Published in the Third International
Symposium on Experimental Robotics
Kyoto, Japan, Oct. 28-30, 1993

Segmentation and Interpretation of Temporal Contact Signals

Brian S. Eberman and J. Kenneth Salisbury

MIT Artificial Intelligence Laboratory
Cambridge MA 02139

Abstract— The temporal structure of the force, or strain, signals from an internal force-torque sensor can be a rich source of information about robot/environment contact conditions. We present a procedure for representing and segmenting the individual signals with an auto-regressive model. This procedure produces segments of approximately constant spectrum, and is a powerful technique for detecting abrupt changes in noisy signals without prior calibration.

1. Introduction

It has long been recognized that manipulation is a problem involving discontinuous changes in contact conditions. The slot of a screw interacting with a screwdriver, the detent in a switch, the placing of an object on a table all have different contact conditions depending on which surfaces are mated. In addition, manipulating a grasped part will cause changing contact conditions at the grasp contact points.

Controlling the contact conditions requires the ability to recognize a given contact state, and recognition requires an adequate representation of the measured signals. Some interactions can be modeled by the constraints imposed by rigid bodies with friction. However, in more general environments this representation always becomes computationally complex and is often not even sufficient. Consider the difficulty of modeling the contact forces created by grasping a pile of paper clips, by sweeping a hand through mud looking for a rock, by shearing a sheep, or even by just snapping on a pen cap.

Humans manage these tasks by integrating the spatial-temporal output of four different tactile sensory channels with kinesthetic channels, vision, hearing, and prior knowledge. Using an unknown mechanism and representation, information is extracted on at least: 1) contact locations, 2) contact forces, 3) local tangential motions at the contact (slip), 4) surface texture, 5) contact geometry, and 6) tool motion constraints.

This is a difficult problem which will require a variety of different sensors and techniques. We have started looking at the temporal strain and position signals generated by idealized tool interactions. If manipulation is performed

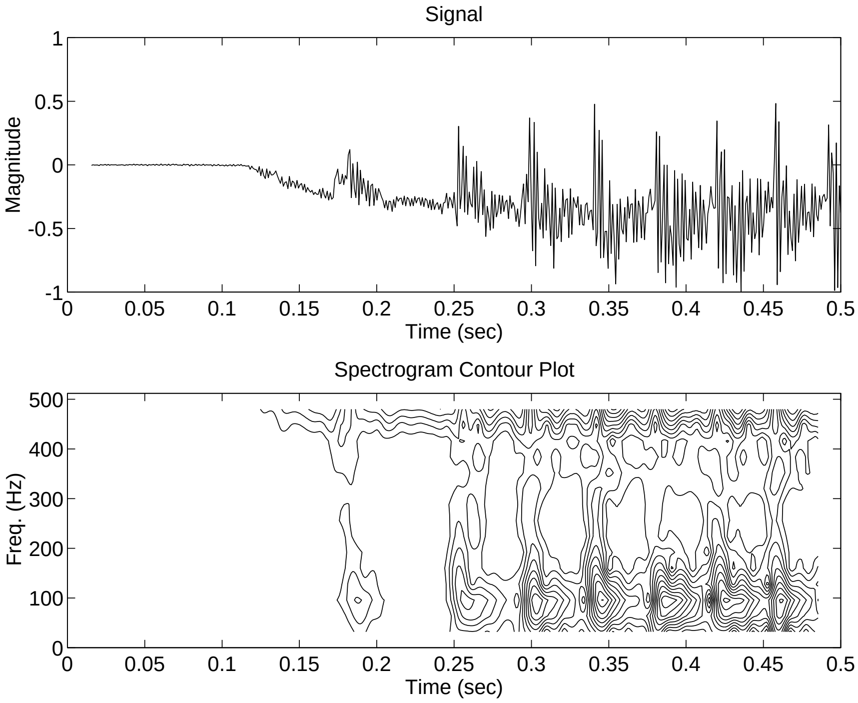

Figure 1. Spectrogram of a sequence of impact events. The figure shows a contour plot of the energy in frequencies from 32-480 Hz as a function of time. The signal was sampled at 1024 Hz. Thirty-two points windowed with a Hamming window were used for each Fast Fourier Transform (FFT). The FFT was computed for every new data point. Note the peaks that occur at each impact event and the short time scale of each event.

through a stick (or tool), the stick encodes the spatial-temporal contact forces into a single temporal signal. Humans are able to extract a great deal of information about the nature of the stick's contact from this signal [1].

Our research is aimed at developing techniques for providing a robot with a force-torque sensor and position sensors the same interpretational capabilities. This should be possible, because, although the measurement channels are different the same information is being measured by both the robot and the human hand.

Our goal is to develop interpretation algorithms that:

1. Are based on the discontinuous nature of manipulation,
2. Are as free of calibration requirements as possible,
3. Are hierarchical where each level produces simplified features for the next highest level, and

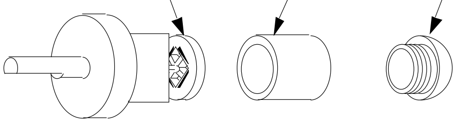

Figure 2. 6-axis fingertip force-torque sensor

4. Use geometry only at the highest level.

Because of the discontinuous way in which contacts are made and broken, the contact force signals consist of a sequence of regions separated by these contact events. Therefore, the algorithm should use these events to produce a sequence of force signal segments. Calibration is time consuming and can introduce systematic bias into procedures. By eliminating calibration, we hope to make the procedures easier to use and more robust. A multi-layered approach can be broken into many cooperating programs and processes for faster computational performance. Finally, the association of forces with geometry is computationally expensive in both cycles and memory. By using geometry only at the highest level, we hope to minimize the computational cost by providing higher level features than the raw force, and increasing the time between geometric calculations.

This paper reports on our experimental development of techniques for representing and segmenting scalar strain signal processes using an auto-regressive model. Vector processes are treated as independant scalar processes. Segmentation is performed with forms of the sequential likelihood ratio test, which is a statistical technique for testing for jump changes in stochastic processes. We concluded by discussing approaches to improve the techniques using estimated constraint surface models.

2. Intrinsic Contact Sensing

Intrinsic contact sensing is the measurement of information about the contact point from an internal, or intrinsic, force-torque sensor [2]. Our investigation used the sensor shown in figure 2. This sensor has a 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 700 Hz when the base is fixed and the shell free.

A tremendous amount of information, besides just the contact force, is available in the strain signals measured by such a force/torque sensor. Figure 1 shows a spectrogram from a single strain signal caused by a sequence of impacts. Each impact results in an energy at all frequencies locally around the event, and a persistent residual vibration at the sensor's natural frequency.

This information is carried not only in the values of the strains at a point in time, but also in the temporal structure or correlation of the strains through time. We are interested in capturing the temporal structure of the signals and detecting jump changes in that structure in order to detect events.

3. Scalar Signals

The correlation structure of a scalar (or vector) signal can be captured by modeling it as a autoregressive (AR) process. The scalar autoregressive model assumes the current strain signal $y(t)$ is generated by

$$y(t) = \sum_{i=0}^{n} a_i y(t-i) + v(t). \tag{1}$$

where $v(t)$ is a white noise process with unknown covariance Q, and the a_i are the lag autoregressive coefficients. When the contact conditions change, the parameters of this model are assumed to discontinuously jump.

An example of this type of model jump is shown in figure 3. In this figure, the sensor was dragged at constant velocity across a smooth plastic surface followed by a writing paper surface.

3.1. Automatic Segmentation

Segmentation of scalar signals based on the spectrum has a wide variety of applications and several techniques have been developed. Our technique is based on work on failure detection and speech segmentation [3, 4, 5]. We use a fast square-root algorithm for recursively computing the reflection coefficients of the process [6]. Then, our algorithm chooses the optimal model order using a minimum description length (MDL) length penalty [7]. Finally, sequential segmentation and model resetting is performed with the sequential likelihood ratio test [3].

The coefficients a_i fit an all pole spectrum to the data used in the estimation. Therefore, by using an appropriate segmentation procedure, a sequence of AR models can be generated which captures the spectrum over a window where the spectrum is approximately constant.

3.1.1. Fast Square-Root Algorithm and Model Order

Linear models, like equation 1 can be written in the form

$$y(t) = \Psi^T(t)\theta + v(t)$$

where $\Psi(t)$, a column vector, is formed from the lagged values of $y(t)$. In the case of equation 1 $\Psi^T(t) = [y(t-1), ..., y(t-n)]$ and $\theta^T = [a_1, ..., a_n]$. The least squares estimate at time t, $\hat{\theta}(t)$, is computed by sequentially solving

$$\begin{aligned} I(t) &= I(t-1) + \Psi(t)\Psi^T(t) \\ X(t) &= X(t-1) + \Psi(t)y(t) \\ I(t)\hat{\theta}(t) &= X(t). \end{aligned}$$

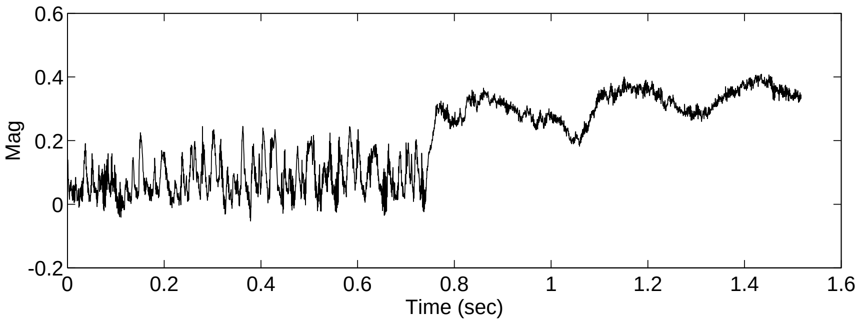

Figure 3. The figure shows one strain signal, sampled at 2700 Hz, that results from moving the probe across a relatively smooth plastic surface followed by a writing paper surface. The first texture was fit with sampled data poles at $[-.30 \pm 0.91i, 0.97, 0.79]$ at a variance of $2.2\,10^{-4}$ and the second was fit with poles at $[-0.30 \pm 0.88i, 1.0, 0.54]$ and a variance of $5.0\,10^{-5}$.

The equations can also be rewritten to avoid the linear accumulation, but this obscures the procedure.

The calculations can be improved by generating an orthogonal set of regressors from $\Psi(t)$ and finding the parameters that correspond to this orthogonal set. These parameters are often called the reflection coefficients or linear predictor coefficients (LPC). The LPC's are computed by the following equations

$$
\begin{aligned}
R(t)R^T(t) &= R(t-1)R^T(t-1) + \Psi(t)\Psi^T(t) \\
R(t)\,[\,\tilde{k}(t) \quad \bar{\Psi}(t)\,] &= [\,X(t) \quad \Psi(t)\,] \\
k_j(t) &= \tilde{k}_j(t)/R_{j,j}(t)
\end{aligned}
$$

where $R(t)$ is the lower triangular square-root of $I(t)$ and $k(t)$ are the reflection coefficients. The first equation can be solved with order n^2 multiplies and n square-roots by applying a series of Householder transforms, which together form T, to solve

$$[\,R(t-1) \quad \Psi(t)\,]\,T = [\,R(t) \quad 0\,]\,.$$

Ladder algorithms are slightly faster than this algorithm for AR models, but this method permits a general form for $\Psi(t)$.

Given the reflection coefficients $k(t)$, the optimal model order can be selected by minimizing the minimum description length over the number of parameters m

$$
\begin{aligned}
L_m(t) &= -t\log(\hat{Q}_m) - b(m+1)\log(t) && (2) \\
&\text{where,} && (3) \\
\hat{Q}_m(t) &= \left[E(t) - \sum_{i=1}^{m} \tilde{k}_i^2(t)\right]/t && (4)
\end{aligned}
$$

where $E(t) = E(t-1) + y(t)^2$ is the signal energy, and b is the parameter cost. The parameter cost is a design variable, we generally used a value of 2.4. Asymptotically a value of 1 can be shown to give unbiased estimates of the optimal model size, however the estimated model order is relatively insensitive to b for small n. Finally, the residuals can be computed in the orthogonal frame via

$$\nu_j(t) = \nu_{j-1}(t) - \bar{\Psi}_j(t)\tilde{k}_j(t) \tag{5}$$

where $\nu_0(t) = y(t)$. The residuals are used for segmentation.

3.1.2. Segmentation

Sequential segmentation of both the AR model and the constraint surface models is performed using a sequential log-likelihood ratio test. This test can be shown to have the smallest delay to decision over all tests with a given error rate on a two hypothesis testing problem. If we assume that the error in the AR model is Gaussian, then the prediction residuals are white, zero-mean, Gaussian random variables with variance $\hat{Q}$. A statistical test which checks for changes in the variances of size $\alpha\hat{Q}$ was used.

Consider a hypothesis testing problem between two possible sources for an n dimensional stochastic process ν: $\mathcal{N}(0, Q)$ and $\mathcal{N}(0, \alpha Q)$ where $\mathcal{N}$ is the multivariate normal distribution. The log-likelihood ratio between these two hypotheses, for a sequence of measurements ν_r^t, is

$$\begin{aligned} l(\nu_r^t) &= \sum_{i=r}^{t} \log \mathrm{p}(\nu_i|\alpha Q) - \log \mathrm{p}(\nu_i|Q) \\ &= -\frac{1}{2}\sum_{i=1}^{t} n \log \alpha + \left(\frac{1}{\alpha} - 1\right) \nu(i)^T Q^{-1} \nu(i) \end{aligned}$$

The sequential log-likelihood ratio test maximizes this value over all values of r between 1 and t. This can be done by monitoring the test statistic $g(t)$ generated by the sequential rule

$$\begin{aligned} \gamma(t) &= -\frac{1}{2}\left(n \log \alpha + \left(\frac{1}{\alpha} - 1\right) \nu(t)^T Q^{-1} \nu(t)\right) \\ g(t) &= \max(0, g(t-1) + \gamma(t)). \end{aligned}$$

Under the no change hypothesis, $g(t)$ is a cumulative sum test with resetting at 0 since $\mathrm{E}(\gamma(t)) < 0$.

The last time g was 0 is the optimal change time. Changes are detected by thresholding g. Reasonable values of the threshold can be picked with a few experiments, or by noting that γ is a function of a chi-squared random variable. The fineness of the segmentation is controlled by the level of the decision parameter. This test is optimal if the two distributions for ν are known. Further details on this test and more powerful tests are given in [8].

In practice, we substitute the current value of $\hat{Q}$ for Q in the test and run two simultaneous tests. One test looks for changes of size αQ and the other test looks for changes of size Q/α. α is used as a design parameter. We used a value of 4 for α and 15 for the threshold in the AR segmentation.

3.2. Experimental Results

The experiments were carried out with the single degree-of-freedom device. An 8 inch glass-epoxy link connects a brushless DC motor, with an integral reflection torque sensor, to the sensor. The output torque from the motor is controlled to one part in 1000 using an integral controller. The motor's design and control is described in [9]. The glass-epoxy link provides a very stiff, well damped linkage. Test specimens can be placed under the sensor or on blocks at the end of its travel. Both the angle and force measurement have 12 bits of resolution.

To begin processing the signals, we applied a finite impulse response low-pass filter of width 32 and decimated the measurement at 2:1 to create a new measurement signal at 512 Hz. This eliminated most of the components of the signal that resulted from resonance of the sensor. We then experimented with different models for the measurement. We tried AR models which were extended with a mean, and with a slope and a mean. These additional terms caused difficulties. In some cases a first order AR process was more useful then a mean processes, and in other cases the mean was more useful. Since the model order selector cannot reorder the regressors, no one order for the regressors will work in both cases. There are two possible remedies.

First, additional models with different orderings can be run simultaneously and the optimal model can be chosen. This is the approach used in [8]. This has the advantage that a constant flat signal, which is generated in free space or while at rest, can be selected. Second, an eigenvalue decomposition of I can be performed to find the optimal regression directions. Unfortunately, both techniques are computationally expensive. Furthermore, during manipulation the mean only model will seldom be selected.

We also tried to use the difference between the lowpass signal and the raw measurement as an indicator of impacts. This high pass response does produce a distinct jump when impacts occur, but because of the sensor's poor damping the signal is blurred over a long time window. This makes time marking of the event difficult, especially when a sequence of impacts occurs.

The AR model does produce useful tests. In one experiment, the sensor was commanded to move at constant velocity over a polished aluminum plate, which had a single small scratch. A stylus was attached to the fingertip to decrease its effective size. Figure 4 shows the algorithm's response to the burr. All of the change information in this experiment is in the variance estimate. The burr produces signals which are improbable given the current estimate of the variance. This causes an accumulation in the decision statistic and after a few measurements a change is detected. The AR parameters for the low frequency signal are all at approximately 1.0 before and after the marked change caused by the burr.

The data shown in figures 1 and 5 was generated by commanding the sensor to move at constant velocity over a flattened hose-clamp. Each ridge in the

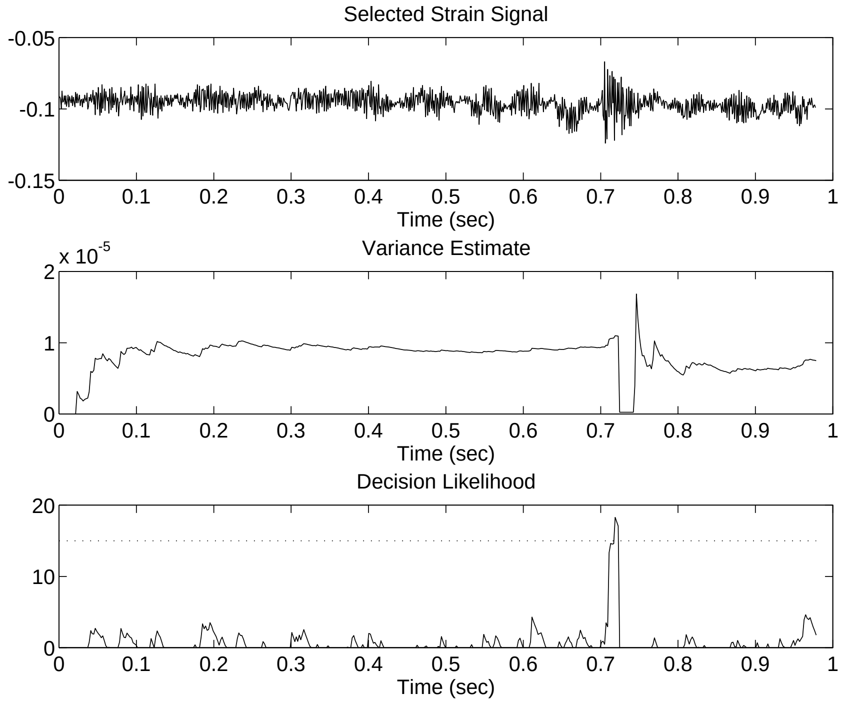

Figure 4. Burr Experiment. One of the eight strain signals is shown in the top figure. The burr occurs at approximately 0.75 seconds. The AR parameters for all the strains use a single AR parameters of just less than 1.0 (not shown). After regressing with this parameter, the variance estimate shown in the second figure is computed. The bottom figure shows the change likelihood and the threshold of 15. The beginning of the burr is marked correctly by the last time the likelihood was at zero. There is a lag of 2n steps after each detected change before a new variance estimate is produced.

hose-clamp produces a spike in the data. The AR model segmentation, detects a constant frequency spectrum consisting of poles at $[0.30 \pm 0.54i \quad -0.56]$. The additional reflection coefficients are introduced by the MDL model selection criteria which is computed for every new measurement. The change likelihood crosses the threshold at the beginning and end of the texture segment, because there is a marked change in the signal variance as the sensor enters and leaves the contact region.

Figures 6 and 7 shows a similar experiment with a roughness calibration plate. This experiment was performed by moving the sensor by hand and sampling with an effective (after decimation) rate of 1350 Hz. These two figures illustrate some of the problem sources for our experiments. First, even though the sensor was moved relatively slowly, the input signal has a frequency of

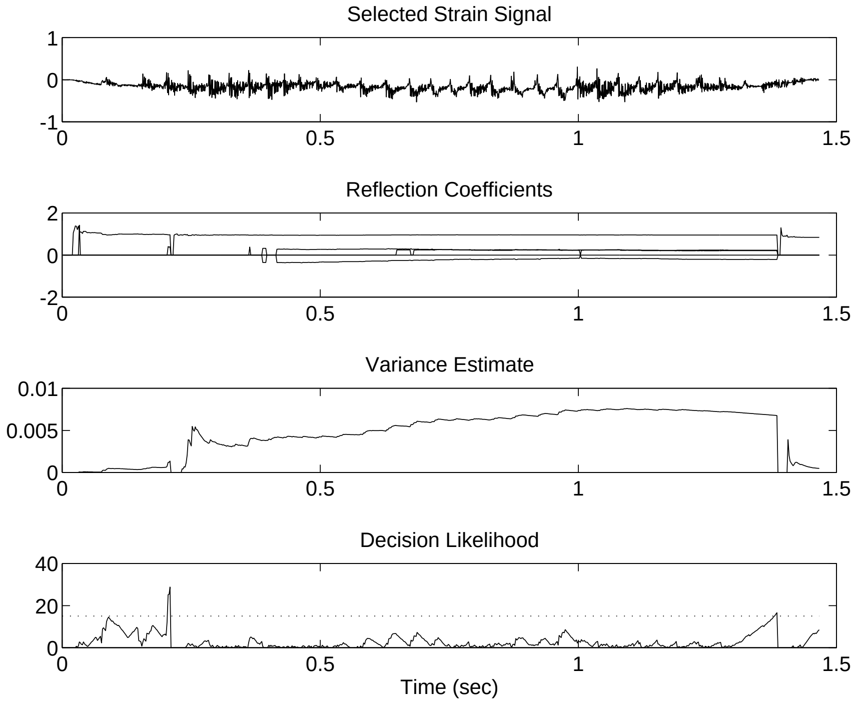

Figure 5. Texture Experiment. One of the eight strain signals is shown in the top figure for the same experiment illustrated in figure 1. The AR model is able to identify a constant frequency pattern starting at approximately 0.4 seconds represented by reflection coefficients [0.95, 0.21, -0.21,0.23]. The likelihood test found three regions where the variance is significantly different (third figure). The bottom figure shows the change likelihood and the threshold of 15.

approximately 40 Hz. This gives only 13 points per cycle at a 512 Hz sampling rate. Second, the signal is obscured by the senor's long vibration decay time. This is most evident in the second plot in figure 7. Therefore, a good system needs a high sampling rate (order 1-2 kHz) with well damped sensory characteristics across the entire frequency range.

4. Discussion

In general, the AR procedure was able to reliably pick out locations in the signal where the spectral shape or energy of the signal had appreciably changed, and is therefore a good representation for segmentation. This is a very useful feature in a layered interpretation scheme. Because it substantially reduces the work load on the higher layers by providing useful segments. The AR parameters are essential for this process, because they make the test significantly more sensitive to changes in energy by whitening the measurement process. For example,

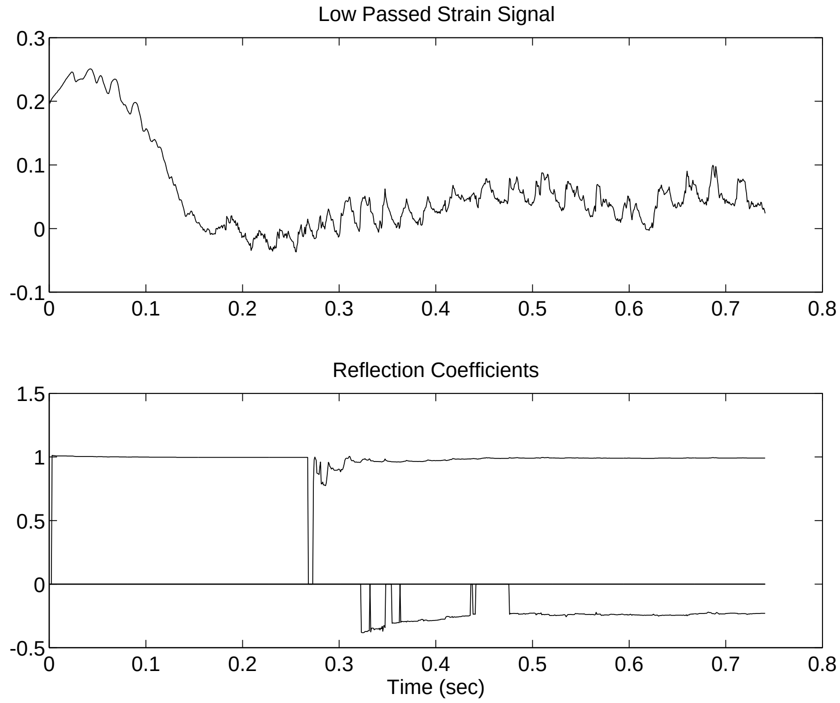

Figure 6. Calibration Plate Experiment. One of the eight strain signals is shown in the top figure. The sensor was moved, by hand, across a 12.5 μmRa turned surface roughness calibration plate. The plate contact beings at approximately 0.15 seconds. The AR model is able to identify a constant frequency pattern starting at around .32 seconds. The value of using an additional reflection coefficient changes depending upon the amount of noise in the signal.

without a whitening scheme the burr signal would have been undetectable.

Whether the representation is rich enough to represent the myriad of different signal forms that could occur during manipulation remains an open question. The procedure represents the power spectral density of a section of data. The length of the section is chosen based upon the data. However, the original signal is not recoverable from this representation because only its statistics are represented.

A second difficulty is that force interpretation, unlike vision or speech, has the unique attribute that the robot's actions directly effect the signal. The AR procedure models this by using a single reflection coefficient at or near 1.0. This essentially removes from the signal the very low frequency terms that correspond to the robot's motions. An alternative would be a quasi-static analysis of the strains. In theory, the low frequency term can be predicted given a model of the geometry, an estimate of the robot's location in configuration space, and

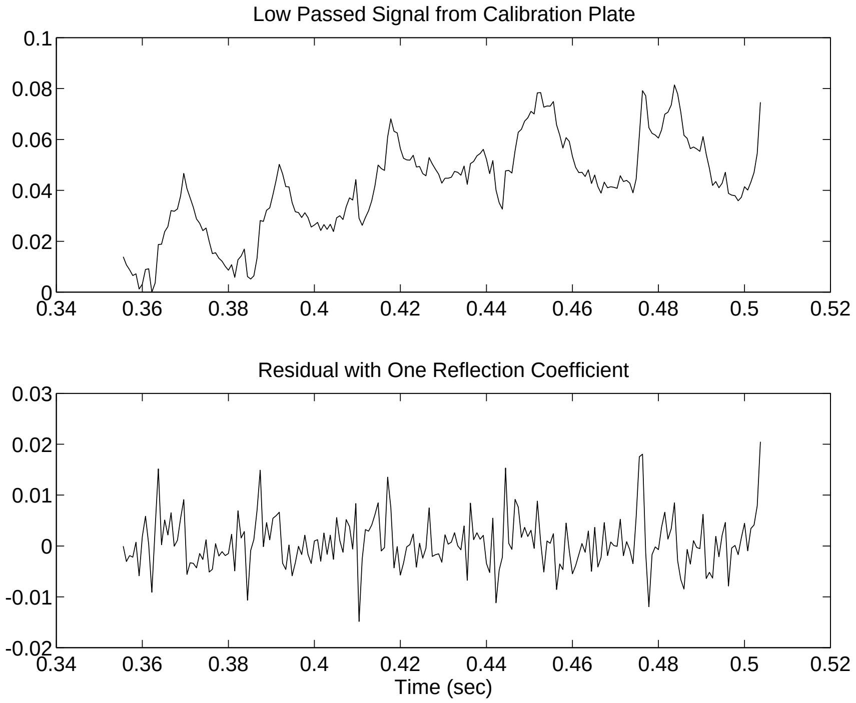

Figure 7. Calibration Plate Experiment. A section of the calibration plate strain signal and the residual after applying the first reflection coefficient is shown. The sinusoidal form of the signal is a result of the sinusoidal texture. The signal was sampled at 2700 Hz and decimated 2:1.

measurements of the robot's applied torques. While this is a powerful method, it is very difficult computationally, and it requires a geometric model.

An alternative approach is to fit a regression relationship between the position (or desired position) of the robot and the strain measurements. This procedure fits a stiffness to a configuration space constraint surface patch. By performing the regression in a combined strain/position space the "stiffness" regression can be made non-singular and symmetric. Our preliminary experiments with this approach in 1 DOF are very promising.

5. Conclusion

The temporal structure of a contact signal provides a rich source of information about the contact conditions. An auto-regressive representation of the signals can be used to segment the signal into constant spectrum regions. This approach is useful for robustly detecting sudden changes in the signal without prior calibration, and for providing useful segments for higher level procedures.

Representing measured textures will require extending the representation to make it invariant with velocity. Adequate performance in measuring and recognizing texture patterns will require a sample rate of 1-2 kHz and a sensor with well damped performance characteristics.

This study is our first step toward building a layered algorithm for interpreting tactual data. Several other layers are required before an interpretation of the signals at the level of interacting parts is possible. In order for a layered approach to succeed the lowest layer must provide a large reduction in signal bandwidth. Our AR procedure can fulfill that function by segmenting the signal into interesting sections.

Acknowledgement

The authors would like to gratefully acknowledge the financial assistance of 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 the Office of Naval Research University Research Initiative Program under Office of Naval Research contract N00014-86-K-0685.

References

[1] R. H. LaMotte. Personal communication, 1993.

[2] Antonio Bicchi, J. Kenneth Salisbury, and David L. Brock. Contact sensing from force measurements. *International Journal of Robotics Research*, To be pulished, 1993.

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

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

[5] Regine Andre-Obrecht. A new statistical approach for the automatic segmentation of continuous speech signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(1):29–40, January 1988.

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

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

[8] Brian Eberman and J. K. Salisbury. Application of change detection to dynamic contact sensing. Memo 1421, MIT Artificial Intelligence Laboratory, 1993.

[9] Michael Dean Levin. Design and control of a closed-loop brushless torque actuator. Technical report, MIT Artificial Intelligence Laboratory, AI-TR 1244, 1990.