

Object recognition with ultrasonic sensors

M. İhsan Ecemiş and Paolo Gaudiano*

Boston University Neurobotics Laboratory

Dept. of Cognitive and Neural Systems, 677 Beacon St., Boston MA 02215 USA
{ecemis, gaudiano}@cns.bu.edu <http://neurobotics.bu.edu>

Abstract

We have recently introduced a novel approach to sonar-based object recognition for robotics. The sonar recognition system—consisting of a Polaroid sonar coupled with an A/D data acquisition board in a LINUX-based PC—uses a fuzzy ARTMAP neural network classification system to recognize objects at varying distances. In this article we report additional results in which we test systematically recognition performance using different kinds of information from the sonar echo, and different object configurations. The results strengthen our claims that sonar can be used as a viable system for real-time object recognition in robotics and other application domains.

1 Introduction

At last year's CIRA Symposium our group introduced a novel system for recognizing objects using the spectral information contained in sonar echoes (Streilein, Gaudiano, & Carpenter, 1998). The system we presented uses the readily available 6500-series Polaroid sonar and an inexpensive data acquisition board that can operate under the LINUX operating system (DAS16-M1, Computer Boards, Inc., with a LINUX driver written by Warren Jasper of North Carolina State University).

Our work was based on the observation that animals such as bats and dolphins can perform remarkable sensory feats using ultrasound signals (Dror, Zagaeski, Rios, & Moss, 1993; Moore, Roitblat, Penner, & Nachtigall, 1991). In contrast, typical robotics applications only use sonar as a range finder, measuring the time-of-flight of the leading edge of the ultrasonic echo to determine the distance of the object that generated the echo (Borenstein, Everett, & Feng, 1996; Leonard & Durrant-Whyte, 1992).

*This work is supported by the Office of Naval Research and the Navy Research Laboratory through grant N00014-96-1-0772. We thank William Streilein for his contribution to an earlier version of this work, and Matt Giamporcaro for his electronic wizardry. P. Gaudiano's current address is: Artificial Life, Inc., Four Copley Place, Suite 102, Boston, MA 02116, USA.

In our first study we used a Fuzzy ARTMAP neural network (Carpenter, Grossberg, Markuzon, Reynolds, & Rosen, 1992) to classify echoes from five objects placed at various distances from the sonar. We chose ARTMAP because of its speed, its ability to learn incrementally and its proven performance on a variety of real-world pattern recognition problems. For a description of Fuzzy ARTMAP please refer to the original publication (Carpenter et al., 1992) or our earlier article (Streilein et al., 1998).

The results we presented last year were very encouraging: the recognition system was able to perform with an accuracy as high as 96% (Streilein et al., 1998). In that work, we used Matlab (Wolfram Research) to calculate the power spectral density (PSD) of each echo. The frequency information was then used as input to the Fuzzy ARTMAP neural network.

In this article we describe some extensions we have made to this work. First, we have written all the code in C to increase processing speed and flexibility in the data pre-processing scheme. Second, we have tested an alternative way of extracting information from each echo, using the envelope of the echo in the time domain rather than frequency content. Finally, we have tested the system's ability to learn to recognize objects at different distances and angles relative to the sonar. As we will show, the results confirm our previous findings about the value of sonar as a sensor for object recognition.

The remainder of this article is organized as follows: Section 2 describes the system, including the hardware and software components. Section 3 describes the results. The article closes with a short discussion of the results.

2 Data Collection and Preprocessing

The hardware system consists of an inexpensive Polaroid sonar and ranging module, (series #6500) and a data acquisition board that can operate under the LINUX operating system (DAS16-M1, Computer Boards, Inc., with a LINUX driver written by Warren Jasper of North Carolina State University).

For all results reported here, the sonar module was placed on a movable cart, at approximately the same height as the stand upon which each object was placed. The frontal surface of the stand was covered with foam rubber to minimize spurious echoes, though we have found from our previous experiments that recognition is quite robust even if other objects are ensonified by the sonar. The distance to each object was measured manually and later confirmed directly from the sonar echoes.

Each echo was sampled at a frequency of 500KHz. The raw echo is processed with a digital bandpass filter with cutoff frequencies at 15KHz and 93KHz. The filter is implemented by calculating the inverse Fast Fourier Transform (FFT) of an ideal frequency response (with an amplitude of 1.0 in the desired range and 0.0 elsewhere). A Hamming window is applied to the inverse FFT to smooth the filter in the time domain. Finally, the filter is normalized, truncated to 255 points, and convolved with the echo.

There are several reasons for using this digital filter:

- The filter suppresses aliasing errors
- Certain artifacts, including occasional cropping of the largest part of the echo, are removed.
- The filter significantly reduces noise present in the signal before the echo arrives. This makes the detection of the echo highly reproducible, with an accuracy of 1 data point (0.002msec).
- By removing the low-frequency components of the echo, the resulting signal is insensitive to 60Hz line noise and to overall fluctuations that occur for example when the battery that triggers the sonar is running low.

Figure 1 shows the filtered sonar echo returned by a 1-gal plastic bottle located 100cm in front of the sonar. The vertical lines demarcate the data extracted for calculation of the PSD function, as described below. For the envelope extraction, the data are truncated closer to the echo onset, as described later.

2.1 Extracting frequency information

In our original report each (unfiltered) echo was transformed into the frequency domain using Matlab's PSD function. The PSD is well suited as it averages frequency information across time windows, thereby obtaining more reliable measures of frequency content. We have tried using a simple FFT and found it to be much less reliable.

For the results presented here we calculated the PSD using the method of Welch, an approach that combines averaging and windowing. Specifically, we used 18 Hamming windows of 256 points each, with an overlap of 90%

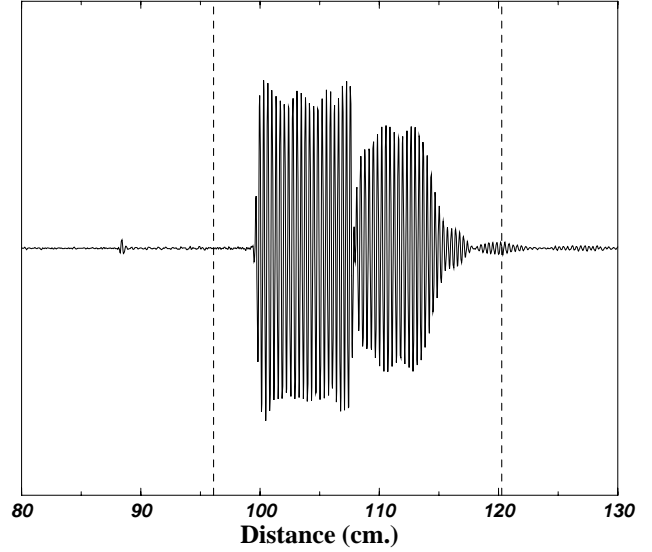


Figure 1: Typical sonar echo after application of the digital bandpass filter. The vertical dashed lines demarcate the portion of the echo used for calculating the PSD function, as described in the text.

between adjacent windows, covering a total of 698 points (which corresponds to 1.4msec of data). The PSD is calculated by summing the FFT of all 18 windows and dividing the total by 18.

The 256-point Hamming windows yield a resolution of $500\text{KHz}/256=1,953\text{Hz}$. Each PSD vector is truncated to the 40 elements in the frequency range [15,625Hz–91,797Hz], reflecting the characteristics of the band-pass filter. The 40-D vector is used as the input vector for the ARTMAP neural network.

The solid line in Figure 2 illustrates the average PSD obtained from the echo of the 1-gal plastic bottle located at a distance of 100cm from the sonar. Each point along the solid line is the average of 50 measurements. The dashed line near the bottom of the figure is the variance in the 50 measurements. This averaging was done to test the repeatability of the PSD function for a given object at a given distance. Our results (not shown here) demonstrate that PSD functions are highly repeatable for the same object at the same distance, though they can change significantly as the object is moved or rotated relative to the sonar.

2.2 Using time-domain information for recognition

The results obtained in our original experiments showed that the frequency content of each echo contains a significant amount of information about the object. However,

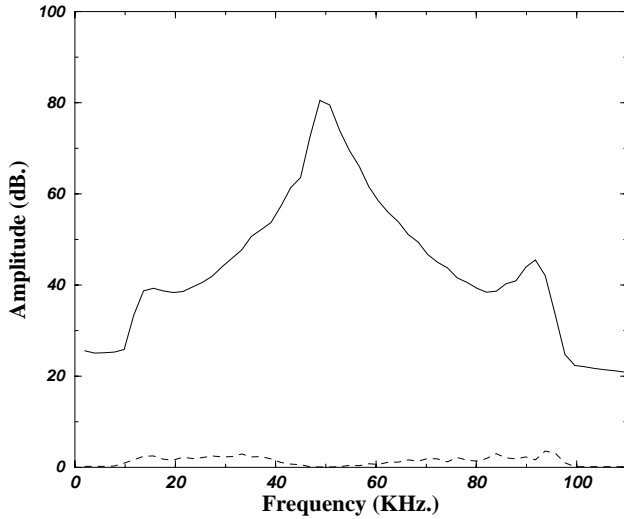


Figure 2: Average PSD (solid line) and variance (dashed line) calculated from 50 echos generated by the 1-gal plastic bottle at a distance of 100cm from the sonar.

we decided to try object recognition based on time-domain information for two reasons.

First, Sillitoe, Visioli, Zanichelli, and Caselli (1996) had shown excellent performance using a decision tree classifier to recognize 22 small objects based on their sonar echo. Only two points were extracted from each echo: the time of arrival and the amplitude of the echo's envelope at a fixed time interval after the echo onset. Surprisingly, even with this minimal information (collected from up to five different viewpoints at a fixed distance) an overall accuracy of over 98% was achieved. This suggested to us that the shape of the echo's envelope contains a significant amount of information.

Second, given the practical limitations of typical data acquisition hardware, frequency information may not be an ideal source of information. With a sampling frequency of 500KHz, the Nyquist sampling theorem suggests that we can measure frequencies up to 250KHz. In practice a more reasonable range might be 100KHz. The natural frequency of the Polaroid sonar is 55KHz. What is the useful range of frequencies we might want to measure? Any variations in the echo waveform should arise from two sources: (1) resonances set up by the sonar waveform, which depend on the material reflecting the waveform and its frequency response characteristics; (2) interference patterns resulting from different surfaces of the object being ensonified at slightly different times, which depend on the shape of the object.

In the first case, unless a material happens to have a natural frequency near 55KHz, the changes near the peak frequency are likely to be small. In fact, our informal obser-

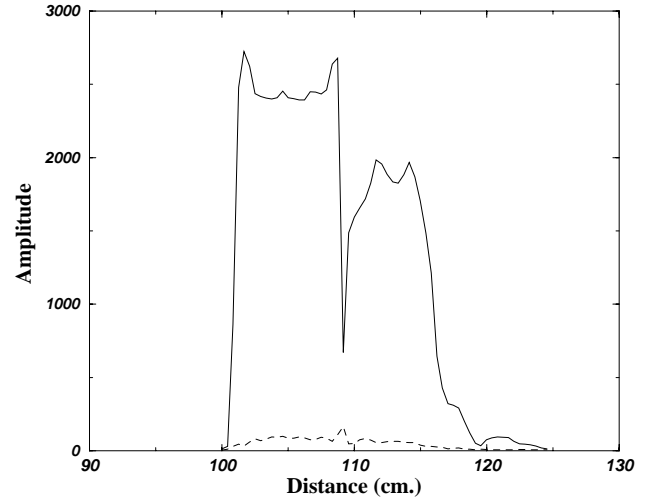


Figure 3: Average envelope function (solid line) and variance (dashed line) calculated from 50 echos generated by the 1-gal plastic bottle at a distance of 100cm from the sonar.

vations confirm this.

In the second case, consider an object with two surfaces separated by a distance of about 3.5cm. The speed of sound in air is approximately 350m/sec, so the echo from the second surface will be delayed by 0.2msec relative to the echo from the first surface (notice that the time is doubled to account for the round-trip distance). This corresponds roughly to a "feature" at a frequency of about 5,000Hz, which, given our Hamming windows of 256 points (i.e., a resolution of 1,953Hz), is close to the theoretical sampling limit. This suggests that unless we increase the Hamming window size and the amount of data collected, which can become costly in terms of processing speed and introduces other problems, we can only identify "shape" features in the order of a few cm.

Based on these observations we decided to use the shape of the echo waveform in the time domain as the input vector to the neural network classifier. The high precision with which we can locate the onset of the echo suggested that shifting in the time domain would not be a problem.

Figure 3 shows the average and variance of 50 echoes from the 1-gal plastic bottle located at a distance of 100cm from the sonar. Each envelope is calculated as follows: Starting 30 data points (0.06msec) prior to the echo onset, the code finds the maximum value in each of 60 non-overlapping and contiguous windows of 12 data points each. Therefore the entire envelope function consists of 60 points spanning $720/500\text{KHz}=1.44\text{msec}$, which corresponds to a spatial range of about 25cm from the front edge of the object. This procedure effectively performs down-

sampling and half-wave rectification of the original waveform.

The entire envelope function is passed as a 60-D input vector to the ARTMAP neural network for classification. Please note that distance information is not passed to the neural network implicitly or explicitly: all inputs from a given object are classified to the same output node. As described in the next section, classification is always better with the envelope than with the PSD.

3 Results

In our first report we presented results for recognition of five objects at different distances. Tests were conducted using up to 50% of all the collected data for training and the remaining data for testing.

In this article we performed a similar experiment to determine, using our PSD and envelope functions, what accuracy can be achieved in recognizing an object independent of its distance from the sonar. In addition, we tested the system's ability to recognize objects at arbitrary distance and orientation relative to the sonar.

3.1 Distance-independent object recognition

The first experiment is designed to test how well ARTMAP can generalize to recognize objects at distances it has not seen during training. To this end we collected echoes for four different objects: a one-liter plastic water bottle, a metal trash can, a styrofoam sheet measuring approximately 34x63cm, and a lego wall measuring approximately 38x10cm. For each object we collected 50 echoes at 11 distances ranging from 50cm to 150cm in 10cm increments. In addition we collected 50 "distractor" echoes from two other objects at 100cm only.

The Fuzzy ARTMAP neural network was trained using a randomly selected subset of five processed echoes from each of the four objects at distances of 50, 70, 90, 110, 130 and 150cm, plus five randomly selected echoes from the two "distractor" objects at 100cm. For each training input vector the desired output class was set to one of six nodes indicating the object whose echo was being presented.

Learning was set to a single epoch (fast learning mode) with a vigilance level of 0.95. Because ARTMAP is sometimes sensitive to the order of input presentation in fast learning mode, we repeated each experiment 10 times (each time drawing a different random set of 5 processed echoes for each object) and report the average results. However, we found that there was little variance across individual experiments.

Testing was performed only for the four main objects, using the remaining 45 echoes from the distances of 50,

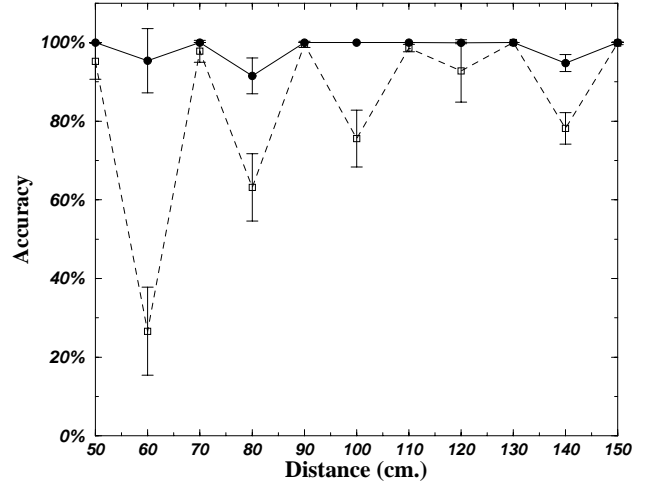


Figure 4: Average recognition accuracy as a function of distance using the envelope (solid line with circles) or PSD (dashed line with squares) as input vectors.

70, 90, 110, 130 and 150cm, and all 50 echoes from the distances of 60, 80, 100, 120 and 140cm. Hence, leaving aside the distractors, the neural network was trained with only approximately 5% of the data (none of the training data was used for testing).

Figure 4 shows the results of this experiment in terms of percent accuracy (across all four objects) as a function of distance. The solid line with circles shows the average results using the 60-D envelope function as input, while the dashed line with squares shows the average results using the 40-D PSD function as input. The error bars represent the variances in the 10 experiments. Several points merit discussion.

First of all, the envelope function yields better results than the PSD, as we had suspected. The envelope function yields 100% recognition at all the distances on which it was trained, and also on two of the distances on which it was not trained (100cm and 120cm). Performance always remains above 90% at all measured distances. The PSD function also yields 100% accuracy at those distances on which it was trained, but performance is considerably worse at intermediate distances.

Another interesting observation is that both the envelope and PSD yield better results around 120cm than around larger and smaller distances. We have verified informally that this peculiar behavior is the result of an automatic gain mechanism of the Polaroid sonar, which increases the gain of the receiver in several steps over time to overcome the dissipation of the ultrasonic wave as it travels through the air. The gain function is not uniform in the frequency domain and changes from step to step. This induces distance-

dependent changes in the PSD and in the envelope (to a lesser extent) that result in sudden changes in recognition performance across gain steps. The data points around 110cm happen to be about in the middle of one of the gain steps, so there is less change in the waveform and thus superior recognition performance in this range.

These results confirm and actually extend the results in our original report (Streilein et al., 1998). From a practical point of view, this suggests that using the envelope function it is sufficient to train the system to recognize an object every 20cm or so with only a few points from each distance. With a sonar firing every 100msec, this means that a robot approaching an object head-on can quickly learn to classify the object.

3.2 Recognition during unconstrained movement

One important restriction of the results in the previous section is that they are based on the object being “viewed” from a single angle. Given the directional nature of acoustic waveforms, one can expect dramatic changes in the PSD and envelope functions as objects are rotated by different angles. Streilein (1998) reports encouraging preliminary results using a variant of ARTMAP that accumulates information from multiple views of an object. In that situation the neural network was trained as the object was rotated in regular increments, and then performance was tested by presenting echoes from a few different angles.

In this article we tested object recognition under a fairly unconstrained configuration meant to imitate what might happen with a mobile robot. The sonar was mounted on a rolling cart and could thus be moved relative to each object. For this experiment we used five objects: a 1-gal plastic bottle, a cardboard box measuring 15x25x20cm, a styro-foam sheet measuring approximately 34x63cm, a lego wall measuring approximately 38x10cm, and a bucket-like box (the lego bucket) measuring approximately 18x18x25cm.

For each object, at the start of data collection the cart was moved back-and-forth across angles of ± 45 -deg and distances ranging between about 75cm and 130cm from the object. Because of the weight of the cart and the wheel configuration, the sonar was not always pointing directly at the object. This experiment was meant to replicate a scenario in which a robot is moving around an object.

Data collection for each object lasted about 20 seconds, collecting one sonar echo every 100msec or so, for a total of 200 echoes for each object. During this time the cart was moved from one extreme to the other approximately 5 times.

Training was performed using anywhere between 5 and 100 randomly chosen echoes (out of a total of 200) for each

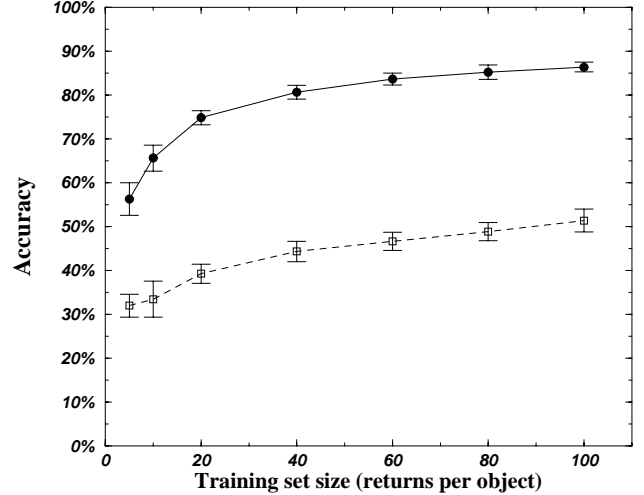


Figure 5: Recognition accuracy as a function of training set size for all five objects using the envelope (solid line with circles) or PSD (dashed line with square) function as input.

object. Testing only used the echoes that were not seen during training. ARTMAP vigilance was set to 0.9, training lasted one epoch (fast learning), and each experiment was repeated ten times with ten different random seeds. Again, the results were very stable across experiments.

Figure 5 shows recognition accuracy as a function of training set size (out of 200). As before, the envelope function results are shown as a solid line, and PSD results with a dashed line, while the variances are shown by the error bars. Two points are important. First, the envelope function clearly outperforms the PSD, in most cases nearly doubling the accuracy of recognition. Second, even in this relatively unconstrained case, ARTMAP performs remarkably well, achieving an overall recognition accuracy of over 55% with only 5 training vectors per object (2.5% of the data set), and nearly 90% accuracy with 100 training vectors per object (50% of the data set).

It is also interesting to consider the efficiency of the Fuzzy ARTMAP neural network. When 40 envelope vectors per object are used for training (20% of the data) and vigilance is set to 0.9, ARTMAP creates a total of 110 category nodes to classify all five objects at all distances and angles, achieving an overall accuracy of 82%. We also tried varying the vigilance parameter but found the results to vary only slightly.

These results strengthen our claim that sonar can be used as an effective sensor for object recognition.

4 Conclusions

We have presented results showing that sonar can be an effective sensor for object recognition. Our system can easily work in real time, making it possible to digitize, process and recognize each echo several times per second. Clearly, there are many ways in which we could try to improve our results, for instance by adjusting the pre-processing scheme, using a different classification method, or collecting larger data sets. We could also increase efficiency by using dedicated hardware for some of the pre-processing. However, we feel that the simplicity and robustness of this system are part of its appeal.

This is not the first proposal for the use of sonar for object recognition tasks. Kleeman and Kuc (1995) used a novel sonar array for classification of multiple targets into four reflector types (which are planes, corners, edges, and unknown), by combining the ranging information from two transmitters and two receivers. Sillitoe, Mulvaney, and Visioli (1995) have used a radial basis function neural network to recognize corners, poles, and other shapes typically found in indoor environment using a *bistatic* sonar array (a bistatic sonar is one in which the transmitter is separate from one or more receivers). Sobral, Johansson, Lindstedt, and Olsson (1996) perform recognition of simple objects by generating transfer functions or impulse responses from the envelope of the sonar echo of each of four objects at eight orientations, then using regression to find the best matching object for a given novel input vector (i.e., an unknown object). All of these approaches, however, utilize specialized hardware and are thus not easy to replicate.

Our goal eventually is to migrate the entire system on-board one of our robots. So far we have worked with a stand-alone sonar because of our frequent need to modify the setup, but all the components could easily be placed inside any robot with sonar sensors and an on-board PC, such as ActivMedia's Pioneers, Real World Interface's (now a division of IS Robotics) Bxx and ATR lines, or Nomadic Technologies' Nomad Robots.

One practical advantage of using the envelope function instead of the PSD is that we can sample at much lower frequencies and thus reduce the cost of the data acquisition card. As shown by Sillitoe et al. (1996), it should be possible to reduce the sampling dramatically and still obtain good results.

References

- Borenstein, J., Everett, H., & Feng, L. (1996). *Navigating mobile robots*. A.K. Peters, Wellesley, MA.
- Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J. H., & Rosen, D. B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks*, 3(5), 698–712.
- Dror, I. E., Zagaeski, M., Rios, D., & Moss, C. F. (1993). Neural network sonar as a perceptual modality for robotics. In Omidvar, O., & van der Smagt, P. (Eds.), *Neural Systems for Robotics*, pp. 1–16. Academic Press, San Diego.
- Kleeman, L., & Kuc, R. (1995). Mobile robot sonar for target localization and classification. *The International Journal of Robotics Research*, 14, 295–318.
- Leonard, J. J., & Durrant-Whyte, H. F. (1992). *Directed sonar sensing for mobile robot navigation*. The Kluwer international series in engineering and computer science. Kluwer Academic Publishers, Boston.
- Moore, P. W. B., Roitblat, H. L., Penner, R. H., & Nachtigall, P. E. (1991). Recognizing successive dolphin echoes with an integrator gateway network. *Neural Networks*, 4, 701–709.
- Sillitoe, I., Mulvaney, D. J., & Visioli, A. (1995). The classification of ultrasonic signals using multistage radial basis function neural networks. In Namazi, N. M. (Ed.), *Proceedings of the IASTED International Conference on Signal and Image Processing*, pp. 135–139 Anaheim, CA, USA. IASTED, ACTA Press.
- Sillitoe, I., Visioli, A., Zanichelli, F., & Caselli, S. (1996). Experiments in the piece-wise linear approximation of ultrasonic echoes for object recognition in manipulation tasks. In *Proceedings of the IEEE ICRA*, Vol. 1, pp. 353–359. IEEE Press.
- Sobral, L., Johansson, R., Lindstedt, G., & Olsson, G. (1996). Ultrasonic detection in robotic environments. In *Proceedings of the IEEE ICRA*, Vol. 1, pp. 347–352. IEEE Press.
- Streilein, W. W. (1998). *Neural Networks for Classification and Familiarity Discrimination, with Radar and Sonar Applications*. Ph.D. thesis, Boston University. Available through UMI, Ann Arbor, MI.
- Streilein, W. W., Gaudiano, P., & Carpenter, G. A. (1998). A neural network for object recognition through sonar on a mobile robot. In *Proceedings of the International Symposium on Computational Intelligence in Robotics and Automation*, pp. 271–276.