0.85$, $\zeta = 0.1$, $\beta = 1$, $\eta = 5.5$

Simulations of this model result in some of the phenomena displayed in the actual data. Figure 6 shows the sudden increase in displacement, the slow exponential decay in the signal, and the discontinuous effect of secondary impacts. However, the simulation does not display the disorder seen in the actual signal, and is a poor match to the data. A more detailed model for this process treats the process as an initial condition response for a autoregressive (AR) model of unknown order. We were able to get good fits to each impact event with a fourth order auto-regressive model. An impact event lasts for about 50 samples. The first four values of the impact signal were used to create the initial conditions and the parameters of the autoregressive model were estimated using maximum likelihood (MLE) over the remaining samples.

The majority of the impact signal energy is in the first four samples. The remaining energy is in the extended ringing that follows the impact event. The energy in the initial part is captured in the model by using the first four values as initial conditions. The ringing is captured by the autoregressive coefficients. Unfortunately, the MLE estimator is nonlinear and uses all of the data. Therefore the estimates have to be generated through numerical search which make the model impractical for real-time implementation. Instead we adopted the simpler approach of computing a Karhunen-Loève (KL) expansion for the impact shape based on empirically segmented training data.

Seventy-two training example were generated using the impact apparatus. The start of the impact was found by looking for the first point in the signal that was twenty or more standard deviations away from the calibrated initial conditions. Forty points were then taken from this starting point to get seventy-two examples of length forty. The mean was removed from each sample individually, and then each sample was normalized to have unit energy. Normalization prevented any one signal from dominating the result and since the test is a ratio between different hypotheses the energy is not important. The Karhunen-Loève expansion, which is an eigenvalue expansion of the sample covariance, was then computed. The majority of the energy was contained in the first four eigenvalues; therefore, these were made additional features for the impact model. The features were extended beyond forty samples by adding zeros to the end of the feature signal. The residual after fitting the mean and four features was modeled as white normal noise. The resulting impact model for the measurement $y(t)$ is

$$y(t) = \sum_{k=1}^{4} \nu_k s_k(t) + \mu + e(t) \quad (1)$$

$$e(t) \sim N(0, V). \quad (2)$$

In this model, the measurement, y, is a mean μ plus a linear combination of four features s_k, which are treated as being orthogonal to μ, and a normal process noise $e(t)$. The shape of the impact signal and its residual vibration is captured by s_k, the new mean is captured by μ. In building this model, the implicit assumption is that the shape of the impacts is approximately the same. This turned out to be true only in the ideal case.

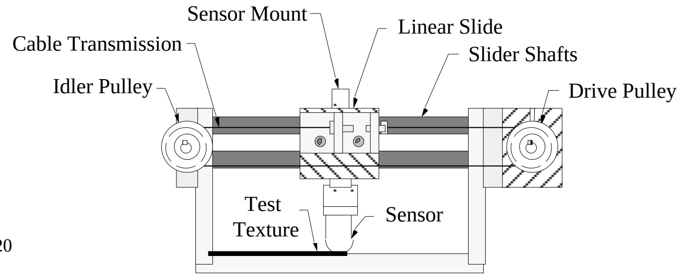

Figure 7: Slip Test Apparatus

4.1.2 Slip Model

Slip data was collected with the experimental apparatus shown in figure 7. The sensor was pulled along the base with constant normal load. The load could be changed by adding weights to the sensor mount. Pieces of sandpaper were used to control the surface roughness. These were placed part way along the base of the apparatus. A section of the signal generated from passing over the sandpaper is shown in figure 8. The FFT shows that the spectrum is essentially flat out to the bandwidth of the sensor at which point it rapidly decays. An approximate model is then given by a normal white noise process

$$y(t) \sim N(\mu, V).$$

This model was also confirmed by a histogram and the autocorrelation function.

4.1.3 No Contact Model

A sample of the strain gauge signal without any contact is shown in figure 9. Although the sensor is not being contacted, background vibrations are picked-up through the base. In this mode, the sensor is acting as an accelerometer. The FFT of the signal is again fairly flat, and the autocorrelation also indicates that the signal is approximately white. For small vibrations, quantization noise would be the dominant error which has a uniform distribution. However, the vibration levels are such that a normal model is a better fit to the data as indicated

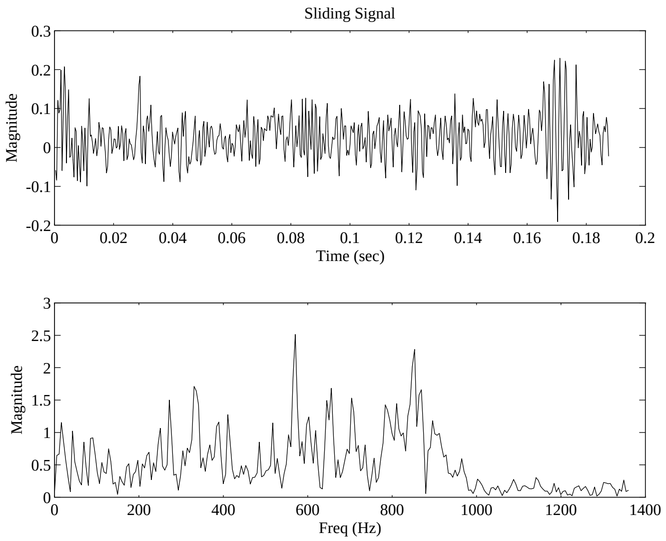

Figure 8: Representative section of signal generated by sliding over a piece of sandpaper. 512 samples taken at 2.7 kHz and the corresponding FFT using a square window. The signal is fairly flat out to the bandwidth of the sensor.

by the histogram. Therefore an appropriate model of the signal is

$$y(t) \sim N(\mu_0, V_0).$$

This model is distinguished by the particular values of μ_0 and V_0. To recognize this model these values must be calibrated.

4.1.4 Grasping Contact

Finally, the fingertip sensor was grasped by hand and forces were applied (figure 10). A slip free grasp was maintained by relying on human capabilities for detecting slip. The FFT of the signal shows a lowpass response, and an appropriate model is an autoregressive process

$$\begin{aligned} y(t) &= \sum_{j=1}^{p} a_j y(t-j) + e(t) \\ e(t) &\sim N(0, V). \end{aligned}$$

4.2 Hypothesis Models

The goal of this work is to segment any strain signal into pieces which can be identified with one of the experimentally selected models. Segmentation can be achieved by designing a decision algorithm which is tuned to our model set. Therefore we collected the parametric models

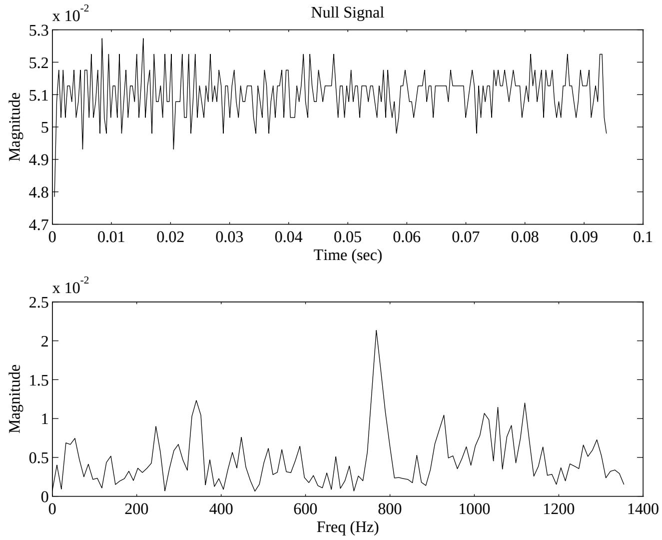

Figure 9: Representative rest signal from one strain gauge bridge sampled at 2.7 kHz and the corresponding FFT using a square window.

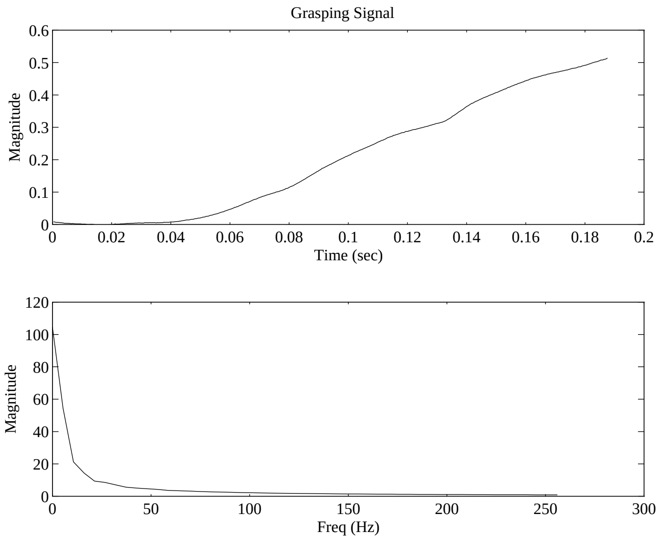

Figure 10: Representative grasp signal. 512 samples at 2.7 kHz and the corresponding FFT using a square window. The fingertip sensor was clamped in place and then loads were applied by grasping and twisting the sensor housing.

of the signal into six basic signal sources, or hypotheses:

Null State (H0):	$y(t) \sim N(\mu_0, V_0)$
New Mean (H1):	$y(t) \sim N(\mu, V_0)$
New Noise Level (H2):	$y(t) \sim N(\mu_0, V_1)$
New Mean and Noise Level (H3):	$y(t) \sim N(\mu_1, V_1)$
Impact Signal (H4):	$y(t) = \sum_{k=1}^{4} \nu_k s_k(t) + \mu + e(t)$ $e(t) \sim N(0, V_1)$
Grasping Signal (H5):	$y(t) = \sum_{j=1}^{p} a_j y(t-j) + e(t)$ $e(t) \sim N(0, V_1)$.

This set of source can be viewed as asking a set of questions about the statistics and spectral properties of the strain signal:

1. Is the signal white or is there significant correlation?
2. Has the mean of the driving noise changed?
3. Has the variance of the driving noise changed?
4. Is the mean and/or variance from the base case of no contact?
5. Are there any impact signatures in the time series?

A segmentation and identification procedure was derived based on these parametric models. The procedure is based on the generalized likelihood ratio test coupled with the minimum description length principle. This approach offers a number of advantages over more adhoc procedures. First, its model based. The decision procedure follows directly from the models described above. Second, it is an optimal procedure, within the context of the models, when such a procedure exists. Lastly, it explicitly estimates the time of events, which is a property that most filter based approaches lack.

The next section 5 presents the theoretic basis for our decision algorithm. It begins with a slightly more abstract form of the problem of statistical segmentation and identification in order to frame the discussion. The implicit assumptions about the measurement process used in the generalized likelihood ratio are then justified. The test is then presented with a sequence of examples each of which is interesting in its own. For example, the procedure for testing for a change in mean from a known value is the optimal guarded move sensing strategy.

Finally section 5.4 discusses the problem of labeling multiple parameterized models which is the problem presented by our set of hypotheses. In this case, the problem of uniformly penalizing the free parameters arises. We discuss the two basic approaches that have been used in the literature and present a justification for the choice of the minimum description length principle (MDL). Finally, the theory section concludes by presenting the algorithm in detailed form including a discussion of efficient parameter estimation algorithms. Results are presented in section 6. We applied the algorithm to both the training examples discussed above and on more general tasks.

5 Sequential Hypothesis Testing

In order to develop algorithms for processing dynamic contact information, we introduce a sequential hypothesis testing model of the sensing process. This area has been an active area of research in statistics and signal processing since its initial development by Wald [28]. A mathematical review is given by Siegmund [25]. There have been a number of important results during the last decade [3, 29]. These methods are relevant to any signal processing task which can be modeled as a stochastic measurement process on an underlying system which undergoes discontinuous changes. The methods are particularly useful when accurate and rapid decisions about the time of change are required. This includes edge detection, continuous speech segmentation, and dynamic contact sensing.

The most powerful hypothesis testing procedure, i.e. the one that uses the most prior information, is Bayesian decision theory on a semi-Markov chain. An approach using this model would estimate the probability of being in every node in the chain and would consider every possible sequence of transitions on the graph to explain the sequence of measurements. This can be computationally complex. In addition, this approach requires a prior probability distribution for the transition probabilities, the holding times, and any parameter values which are difficult to develop. An alternative procedure is the sequential likelihood ratio.

In sequential hypothesis testing it is assumed that the time for the algorithm to detect a transition is short compared to the holding time before a second transition. Therefore it is assumed: 1) that transitions can be detected by considering only the data, 2) that at any time only one hypothesis needs to be assumed to be true, and 3) only one transition from this hypothesis needs to be considered. These assumptions make the problem complexity at most linear in the number of samples.

The next subsection discusses the sequential hypothesis testing approach in general. Then, a set of important special cases is presented. First, the simple hypothesis testing problem of testing between known distribution is presented. This problem arises often in practice and is the easiest to treat theoretically. As an example, we show how easily contacts can be detected optimally, and demonstrate the increase in performance over an optimally designed filter and threshold approach. Then the procedure for two known signals in Gaussian noise is presented. The form of the computations for this test appears in all the more complex algorithms.

Next we consider changes between two parameterized distributions. As a special case we consider testing for an unknown change in mean. This problem is very important. Many problems can be reduced to this problem by appropriate preprocessing. In addition, all estimation procedures give rise to an asymptotically local procedure which looks like a test for change in mean [4].

Finally, we consider the problem of multiple parameterized distributions. This is the form of our dynamic sensing algorithm takes. The procedure requires a technique for uniformly penalizing the number of free parameters, and we adopt the minimum description length principle. This section concludes by presenting the algorithm in detail including a discussion of efficient parameter and model order estimation.

5.1 General Theory

In general, the measurement could be caused by any one of a set of m hypotheses (states) $\mathcal{H} = \{\mathrm{H}_i\}$. Each state provides a statistical description $\mathrm{p}_i(y(k), ..., y(l))$ of the measurement process. $y(l)$ is the first sensor signal generated from state i and $y(k)$ is the last. We consider only discrete measurements. As a series of sensor measurements $y_0^n = \{y(0), ..., y(n)\}$ are taken, the problem is to determine the sequence of states $x_0^n = \{x(0), ..., x(n)\}$ from which the measurements were taken. Since it is assumed that the time between events is sufficiently long for an algorithm to detect the transition, the algorithm can run forward in time, from the last detected event and state and look only for a single transition within the data under consideration.

Initially the algorithm is given that hypothesis H_p is true. Then, for every time r from 0 to n the likelihood [1] that the measurements were generated by H_p from time 0 to time $r-1$ and then by a different state H_q from time r to n is computed. This is compared to the assumption that all the data came from H_p. Because of the independence assumption of state transitions and the measurement densities, the likelihood of hypothesis p followed by hypothesis q is

$$L(p, q, r, y_0^n) = \mathrm{p}_p(y(r-1), ..., y(0))\mathrm{p}_q(y(n), ..., y(r)),$$

and the likelihood that all of the measurements came from state p is

$$L(p, y_0^n) = \mathrm{p}_p(y(n), ..., y(0)).$$

The optimal test is the likelihood ratio for the measurements

$$\frac{L(p, q, r, y_0^n)}{L(p, y_0^n)} = \frac{\mathrm{p}_p(y(r-1), ..., y(0))\mathrm{p}_q(y(n), ..., y(r))}{\mathrm{p}_p(y(n), ..., y(0))}.$$

The decision function is the maximum of the log of this ratio over the possible new states

$$DF(p, q, y_0^n) = \max_{r\in[0,n]} \log \frac{L(p, q, r, y_0^n)}{L(p, y_0^n)}.$$

The most likely new state is $\hat{q}$ which equals

$$\hat{q} = \arg\max_q DF(p, q, y_0^n).$$

This yields the test

$$DF(p, \hat{q}, y_0^n) \underset{\mathrm{H}_p}{\overset{\mathrm{H}_{\hat{q}}}{\gtrless}} T_p^2.$$

This rule says that $\mathrm{H}_{\hat{q}}$ will be chosen as the new state if $DF(p, \hat{q}, y_0^n)$ becomes larger than T_p^2, otherwise H_p will be maintained as the current hypothesis. T_p^2 is the decision threshold and is a design parameter that controls the essential trade-off between the two types of errors.

[1]The likelihood is the conditional probability of receiving the measurements given the hypothesis. It is related to the probability that the hypothesis is true through Bayes rule which required a prior distribution. The likelihood is used when the prior distributions are either unavailable, or assumed to be noninformative.

There are two important characteristics of this test: 1) the false alarm rate, 2) the delay to detection. The earliest time at which the decision function exceeds the threshold, given that the system is still in state p, is the false alarm time $t_f(p) = \inf(n : DF(p, \hat{q}, y_0^n) > T_p^2)$ which has distribution $\mathrm{Pr}_{FA}(n, p)$. The probability of no alarm at time n is $\mathrm{Pr}_{NA}(n, p) = 1 - \mathrm{Pr}_{FA}(n, p)$. The asymptotic false alarm rate is defined to be $f(p) = 1 - \lim_{k->\infty} \frac{\mathrm{Pr}_{NA}(k)}{\mathrm{Pr}_{NA}(k-1)}$. This reflects the rate at which false alarms will occur over the long-term. In contrast, the delay to detection is a transient performance measure. The delay to detection, given that a change to state q occurred at time 0, and the decision function is on state q, is $t_D(p, q) = \inf(n : DF(p, q, y_0^n) > T_p^2 | x(t) = \mathrm{H}_q)$. The distribution of $t_D(p, q)$ is $\mathrm{Pr}_D(n, p, q)$ and its expected value is $\bar{t}_D(p, q) = \sum_{t=0}^{\infty} t\,\mathrm{Pr}_D(t, p, q)$. Both statistics are controlled by T_p which is a design parameter. Increasing T_p decreases the false alarm rate and increases the time to detection. Determing both of these relationships requires solving a first passage problem. Closed form solutions to this type of problem are rare and difficult to derive. However, for some of the special cases considered below approximations can be determined and are presented.

5.2 Changes Between Two Known Distributions

The simplest and most important special case is detecting changes between two known, conditionally independent, probability distributions for a signal. This problem contains all of the essential features of the statistical decision procedure. More generally, many sequential decision problems can be treated by designing a prefilter which changes the problem into this binary testing problem.

Assume that $y(t)$ is an independent sequence with distribution $\mathrm{p}_0(y(t))$ under hypothesis 0 and $\mathrm{p}_1(y(t))$ under hypothesis 1. Further, assume that H_0 is the initial hypothesis. Because $y(t)$ is independent, the probability density of receiving a sequence of measurements y_k^j under either hypothesis, conditioned on the value of H_i is:

$$\mathrm{p}(y_k^j | \mathrm{H}_i) = \prod_{t=k}^{j} \mathrm{p}_i(y(t))$$

The likelihood ratio between the two hypotheses is

$$\mathcal{L}(0, 1, r, y_0^n) = \frac{\mathrm{p}(y_0^{r-1} | \mathrm{H}_0)\, \mathrm{p}(y_r^n | \mathrm{H}_1)}{\mathrm{p}(y_0^n | \mathrm{H}_0)} \tag{3}$$

$$= \prod_{i=r}^{n} \frac{\mathrm{p}_1(y(t))}{\mathrm{p}_0(y(t))}. \tag{4}$$

To simplify the calculations let $\gamma_0(t) = \log(\mathrm{p}_0(y(t)))$, $\gamma_1(t) = \log(\mathrm{p}_1(y(t)))$, $\Xi_{01}(t) = \gamma_1(t) - \gamma_0(t)$, and $S_k^j(0, 1) = \sum_{t=k}^{j} \Xi_{01}(t)$. Then the decision function for a change from state 0 to state 1 is

$$DF(0, 1, y_0^n) = \max_{r\in[0,n]} \log \mathcal{L}(0, 1, r, y_0^n)$$

which results in the binary rule

$$DF(0, 1, y_0^n) \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} T^2. \tag{5}$$

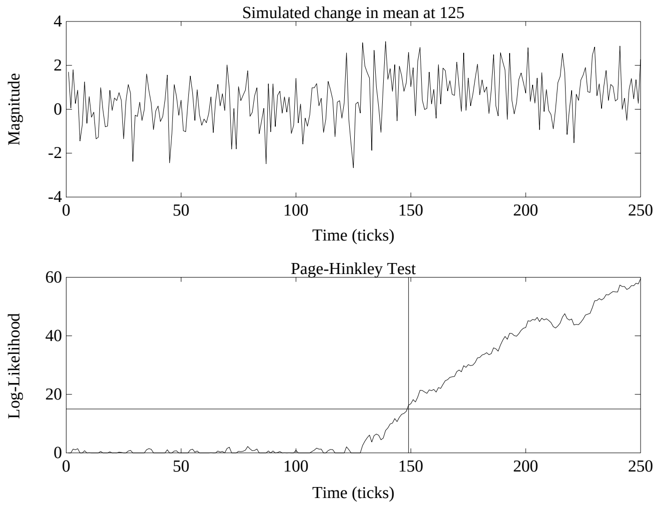

Figure 11: Behavior of the Page-Hinkley stopping rule to a simulated change in mean at tick 126 for a Gaussian process. Signal has standard deviation of 1 before and after the change, and mean of 1.0 after the change. Change is detected with a threshold of 15 at tick 149. The estimate of the time of change is the last time the test equals zero which is at tick 128.

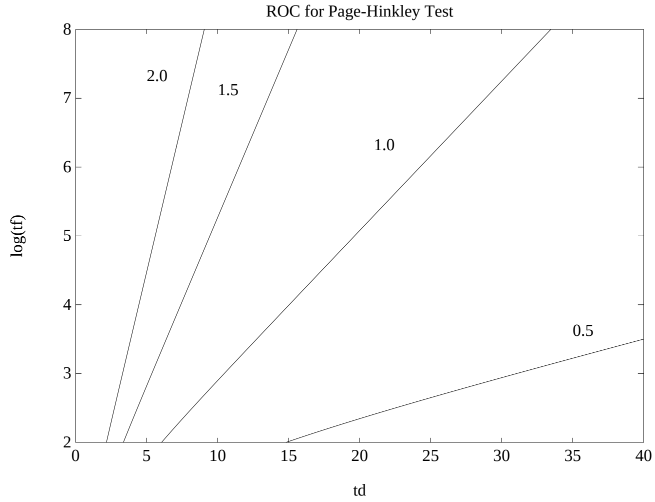

Figure 12: Receiver operating characteristic (ROC) of Page-Hinkley test between two Gaussian distributions with different means and the same variance as a function of the signal to noise ratio $s = \frac{\Delta\mu}{\sigma}$. The $\log_{10}(\bar{t}_f)$ is shown as a function of the mean time to detection $\bar{t}_d$ for s =0.5, 1.0, 1.5, and 2.0.

This is equivalent to the Page-Hinkley (PH) cumulative sum stopping test

$$DF(0,1,y_0^n) \quad = \quad S_0^n(0,1) - \min_{0\le j\le n} S_0^j(0,1).$$

This test minimizes the time taken to reach decision H_1 over all tests that have the same false alarm rate [25]. Further, it is easily computed recursively by

$$DF(0,1,y_0^n) = \max(0, DF(0,1,y_0^{n-1}) + \Xi_{01}(n)).$$

5.2.1 Change in Mean in Gaussian Noise

An application of the PH test to a change in mean in Gaussian noise is shown in figure 11. This figure was generated by adding a mean of 1.0 to a zero mean Gaussian random sequence with variance 1. With a change threshold of 15, the test detects the change at tick 149 for a delay of 24 ticks. The change time is estimated to be 128. In this particular case

$$\begin{aligned}\Xi_{01}(n) &= \frac{(\mu_1-\mu_0)}{V}(y(n) - \frac{\mu_1+\mu_0}{2}) \\ &= \frac{\Delta\mu}{V}(y-\bar{\mu})\end{aligned}$$

where V is the variance of the signal, and μ_i are the means under the two hypotheses. DF is simply the cumulative sum or integration, of Ξ_{01} with resetting at 0. The computation has the same computational cost as the alternative of lowpass filtering the signal and thresholding the result at the optimal value $\bar{\mu}$.

5.2.2 Comparison to the filtering approach

The performance of the PH test can be compared to lowpass filtering followed by thresholding on tests with equivalent false alarm rates. The asymptotic false alarm rate ${}^{PH}f$ and time to detection ${}^{PH}\bar{t}_D$ for the Page-Hinkley test can be approximated by applying Wald's identity and approximations [25]. The results are

$$\begin{aligned}{}^{PH}\bar{t}_F &\approx |e^{T^2} - T^2 - 1|/\beta_0 \\ {}^{PH}\bar{t}_D &\approx (e^{-T^2} + T^2 - 1)/\beta_1\end{aligned}$$

where

$$\beta_i = \int \log\left[\frac{\mathrm{p}_1(\zeta)}{\mathrm{p}_0(\zeta)}\right] \mathrm{p}_i(\zeta)d\zeta.$$

Since the false alarms are the interarrival times of a Bernoulli process they are geometrically distributed. Therefore the asymptotic false alarm rate is

$${}^{PH}f = \frac{1}{{}^{PH}\bar{t}_F}.$$

For the change in mean between two Gaussian processes with the same standard deviations σ, β_i is

$$\beta_i = 1/2\left(\frac{\Delta\mu}{\sigma}\right)^2.$$

A plot of the trade-off between the time to detection, $\bar{t}_d$, and the time to false alarm,$\bar{t}_f$ is called the receiver operating characteristic (ROC). It is a function of the signal-to-noise ratio $s = \frac{\Delta\mu}{\sigma}$. Graph 12 shows the value of $\bar{t}_d$ and $\log_{10}\bar{t}_f$ parameterized by T for a fixed value of s. The ROC for this test is shown in figure 12 for $s = 0.5, 1.0, 1.5, 2.0$. Both the mean time to a false alarm and decision increase with increasing threshold. At a fixed false alarm time, an increase in the signal to noise ratio will decrease the time to detection.

The performance of the alternative test of lowpass filtering followed by thresholding can be bounded using the following asymptotic approximation derived by Hall [10]. The approximations are valid in the limit of an increasing threshold and short sampling time. Consider a filter realized by a stable, linear, time invariant vector process x

$$x(k+1) = Ax(k) + w(k+1) + \Delta\mu u_{-1}(k-r)$$

driven by a white, zero-mean, Gaussian noise $w(k)$ with noise intensity Q. A change of size $\Delta\mu$ is applied by the unit step u_{-1} at time r. The covariance of x satisfies the discrete Lyapunov equation $S = ASA^T + Q$ and the decision function is $DF(k) = x^T(k)S^{-1}x(k)$. In principle it is possible to determine $\Pr_{FA}(k)$ by propagating the density for $x(k)$, $\mathrm{p}(x,k)$, forward in time and then integrating over the decision region. The propagation equation is

$$\mathrm{p}(x, k+1) = \int_D \mathrm{p}_w(x - A\zeta)p(\zeta, k)d\zeta$$

where $D = \{x : DF(k) \leq T^2\}$. Then $\Pr_{FA}(k)$ is given by

$$\Pr_{FA}(k) = 1 - \int_D \mathrm{p}(u,k)du.$$

Unfortunately there are no closed form solutions to this problem. However by treating the discrete system as a sampling of a continuous system, an approximation valid for large k can be determined. Using this approximation, the steady state false alarm rate f is found to be asymptotically bounded by

$$f \leq 1 - \exp\left(\frac{\ln(\det(A))T^p}{\Gamma(p/2+1)}\exp^{-T^2/2}(1 - p/T^2)\right)$$

where p is the dimension of x. In the case of a first-order lag filter $x(k+1) = ax(k) + w(k)$, the bound is

$$f_0 \leq 1 - \exp\left(\sqrt{\pi/2}\ln(a)T\exp^{-T^2/2}(1 - 1/T^2)\right).$$

This is the bound for $x^2/S > T$. The PH test is equivalent to $X/S^{1/2} > T$ which has a false-alarm rate bounded by $f_0/2$.

To approximate $\Pr_D(k)$ note that $DF(k)$ is a noncentral chi-squared random variable with p degrees of freedom and noncentrality parameter $\delta^2(k) = \bar{x}^T(k)S^{-1}\bar{x}(k)$ [2]. The process mean $\bar{x}$ satisfies

$$\bar{x}(k+1) = A\bar{x}(k) + \Delta\mu$$

with initial condition $\bar{x}(0) = 0$ for a change in mean of $\Delta\mu$, where we have assumed for simplicity $r = 0$. If the cumulative noncentral Chi-square distribution of DF at value T^2 is denoted by $F(T^2, \delta^2, p)$, then $\Pr_D(k)$ is bounded by

$$\Pr_D(k) \geq 1 - F(T^2, \delta^2, p)$$

which can be computed numerically or approximated.

For a scalar, first-order lag-filter, the ROC can be computed as a function of the the signal-to-noise ratio s as in the PH test. In this case, the values of $\bar{t}_d$ and $\log_{10}\bar{t}_f$ are parameterized by a. The optimal threshold for the test is $\frac{\Delta\mu^2}{4S}$ where $S = \frac{(1-a)}{(1+a)}\sigma^2$. This gives a

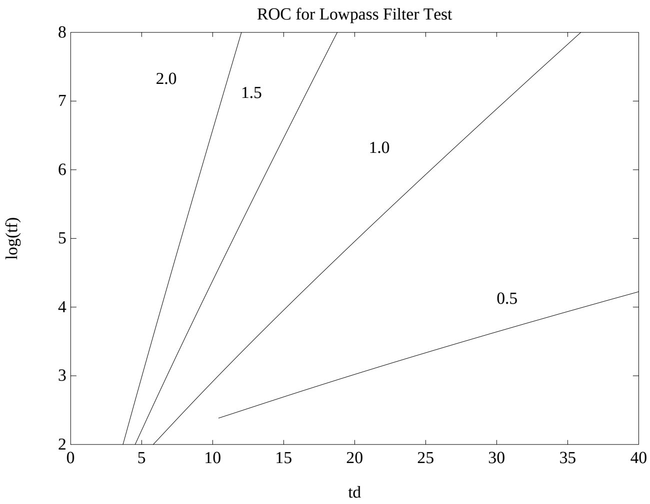

Figure 13: Receiver operating characteristic (ROC) of first order lag filter test with threshold between two Gaussian distributions with different means and the same variance as a function of the signal to noise ratio $s = \frac{\Delta\mu}{\sigma}$. The $\log_{10}(\bar{t}_f)$ is shown as a function of the mean time to detection $\bar{t}_d$.

threshold of $T^2 = \left(\frac{s}{2}\right)^2 \frac{(1-a)}{(1+a)}$. With the one-sided test, an approximation for $\Pr_D(k)$ is simply the probability of drawing a value greater that $\Delta\mu/2$ from a Gaussian random sample which has mean $\bar{x}(k)$ and variance S, given that the test has not already terminated. The probability of terminating at time k given that the test has not already terminated is

$$F(k) = 1 - \mathrm{erf}\left(\frac{s}{2}\sqrt{\frac{1-a}{1+a}}\right).$$

The probability of terminating at time k is then given by the recursion

$$\begin{aligned} P_D(0) &= F(0) \\ P_D(k) &= F(k)(1 - P_D(k-1)). \end{aligned}$$

This gives an underestimate of the termination time. An overestimate is given by the rise time for $\bar{x}(k)$ to $\Delta\mu/2$. Figure 13 shows the logarithm of $\bar{t}_f$ as a function of $\bar{t}_d$ for a signal-to-noise ratio of $s = 0.5$, 1.0, 1.5, and 2 computed using these two approximations. The curve for $s = 0.5$ has been cut short, because the approximation is not valid for small $\bar{t}_d$.

Figure 14 indicates that the lowpass filter approach has a longer delay to detection compared to the PH test when they have the same false alarm rate. The test shown in figure 11 will signal an alarm on average every 6×10^6 samples and the change will be detected after 28 samples. To get equivalent performance from the lowpass filter, a must equal 0.98. With this value, the estimate of $\bar{t}_D$ is 29.5 and the rise time is 34.5. These results demonstrate that the PH test gives an improvement in performance without an increase in computational cost. In addition, an estimate of the change time is possible with a small amount of additional storage.

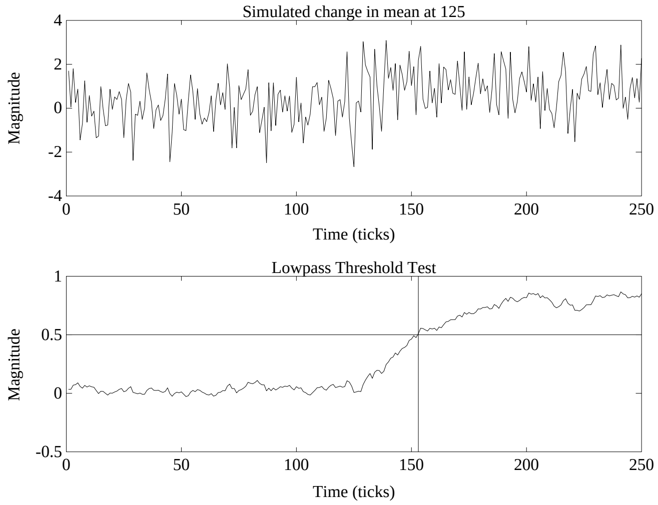

Figure 14: Lowpass filter of $x(n+1) = 0.98x(n) + 0.02y(n+1)$ on the same signal as figure 4. The threshold is 0.50. The change is detected at 153. This is a slower response then the response for the PH test. Further, an estimate of the change time is not computed.

5.2.3 Change between Deterministic Signals in Additive Gaussian Noise

The sequential hypothesis approach can be extended to derive tests for characteristics that are detectable with a single linear filter. A direct extension to the binary problem is given by the hypotheses:

$$\begin{array}{llll} \mathrm{H}_0: & y(t) & = & s_0(t) + v_0(t) \qquad t = 0, \ldots, n \\ \mathrm{H}_1: & y(t) & = & s_1(t) + v_1(t) \end{array}$$

where s_i are known deterministic signals. A test equivalent to the binary case results with $\gamma_i(t)$ redefined as

$$\gamma_i(t, r) = \log(\mathrm{p}_{v,i}(y(t) - s_i(t-r)))$$

where $\mathrm{p}_{v,i}$ are the known probability densities for the noise process. In this case, $DF(0, 1, y_0^n)$ cannot be computed recursively. Instead, the maximization must be computed by exhaustive search over a growing window of length n. To limit the increase in computational complexity, a suboptimal approach is usually taken where the search window is constrained to a moving window of fixed length.

Because of the time invariance of s_i, the complexity of the calculation of $S_r^n(0,1)$ is order Lm where L is the number of points in the window and m is the cost of computing $\Xi_{01}(n, r)$ for a single point. To see this, suppose that $S_r^n(0,1)$ has been computed over the window and stored in a vector $\bar{S}_{01}(n)$ of size L. When data point $y(n+1)$ arrives $\Xi_{01}(t, r)$ must be computed for every r and t in the window and then summed. However, the sum of $\Xi_{01}(t, r)$ for r and t less than $n+1$ has already been computed and stored in $\bar{S}_{01}$. Thus by shifting $\bar{S}_{01}(n)$ by one and then adding $\Xi_{01}(n+1, r)$ to the shifted result we obtain $\bar{S}(n+1)$. This requires order Lm operations. With a vector processor with L elements, the entire calculation can be done in parallel in m operations. $\bar{S}_{01}(n+1)$ must then be searched for the maximum element. This structure for the calculations is preserved in the more general case of an unknown change magnitude in Gaussian noise, which is considered in the next section.

5.3 Unknown Parameterized Distributions

In most applications the magnitude of the change is unknown. For this problem, there are two probability models for $y(t)$

$$\begin{array}{lll} \mathrm{H}_0: & y(t) & \mathrm{p}_0(y(t), \theta_0) \\ \mathrm{H}_1: & y(t) & \mathrm{p}_1(y(t), \theta_1) \end{array}$$

where θ_0 is a known parameter vector and θ_1 is unknown. θ_0 is an element of $\Re^{q_0}$, θ_1 is element of $\Re^{q_1}$, and y is an element of $\Re^p$. The probability densities p_0 and p_1 may or may not be from the same family. If the dimension and interpretation of θ_0 and θ_1 are the same, one approach for detecting the change is to choose a minimum change value $\delta\theta > 0$ *a priori* and then run two Page-Hinkley tests in parallel with $\theta_1 = \theta_0 \pm \delta\theta$. The test which terminates first is used as the decision rule. Changes of parameter greater than $\delta\theta$ will trigger a decision, but the delay to decision will be longer than with the correct $\delta\theta$.

One approach to choosing a value for $\delta\theta$ is to select a distribution for θ_1 *a priori* . In this case the measurement distribution, under H_1, is the convolution

$$\mathrm{p}_1(y(t)|\mathrm{H}_1) = \mathrm{p}_1(y(t), \theta_1) \otimes \mathrm{p}_{\theta_1}.$$

If θ_1 is treated as a sequence of random variables drawn from p_{θ_1}, then the Page-Hinkley test can be applied as above.

The problem of choosing a distribution for θ_1 can be eliminated by maximizing $\mathcal{L}$ over both θ_1 and r. This yields the generalized likelihood ratio (GLR) test [30]

$$DF(0, 1, y_0^n) = \max_{0 \le r \le n - q_1} \max_{\theta_1} S_r^n(0, 1).$$

There is a delay of at least q_1 samples before detection so that θ_1 can be estimated. If θ_0 is also unknown the test includes one more maximization.

5.3.1 Unknown Change in Mean

As an example of this problem, suppose p_i are both Gaussian densities with $\theta_0 = (\mu_0, V_0)$ the known mean and variance before the change. For simplicity assume $\mu_0 = 0$. After the change the process has a new mean μ_1 but the same variance V_0. Then Ξ_{01} is

$$\Xi_{01}(\mu_1, k) = \mu_1^T V_0^{-1} y(k) - \frac{1}{2} \mu_1^T V_0^{-1} \mu_1.$$

The maximization over μ_1 yields the usual sample mean estimate for μ_1 from r to n

$$\hat{\mu}_1(r, n) = \frac{1}{(n - r + 1)} \sum_{k=r}^{n} y(k),$$

and the decision function becomes

$$\begin{aligned} DF(0, 1, y_0^n) &= \max_{r \in [0, n-1]} \sum_{k=r}^{n} \Xi_{01}(\hat{\mu}_1(r, n), k) \\ &= \max_{r \in [0, n-1]} \frac{(n - r + 1)}{2} \hat{\mu}_1^T(r, n) V_0^{-1} \hat{\mu}_1(r, n). \end{aligned}$$

Again, the maximization is limited to a moving window of size L since it must be solved by exhaustive search.

Under hypothesis 0, DF is a central chi-squared random variable with $p/2$ degrees-of-freedom $DF \sim \mathcal{X}^2(p/2)$ which has mean $p/2$. Therefore, the threshold T must be at least $p/2$. Since under hypothesis 1, DF is noncentral chi-squared with p freedoms and noncentrality parameter $\frac{1}{2}\mu_1^T V_0^{-1}\mu_1$, the detection time can be approximated by the results in section 5.2.2.

The test computation can be computed in order Lm, or in order m steps by a vector processor of size L. At time n the value of $\hat{\mu}_1(r,n)$ is stored for every point in the window in the array $M(n)$. Each column of M is the estimate of μ_1 involving the final j measurements. After receiving $y(n+1)$, all of the columns of M are shifted right by 1 and then each column j is updated by

$$M(j, n+1) = M^*(j,n) + (y(n+1) - M^*(j,n))/j$$

where M^* indicates the shift operation. The decision function becomes

$$DF(0,1,y_0^n) = \max_{1\le j\le L} \frac{j}{2} M(j,n)^T V_0^{-1} M(j,n)$$

with $\hat{r} = n - j_{max} + 1$.

5.4 Multiple Parameterized Models

In the most general case each hypothesis H_p in the collection $\mathcal{H}$ could be parameterized by a different number of free parameters θ_p. This is the form that our dynamic sensing algorithm takes. The decision function is still the likelihood ratio but now with a maximization over three different parameter sets

$$DF(p,q,y_0^n) = \max_r \min_{\theta_p} \max_{\theta_p',\theta_q} \log \frac{L(p,q,r,\theta_p',\theta_q)}{L(p,r,\theta_p)},$$

where θ_p' is the value of the parameters for H_p for the data from 0 to $r-1$. Again, the maximization over r is explicit and so only a maximum number of moving windows are kept.

Two additional issues arise in this problem. First, models with more free parameters have greater explanatory power. That is they are able to fit finite length noise signals better than models with fewer free parameters. Therefore, a method of penalizing the number of free parameters which corrects this problem is required. The next subsection discusses this issue.

Second, the autoregressive model hypothesis, hypothesis 5, does not use a fixed number of parameters, and the parameter estimates depend upon the order of the model. Therefore, an efficient computationally technique for estimating the model parameters for all model orders is needed. This is provided by estimation of the reflection coefficients which is discussed in section 5.4.2.

Finally, section 5.4.3 gives the detailed procedure for our algorithm. This includes initialization, choice of weighting for the number of parameters, the estimation procedures, and the decision procedure.

5.4.1 Model Penalty

A problem arises with the unadjusted likelihood ratio test When the test involves models with different numbers of free parameters. Models with more free parameters, or degrees-of-freedom, have more explanatory power. That is they are able to fit finite lengths of data better than models with fewer degrees-of-freedom. Therefore, the model with the greatest number of parameters will always be the most likely. In order to correct this problem, a uniform method of penalizing the extra freedoms is required. Note that even if all the models had the same number of freedoms, the change hypothesis has twice as many freedoms as the recursive model because because it uses two models.

This problem is ubiquitous to likelihood approaches which involved comparisons between models with different numbers of freedoms. This problem is eliminated in the Bayesian approach by the *a priori* choice of prior probabilities for the number of models and parameters. The log of these prior probabilities adds to the likelihood calculation and essential creates an *a priori* model cost. What the likelihood approach requires is a principle for choosing these model costs.

The minimum description length principle [22, 21], or MDL, is one of the few general principles for choosing the model cost. This principle states that the cost of the model is related to the number of parameters or bits it takes to encode the model. In general, simpler explanations are preferred, therefore the likelihood should be reduced by the complexity of coding the model. For a number of linear parameter estimation problems, Rissanen has shown that the model cost, defined in terms of description length, is asymptotically $\frac{k}{2}\log n$ where k is the number of parameters and n is the number of samples used in the estimation.

An equivalent result was derived by Schwarz [23]. Schwarz derived an asymptotic expansion to the optimal Bayes procedure by assuming a fixed error cost and fairly general conditions for the prior distribution and the sample distributions. Under his assumption the Bayes procedure chooses the *a posteriori* most probable model and parameter values. Asymptotically this is equivalent to maximizing the log-likelihood minus $\frac{k}{2}\log n$. For autoregressive (AR) processes this procedure was shown to be strongly consistent (asymptotically) by Hemerly and Davis [11].

No such result exists for the still popular, and useful, alternative AIC procedure derived by Akaike [1]. Akaike suggested that the decision criterion should be based on maximizing the expected value of the log-likelihood. When all the models use the same underlying probability distribution, the penalty becomes the difference in expected log-likelihood between two competing models when the actual signal is white noise. For parameter estimation problems, this gives a penalty equal to the number of free parameters. For AR processes, Shibata [24] derived the distribution of the number of free parameters estimate, $\hat{k}$, using this penalty. The result shows that the most probable $\hat{k}$ equals k but that the expected value of k is generally closer to $k+1$.

All of these results are asymptotic and so for small sample sizes, with which all our tests work, there is some freedom in choosing the penalty. Because of the stronger theoretical justification of the MDL criterion, our algorithms utilizes a penalty of the form $b\frac{k}{2}log(n)$ where b is a unit parameter cost. In addition, since our procedures use data sizes of between 20 and 40 samples for the moving window the MDL and AIC criterion give nearly equivalent penalties. Values of b larger than one favor simpler hypotheses more heavily than the MDL criterion. We obtained subjectively good results with b from 2.2 to 2.7.

5.4.2 Linear Models and Orthogonal Estimation

This section discusses the efficient computation of the order and parameter estimates for linear predictor models in Gaussian white noise. All of the models used in our dynamic sensing algorithm fit the framework discussed here.

The general form of a linear predictor model is:

$$y(n) = \psi^T(n)\theta + e(n)$$

where $\psi^T(n)$ is a vector of regression coefficients, θ is the parameter vector, and e is a white Gaussian noise process. For example, for the new mean model H1 $\psi^T(n) = 1$. For the AR model with m free coefficients $\psi^T(n) = [y(n-1), ..., y(n-m)]$.

For the linear predictor model the maximum likelihood estimate of the parameters is the least-squares estimate. The least-squares estimate can be written in matrix form by collecting the measurements into a vector as $Y(n) = y_1^n$, and collecting each element of ψ^T into a vector as $s_i(n) = \psi(i)_1^n$. Then the least-squares estimate takes the form

$$\begin{aligned}
\mathcal{I}_{ij}(n) &= \langle s_i(n)|s_j(n)\rangle \\
\mathcal{I}(n) &= \sum_{t=1}^{n} \psi(t)\psi^T(t) \\
\mathcal{X}_i(n) &= \langle s_i(n)|Y(n)\rangle \\
\mathcal{X}(n) &= \sum_{t=1}^{n} \psi^T(t)y(t) \\
\hat{\theta}(n) &= \mathcal{I}^{-1}(n)\mathcal{X}(n),
\end{aligned}$$

where $\langle|\rangle$ denotes inner product. $\mathcal{I}$ is called the empirical information matrix and $\mathcal{X}$ is called the empirical information vector. $\hat{\theta}(n)$ is the parameter estimate at time n.

The parameter estimates depend upon the model order because in general the vectors $\{s_j(n)\}$ are not orthogonal. An efficient estimation procedure for the model order can be performed by first orthogonalizing the vectors $\{s_j(n)\}$ and then reconstructing the estimate for any desired order. The orthogonal parameter estimates are called the the reflection coefficients because they are related to the amount of energy reflected back at a changing impedance junction in an electrical transmission line [19].

When the computation is performed on-line only $\mathcal{I}(n)$ and $\mathcal{X}(n)$ are actual kept in memory. Therefore, the reflection coefficients need to be computed from these matrices. If the vectors $\{s_j(n)\}$ were actually available, a Gram-Schmidt procedure could be applied to the vectors to generate an orthogonal basis. This decomposition represents the matrix $S = [s_i]$ as $S = QD^{1/2}R$ where Q is the orthonormal basis, $D^{1/2}$ is the diagonal matrix of lengths for the orthogonal vectors, and R is the correlation structure of S. The time dependence has been dropped for clarity. Then, the reflection coefficients, K, satisfy $D^{1/2}K = Q^TY$.

Now note that $\mathcal{I} = S^TS = R^TDR$, since Q is orthonormal, and $\mathcal{X} = S^TY = R^TD^{1/2}Q^TY$. Therefore the reflection coefficients are generated by the first stage of Gaussian elimination. Gaussian elimination factors $\mathcal{I}$ into R^TDR. Therefore, the reflection coefficients are the solution to $R^TDK = \mathcal{X}$. This solution is just the result of the first stage of Gaussian elimination. The second step, back substitution, solves $R\hat{\theta} = K$.

Now the reflection coefficients can be used to reconstruct the solution for any model order. Given any order model m, let the first m reflection coefficients be K_m and the corresponding submatrix of R be R_m. Then the original model coefficients for an order m model can be determined from $K_m = R_m\hat{\theta}_m$.

More importantly, the reflection coefficients can be used to immediately determine the optimal model order. Let $E_0 = \langle Y|Y\rangle$ be the original signal energy. Then because of orthogonality, the energy remaining after using the m^{th} reflection coefficients is $E_m = E_{m-1} - k_i^2/D_{m,m}$, and the adjusted log-likelihood of this model is $l(Y, m) = -0.5n(\log(E_m/n)+1) - b\frac{(m+1)}{2}\log(n)$. The model order which maximizes $l(Y, m)$ is optimal under this criterion.

Although the computations can be reorganized for better numerical performance, this is the essence of the ladder or lattice algorithm. In summary the estimation procedure stores $\mathcal{I}$, $\mathcal{X}$, and the energy in the input signal E_0 and then performs the following update procedure:

1. Update the regression vector ψ based on the model equation.
2. Update E_0.
$$E_0 \leftarrow E_0 + y(n)^Ty(n)$$
3. Update the information vector and matrix.
$$\begin{aligned}
\mathcal{I} &\leftarrow \mathcal{I} + \psi^T\psi \\
\mathcal{X} &\leftarrow \mathcal{X} + \psi^Ty(n)
\end{aligned}$$
4. Decompose $\mathcal{I}$ into $\mathcal{I} = R^TDR$.
5. Solve for the reflection coefficients.
$$R^TDK = \mathcal{X}$$
6. Determine the optimal model order m^* by maximizing the adjusted log-likelihood $l(m)$.
$$\begin{aligned}
E_m &\leftarrow E_{m-1} - k_i^2 * D_{m,m} \\
l(m) &\leftarrow -0.5n(\log(E_m/n)+1) - b\frac{(m+1)}{2}\log(n)
\end{aligned}$$
7. Solve for the optimal model parameters.
$$\hat{\theta}_m* \leftarrow R^{-1}K_m*$$

When the number of model parameters is fixed, the fast-gain algorithm can be used to update the parameter estimates with a fewer number of computations [18] but the essentials are the same.

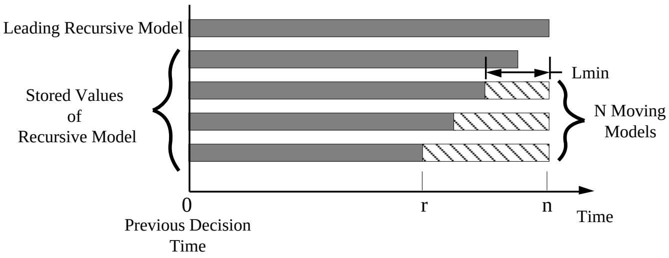

Figure 15: Generalized Likelihood segmentation procedure with multiple hypotheses. A single recursive model and a collection of moving models must be updated for every new measurement.

5.4.3 Computational Organization

The sequential nature of the test leads to an organization of the computations that minimizes the number of computations. A conceptual picture of the computations is shown in figure 15. At any stage in the computation there is one hypothesis being grown recursively and a collection of L moving window models for every alternative hypothesis. A minimum length L_{min} is chosen for the moving windows based on estimation requirements.

Since all of the hypotheses have linear parameters in Gaussian noise, the least squares estimate of the parameters is the maximum likelihood estimate. Therefore, for every window w, both recursive and moving, and hypothesis p we store $D_p(w)$ which contains: 1) the current value of the empirical information matrix $\mathcal{I}_p(w)$, 2) the current value of the empirical information vector $\mathcal{X}_p(w)$, 3) the current value of the parameters $\hat{\theta}_p(w)$, and 4) the current value of the adjusted log-likelihood $l_p(w)$.

All of the moving windows for each hypothesis are collected into a structure $\mathcal{M}_p(n) = \{D_p(1), ..., D_p(lmax)\}$. For window lengths less than $lmin$, the parameters are not estimated. When a new data point $y(n+1)$ is measured the moving windows are updated by:

1. The values in $\mathcal{M}_p$ are shifted by 1 for every hypothesis.
2. $\mathcal{X}_p(w)$, $\mathcal{I}_p(w)$, $\hat{\theta}_p(w)$, and $l_p(w)$ are updated for every window using the algorithm described in section 5.4.2.

Let the current hypothesis by H*. For this current hypothesis a collection of recursive windows $\mathcal{R} = \{D^*(1), ..., D^*(lmax)\}$ is stored. In this case each $D(i)$ structure contains the information vector and matrix based on data from 0 to $n-i+1$. When a new data point $y(n+1)$ is measured the recursive windows are updated by:

1. The values in $\mathcal{R}$ are shifted by 1.
2. The values in $D^*(1)$ are updated using the algorithm described in section 5.4.2 except now the model order is fixed.

After updating all the moving windows and the recursive window, the algorithm examines the results to detect a change. The likelihood of the current recursive model $D^*(1)$ plus the threshold T_p^2 is compared to the sum of the likelihood for all possible pairs of moving models with the matching recursive model. If the change likelihood is greater a change is detected.

Several steps are necessary to update the computation after detecting a change. First the change time is taken as the new time origin. Then, the number of moving models is adjusted so that only models that lie entirely after the detected change time are kept. Finally, recursive models starting at time r, with lengths from L_{min} to $n-r+1$ must be recomputed in a batch mode. The computation then proceeds from this adjusted state.

To initialize the computation, each hypothesis is evaluated on a minimum length window L_{min} in a batch mode. The optimal hypothesis is computed by maximizing the likelihoods after adjusting for a model cost. The maximal hypothesis is then stored as a single recursive model.

This computation produces at every time t, lagged by $L_{max} = L_{min} + L - 1$, a hypothesis model, its associated parameter vector, an estimate of y, and the model's adjusted likelihood. In addition, a sequence of event marks are produced as each change in model is detected. If the hypothesis models are chosen well, the output of the algorithm is essential a sequence of symbols for the data which can be used for further task recognition processing.

6 Experiments and Results

The algorithms developed in the previous section were applied to segmenting the contact signal hypotheses developed in section 4 on both the selected data samples and more general examples. In addition, the algorithms were simplified and specialized to particular problems to show their utility in detecting important but minute changes in real-time. The results should be evaluated in terms of three criteria: 1) do the algorithms divide the signal into segments that make sense, 2) are the hypotheses identified by this context free algorithm the ones we expect, and 3) could the results be interpreted given a context. The results show that the algorithms do a good job of segmenting the signal, but that a context free interpretation is very difficult. However, by fixing the context for simple cases, robust interpretations should be possible.

6.1 Performance on design set

The general multi-model generalized likelihood ratio test was applied to the set of data used for developing the model. There were six basic models

Null State (H0):	$y(t) \sim N(\mu_0, V_0)$
New Mean (H1):	$y(t) \sim N(\mu, V_0)$
New Noise Level (H2):	$y(t) \sim N(\mu_0, V_1)$
New Mean and Noise Level (H3):	$y(t) \sim N(\mu_1, V_1)$
Impact Signal (H4):	$y(t) = \sum_{k=1}^{4} \nu_k s_k(t) + \mu + e(t)$ $e(t) \sim N(0, V_1)$
Grasping Signal (H5):	$y(t) = \sum_{j=1}^{p} a_j y(t-j) + e(t)$ $e(t) \sim N(0, V_1)$.

For each model, moving windows with 15 to 40 samples, inclusively, were used. The number of free parame-

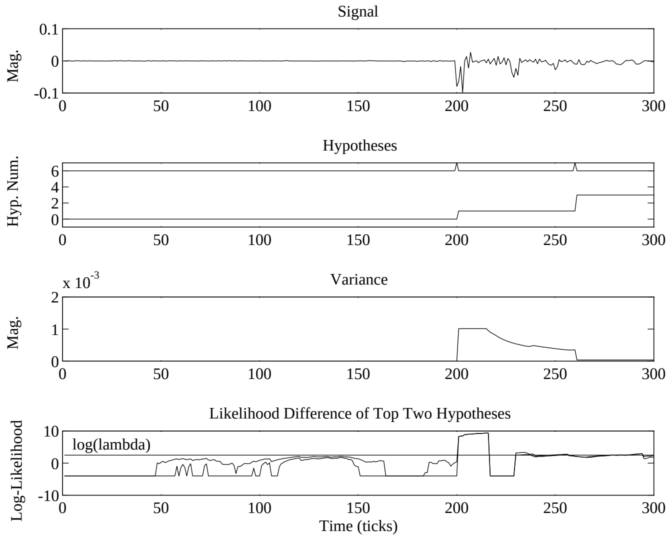

Figure 16: Impact Signal incorrectly marked at tick 200.

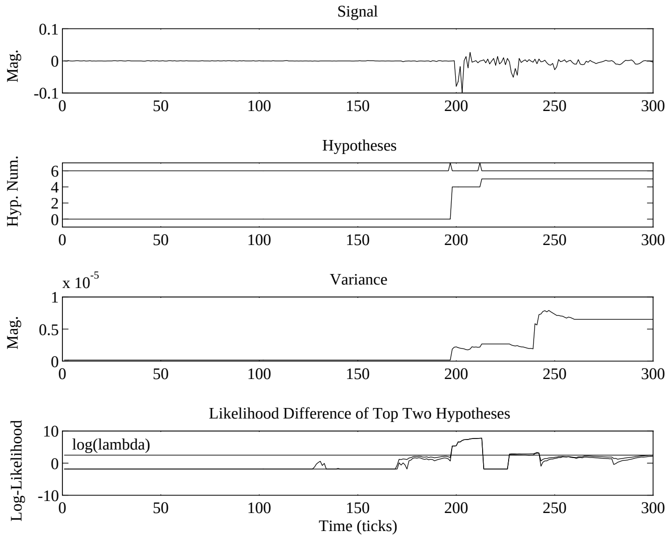

Figure 17: Impact Signal correctly marked.

ters was penalized using the MDL criterion with $b = 2.2$. In addition, impacts cause such large changes in variance that changes were signaled when the first data point of an impact entered a moving window. Therefore, a persistence test of 20 was necessary to get the correct marking of the change time. The length of minimum window was chosen based on the length required for estimation, while the maximum was chosen based on the size of the persistence test. The decision threshold was set at 12.

6.2 Performance on task examples

6.2.1 Impacts

The training set for impacts was used to get a measurement of performance. The classification for 72 examples was:

H0	H1	H2	H3	H4	H5
0	25	0	0	40	7

The results show that impact hypotheses (H4) and the change in variance hypothesis (H1) are difficult to distinguish and suggests the need to consider context. Selection of H4 depends upon this signal having some consistency from sample to sample. This is only occasionally the case, and so the simpler explanation H1 is frequently chosen. In real world impact situations, this problem resulted in confusion of the impact model (H4) with the change in new mean and noise model (H3) and the grasping signal model (H5). In the case of H5, a set of high frequency poles was estimated.

However, the introduction of context should make distinguishing the hypotheses much easier. An example of a typical unsuccessful marking is shown in figure 16. The top plot shows the signal, the next plot shows the sequence of hypotheses and the transition marks, the third plot shows the variance estimate, and the bottom plot shows the log of the loglikelihood for the top two hypotheses. Estimates of the coefficients for each of the models is also produced by the procedure, but is not shown. The figure shows than in unsuccessful marking the variance estimate has a sudden increase followed by a sudden decrease. It also shows that the likelihood indicates an event very clearly, at the impact time, and that the two top hypotheses, which includes the impact hypothesis, are very close during the impact. It should be possible to use this additional information to add context and get more accurate markings. Finally, the residual vibration after the impact is matched to hypothesis 3 because of the change in mean caused by the hammer resting on the sensor. A successful marking is shown in figure 17. The impact signal is modeled as H4 and the residual after the impact is modeled with H5. Again, H1 and H4 are close during the impact event.

The transition marks are generally close to the actual transition. The transition time error for the training set was:

Distance to transition	0	1	2	3	>3
	48	12	2	1	9

The transition times are within three 87% of the time. Errors larger than 3 were caused by small transitions in the signal that occurred just before the impact. In this case, the persistence test prevents a second change at the impact location and thus causes a large error.

The performance on the training examples indicate that impacts are very difficult to detect in a context free form. They are easily confused with jumps in the driving variance for the other models. This is confirmed in a more general impact experiments. To get a sense of performance for general manipulation, the fingertip was held by hand and then struck against an aluminum block. One hundred and three examples were taken, and the impact point was selected by looking for the first outlier that was 20 or more standard deviations away from the initial signal. The initial part of each of these signals is much more variable and the impacts are usually followed by a large change in mean from the sustained contact. For these examples the classification was:

H0	H1	H2	H3	H4	H5
0	37	0	55	1	10

This clearly shows that impacts are confused with H3 and H1. However, the segmentation times are still very good with 93% of the times falling within 3 of the correct time.

Distance to transition	0	1	2	3	> 3
	75	14	5	2	7

Even in the case of an error, the variance shows a characteristic increase and then decrease over time. This can be used to improve labeling at a context level.

6.2.2 Changes in texture

Several basic change in texture experiments were performed. The first used the slip apparatus to move the sensor over the aluminum bottom and then over a piece of 180 grit sandpaper. An example segmentation from this experiment is shown in figure 18. In all of these experiments, the high vibration section was matched with H5. The moment of impact with the sandpaper is generally indicated by a sudden rise in variance. In this particular plot, the contact occurs at tick 1240. It was not marked because the likelihood did not pass the persistence test, however the change in variance was detected. In all the plots at least two levels of vibration, as indicated by the variance estimate, are always marked: one for sliding on the aluminum, and one for the sandpaper. In figure 18 this can be clearly seen in the variance estimate at tick 500 and 1500.

The importance of this result, and the other texture experiments is that substantial (over 500 samples) sections of each signal are marked as statistically similar. This is a huge reduction in the size of the input to a higher level recognizer. The algorithms reduce large pieces of the raw signal into a single variance and set of estimation parameters.

This can be clearly seen in two additional texture experiments. In each of these experiments the sensor was moved by hand in order to get smooth, low vibration, motion. In the first experiment, the fingertip sensor was pulled over a smooth plastic surface and then over 4 sheets of ordinary paper (figure 19). In the second, the fingertip was pulled across a calibrated surface roughness made with a lathe (figure 20).

In figure 19 the transition to sliding over the paper is marked at tick 200. The algorithm detects both a change in the variance, which is caused by the difference in roughness between the two surfaces, and a change in the spectrum of the signal. In figure 20 the repetitive pattern of the lathed surface texture can be seen in the signal. The algorithm models this with the autoregressive coefficients. The algorithm detects the jump in the

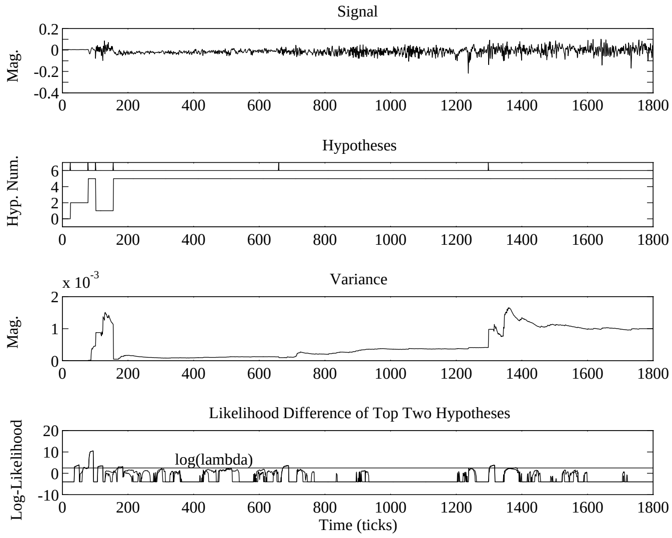

Figure 18: Example of change in texture: Sliding over an aluminum surface followed by sliding over 180 grit sandpaper.

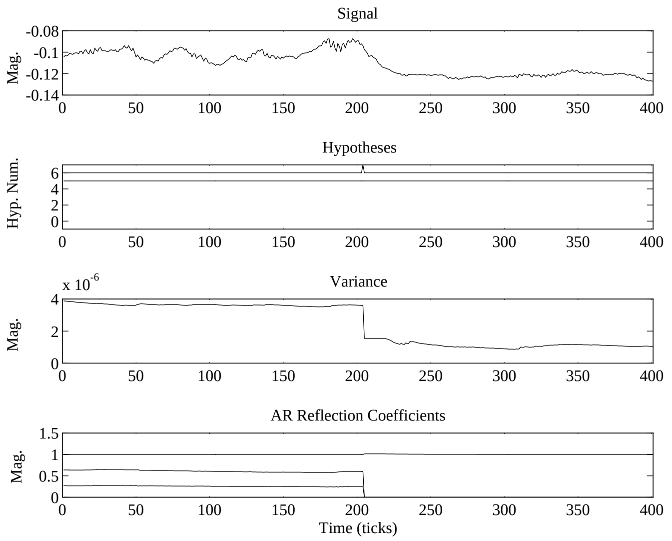

Figure 19: Example of change in texture: Section of sliding over a plastic table-top followed by four sheets of paper. The mark is placed at the point of change between the two surfaces. In addition to a change in variance, the algorithm detects a change in the spectral content of the signal as indicated by the AR coefficients.

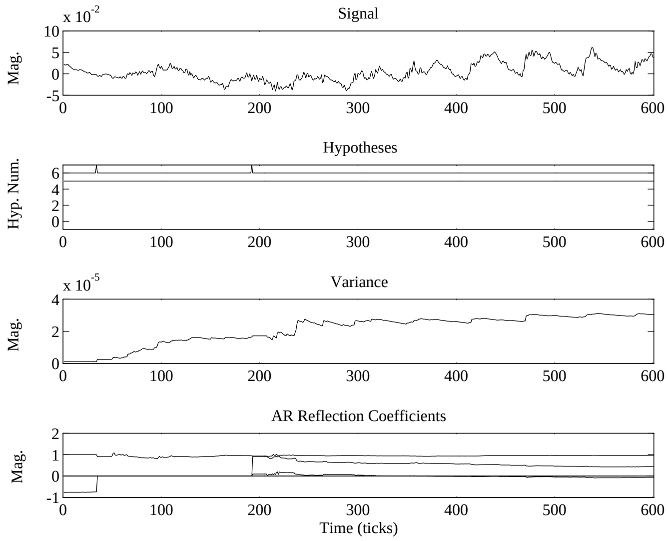

Figure 20: Example of change in texture: Motion across a calibration surface roughness calibration surface. The fingertip begins sliding over the surface at approximately tick 180. The surface was made with a lathe and has half-sinusoidal features of height 0.002 in and width 0.030 in.

spectrum when the fingertip begins to move across the textured surface. If the velocity of the sensor could be controlled, or modelled, the autoregressive coefficients could be potential used for identifying or sorting textures.

6.3 Summary

The results show that the algorithm tests for characteristics that changed when the contact conditions changed. The technique produced generally accurate estimates of the change times, and was sensitive to small variations in the contact conditions. Our experiments showed that changes between small texture variations were detectable when the sensor was moved by hand. It was less sensitive under computer control, possibly because of the additional robot vibration. The results show that the algorithm provides a method of discrimination between different contact conditions.

Identification, or labeling, of the correct signal source was problematic. This paper discussed work aimed at characterizing the statistical source of the data, and then using that source as a label. The autoregressive hypothesis H5 was often selected as the most probable even in the developmental data sets. In more general manipulation, where some low frequency interaction force will always be present, this hypothesis will become even more probable. Therefore, future work will explore using only the AR signal model

$$y(t) = \mu + \sum_{i=0}^{p} a_i y(t-i) + e(t)$$

as a basis for segmentation. This model encompasses almost all of the other models and therefore little segmentation power will be lost.

In addition, this model produces segments by fitting a spectrum to the local data. Therefore, all of the data in a segment will have approximately the same spectral content. Preliminary work shows that it may be possible to use the AR coefficients for texture sorting. The same texture tended to produce an equivalent spectrum when the sensor was moved at the same velocity. Additional work in adjusting for sensor velocity and characterizing the performance of this technique remains to be done.

In addition, a task level explanation of the results requires a task model or task context which was deliberately not introduced into this study. Future work is aimed at examining this problem. A context can provide both a interpretation framework, and a guide for placing better segmentation boundaries. The framework may depend upon position measurements and should predict the expected signal parameters.

7 Conclusion

This paper examined the problem of temporal segmentation of manipulation force signals. Substantially more information, useful for manipulation, then contact/no contact is carried by the contact force signals. A better understanding of the signal characteristics and their relationship to task models would provide a rich framework for controlling manipulation.

The first problem in designing a system to interpret the force signals, is to segment the signal into statistically equivalent classes. The sequential hypothesis testing method was introduced as a general tool for producing segment boundaries, and was applied to the problem of segmenting the individual strains from a 6-axis force torque sensor.

Experiments with isolated simple examples was used to develop a set of statistical source hypotheses for the data. These experiments resulted in the development of six basic statistical models. Impact signals were the most difficult signal to model and detect. A great many procedures and models were attempted before settling on a training data based model and the GLR test.

The GLR test produced accurate segmentation boundaries and event time marks. In addition, the results showed that an autoregressive model was often the most power hypothesis. We expect that this model will be sufficient, by itself, for producing good segmentations of the signal.

In addition, the framework produced some simple tests which can be used in place of threshold procedures and which have better performance and theoretical justification with no additional computational cost. An appropriate filter followed by the Page-Hinkley test can be used to test for many simple conditions. In some preliminary experiments, we were able to detect the onset of fingertip slip, by feeding the absolute value of a highpass signal into the Page-Hinkley test.

One of the most important design lessons is that the high frequency component of the force signal carries essential information. During manipulation, the robot fingers will exert time varying lowpass forces. Therefore,

Level 4:	Associating Dynamics with Objects in the Environment
Level 3:	Characterizing Dynamics of Interaction
Level 2:	Association of Signal Statistics with Regions of Space
Level 1:	Network of Expected Signal Statistics
Level 0:	Characterizing Sensor Signal Statistics

Table 1: Hierarchy of haptic explanations.

any unexpected events, like impact or slip must be detected in the high frequency range. Lowpass filtering the sensors to eliminate this "noise" removes a major information channel.

Soft covers, used for better gripping, for the aluminum fingertips have the effect of lowpass filtering the contact signals. Therefore, a two part sensor that uses a high frequency sensor embedded in the cover, and the fingertip sensor for low frequency signals needs to be developed. The high frequency sensory can be coarsely distributed and have poor directional sensitivity. It's only the vibration magnitude that is important.

Sensors should also be designed with critical damping properties. The fingertip sensor is lightly damped and rings when excited at high frequencies. Light damping makes modeling the force signal source more difficult, because near the natural frequency the signal is a reflection of the sensor's dynamics and not the input signal.

Although the full algorithm implemented in this paper does not run in real-time, we feel that a less complete set of tests based purely on the AR models could be made to run in real-time on a DSP processor. The steps in the algorithm are highly vectorized and could be run in parallel. We are investigating these possibilities.

Additional discrimination is possible by examining the spatio-temporal distribution of the contact signal. Local slip could be distinguished from an overall vibration by examining this distribution. However, this requires the development of new array sensor technologies.

The procedures reported on in this paper provide a powerful front-end to an algorithm for interpreting the state of contact. Such an interpretation will require task context. This can be provided with several approaches each with different levels of explanatory power. We envision the hierarchy of haptic explanations shown in table 1. This sequence of levels of interpretation provides the basis for our long-term research agenda in dynamic haptic sensing. Our current research is aimed at investigating levels 1 and 2 to provide a contextual basis for interpreting the sensor signal statistics, and the bottom level was investigated in this paper.

Level 1 is the lowest level of context. In this level an expected sequence or network of expected hypotheses is associated with a task. For example in a change of texture task, the algorithm would be told that two textures are present and that it is expected that a transition from one to the other will occur. Each texture would be characterized by the parameters that the level 0 algorithm is expected to estimate for each region.

The next level of complexity is to associate the expected hypotheses and their parameters with a region of the workspace. This approach again builds a graph, but now the graph is indexed by regions of space. This approach requires measurements of position.

With the measurement of position, a different graph can be constructed (level 3). This graph would predict relationships between force signals and the position signals. The elements in the graph become the possible dynamic relationships between position and contact force. This would require knowledge of the possible physics available in the task.

Finally, at the highest level of context the dynamics and the signals are associated with objects in the environment. This would require approximate knowledge about the shape and associated physics for the objects that the robot might interact with. This could either be programmed in or be acquired through sensing with a combination of haptic sensors and vision.

This paper addressed the problem of dynamic contact sensing. Dynamic models of different basic contact events were developed and used to derive a statistical segmentation algorithm based on the generalized likelihood ratio teat. This test provides a powerful model-based framework for developing segmentation algorithms. The procedure was shown to be capable of segmenting a number of useful signals into statistically equivalent pieces. This resulted in a huge reduction in the amount of data that has to be given to a higher level recognizer. Finally, we presented a framework for future study of higher level recognizers and discussed the challenges presented by our approach to perception based robot programming.

References

[1] H. Akaike. A new look at statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723, December 1974.

[2] Theodore Wilbur Anderson. *An Introduction to Multivariate Statistical Analysis*. Probability and Mathematical Statistics. John Wiley & Sons, 2nd edition, 1984.

[3] Michèle Basseville. Detecting changes in signals and systems - a survey. *Automatica*, 24(3):309–326, 1988.

[4] Albert Benveniste, Michèle Basseville, and Georges V. Moustakides. The asymptotic local approach to change detection and model validation. *IEEE Transactions on Automatic Control*, 32(7):583–592, July 1987.

[5] Antonio Bicchi, J. Kenneth Salisbury, and David L. Brock. Contact sensing from force measurements. memo 1262, MIT Artificial Intelligence Laboratory, October 1990.

[6] Antonio Bicchi, J. Kenneth Salisbury, and Paolo Dario. Augmentation of grasp robustness using intrinsic tactile sensing. In *Proc. IEEE International Conference on Robotics and Automation*, pages 302–307, 1989.

[7] D. L. Brock and S. Chiu. Environment perceptions of an articulated robot hand using contact sensors. In *Proceedings of the ASME Winter Annual Meeting*, Miami, FL, 1985.

[8] David L. Brock. *A Sensor Based Strategy for Automatic Robotic Grasping.* PhD thesis, MIT, Department of Mechanical Engineering, 1993.

[9] David Dornfeld and Christopher Handy. Slip detection using acoustic emission signal analysis. In *Proc. IEEE International Conference on Robotics and Automation*, pages 1868–1875, 1987.

[10] S. R. Hall. *A Failure Detection Algorithm for Linear Dynamic Systems.* PhD thesis, MIT, Department of Aeronautics and Astronautics, 1985.

[11] E. M. Hemerly and M. H. A. Davis. Strong consistency of the pls criterion for order determination of autoregressive processes. *The Annals of Statistics*, 17(2):941–946, 1989.

[12] Robert D. Howe and Mark R. Cutkosky. Sensing skin acceleration for slip and texture preception. In *Proc. IEEE International Conference on Robotics and Automation*, pages 145–150, 1989.

[13] Robert D. Howe and Mark R. Cutkosky. Touch sensing for robotic manipulation and recognition. In O. Khatib, J. Craig, and Tomás Lozano-Pérez, editors, *Robotics Review 2.* MIT Press, 1991.

[14] Roland S. Johansson. *Tactile Afferent Units with Small and Well Demarcated Receptive Fields in the Glabrous Skin Area of the Human Hand*, pages 129–145. Plenum, 1979.

[15] Roland S. Johansson and A. B. Vallbo. Tactile sensibility in the human hand: Relative and absolute densities of four types of mechanoreceptive units in glabrous skin. *Journal of Physiology*, 286:283–300, 1979.

[16] Roland S. Johansson and Göran Westling. Afferent signals during manipulative tasks in man. In O. Franzen and J. Westman, editors, *Information Processing in the Somatosensory System: proceedings of an international seminar at the Wenner-Gran Center.* Macmillian Press, 1991.

[17] K. O. Johnson and S. S. Hsiao. Neural mechanisms of tactual form and texture perception. *Annual Review of Neuroscience*, 15:227–250, 1992.

[18] Lennart Ljung. *Theory and Practice of Recursive Identification.* MIT Press, 1983.

[19] John Makhoul. Linear prediction: A tutorial review. *Proceedings of the IEEE*, 63(4):561–580, 1975.

[20] Brenan Joseph McCarragher. *A Discrete Event Dynamic Sytems Approach to Robotic Assembly Tasks.* PhD thesis, MIT, Department of Mechanical Engineering, 1992.

[21] Jorma Rissanen. Stochastic complexity and modeling. *The Annals of Statistics*, 14(3):1080–1100, 1986.

[22] Jorma Rissanen. *Stochastic Complexity in Statistical Inquiry*, volume 15 of *Series in Computer Science.* World Scientific, 1989.

[23] Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[24] Ritei Shibata. Selection of the order of an autoregressive model by akaike's information criterion. *Biometrika*, 63(1):117–126, 1976.

[25] David Siegmund. *Sequential Analysis, Test and Confidence Intervals.* Springer-Verlag, 1985.

[26] M. A. Srinivasan, J. M. Whitehouse, and R. H. LaMotte. Tactile detection of slip: Surface microgeometry and peripheral neural codes. *Journal of Neurophysiology*, 63(6):1323–1332, June 1990.

[27] Sharon A. Stansfield. Visually-aided tactile exploration. In *Proc. IEEE International Conference on Robotics and Automation*, pages 1487–1492, 1987.

[28] Abraham Wald. *Sequential Analysis.* John Wiley & Sons, Inc., 1947.

[29] A. S. Willsky. A survey of design methods for failure detection in dynamic systems. *Automatica*, 1976.

[30] Alan S. Willsky and Harold L. Jones. A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *IEEE Transactions on Automatic Control*, 21(1):108–112, February 1976.