BROWN UNIVERSITY
Providence, Rhode Island 02912

Institute for Brain and Neural Systems

December 30 1997

Dr. Joel L. Davis, Program Manager
Office of Naval Research
Code 1142 BI
800 N. Quincy Street
Arlington, Virginia 22217-5000

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

Re: Progress Report N00014-91-J-1316

Dear Dr. Davis:

This is written to provide a semi-annual progress report for the contract N00014-91-J-1316 entitled "Neuromorphic Systems: From Biological Foundations to System Properties and Real World Applications." The major goal of our research is to elucidate the biological mechanisms that underlie learning and memory: to find principles of organization that can account both for experimental data on the cellular level and, when applied to large numbers of neurons that receive sensory and/or interneuronal information, for various higher level system properties. Then to apply this in the construction of advanced neural architectures that can be used in practical applications such as mine detection.

Among our detailed objectives are the following: to clarify the dependence of learning on synaptic modification, to elucidate the principles that govern synapse formation or modification, to use principles of organization that can account for observations on a cellular level to construct neural-like systems that can learn, associate, reproduce such higher level cognitive acts as abstraction and computation, and perform in various situations of practical interest.

The approaches employed to achieve these objectives include both theory and experiment. Our ongoing work has lead to a theory of synaptic plasticity that appears to be in agreement with much visual cortex experimental data. In addition, recent experimental work on slice preparations seems to confirm the underlying hypotheses of the theory- the variation of synaptic modification with post-synaptic de-polarization as well as of the movement of the LTP/LTD crossover point as a function of postsynaptic activity. We have applied this biologically based synaptic modification rule to the design of neural-like systems that have proven their value in practical applications.

Advanced neural-like architectures combining the modification rule of Bienenstock, Cooper and Munro (BCM) with back-propagation architectures and wavelet analysis have been applied to the solution of various complex problems.

Among our objectives is the enhancement and application of BCM-type algorithms, dimensional reduction mechanisms and advanced neural architectures to various problems of practical military and commercial interests such as object and speech recognition, fault detection and multi-sensor fusion and mine detection. In addition we have started exploring the use of recent advances in computer hardware such as the Ni1000 to increase processing speed and allow more detailed representations for the purpose of classification.

1 The effect of Synaptic Dynamics on Spatio-Temporal Receptive Fields in Visual Cortex

Temporal dynamics are a feature of all synapses (Zuker, 1989), however the functional consequences of such dynamics are not clear. Recent work has revealed novel aspects of the temporal dynamics of synaptic transmission in neocortex (Markram and Tsodyks, 1996; Abbott et al., 1997). The major findings reveal short term depression in both cortico-cortical (Markram and Tsodyks, 1996; Abbott et al., 1997) and thalamocotical synapses (Stratford et al., 1996; Gil et al., 1996), where the rate of the depression in thalamocortical synapses seems to be larger than in cortico-cortical synapses. The depression is frequency dependent; the steady-state magnitude of the EPSC is approximately inversely proportional to the frequency of stimulation ($\approx 1/\mu$) (Tsodyks and Markram, 1997; Abbott et al., 1997). Further, it has been found (Markram and Tsodyks, 1996) that potentiation enhances the depression. Thus causing 'potentiated' synapses to be potentiated only in response to the first spikes in a moderate (5-50 Hz) frequency range and actually depressed for the rest of the pulses. The properties of these synapses have implications on spatio-temporal properties of cortical RF's and on how established cortical plasticity mechanisms affect the formation of cortical RF's. In this paper, we investigate the effect such synapses have on a model of simple, single cell exhibiting orientation selectivity. We assume that the dynamical of properties of synapses that were investigated in vitro are not significantly altered in vivo (J.Castro-Almanacos and Connors, 1996). Real cortical cells interact with neighboring cells, these interactions may effect their properties, however it has been shown (Ferster et al., 1996) that non-interacting cells already show orientation selectivity similar to those observed in interacting cortical cells. Furthermore, most models invoking cortical interactions in order to sharpen orientation selectivity require a seed of orientation selectivity at the thalamocortical level (Somers et al., 1995; Ben-Yishai et al., 1995).

The properties of cortical receptive fields are experience dependent [For example see review by Katz and Shatz 1996] , and the most likely candidates for the cellular mechanism that underly this plasticity are Long Term Potentiation (LTP) and Long Term Depression (LTD). There is a long standing debate concerning the nature of this change. One view is that LTP changes the presynaptic probability of release (Stevens and Wang, 1994; Markram and Tsodyks, 1996). Another view is that synaptic efficacy is changed (Liao et al., 1995; Isaac et al., 1995; Isaac et al., 1997). If probability of release is altered by LTP, cortical receptive fields may be composed of a structured probability of release, whereas if efficacy is altered by LTP, receptive fields may be composed of a structured efficacy. In this paper we examine the effects of these, two different, possibilities on the spatio-temporal structure of receptive fields in visual cortex.

The model we propose is composed of several components, temporal dynamics of synaptic conductance, assumptions about receptive field structure and assumptions about the input. The equations of synaptic dynamics used are similar to the ones described as in (Tsodyks and Markram, 1997). The amount of available neurotransmitter resources that can be released to the synaptic cleft is reduced by an amount proportional to the probability of release every time an action potential invades the presynaptic terminal, and recovers to full strength with a relatively long time constant. The neurotransmitter released to the synaptic cleft gates the ion channel which potentiates the postsynaptic cell. The proportionality constant between the amount of released neurotransmitters and the current that flows in is called the efficacy of the synapse. The postsynaptic cell is modeled as a leaky integrate-and-fire neuron with a typical membrane time constant of $50ms$. If the postsynaptic cell potentiates above a threshold ($-55mV$) the cell fires. In the visual system, the information is first processed by retinal ganglion cells then relayed to LGN. We used difference-of-gaussian (DOG) filters to capture the center surround filtering of retinal cells. The ON and OFF channel pathways to LGN is also employed and LGN cell firing rates are proportinal to the luminance of the DOG filtered visual stimuli. The cortical receptive fields can be formed by spatial modulation of one or both of the parameters in the synaptic dynamics, namely *synaptic efficacy* or *probability of release.* We examine two extreme cases as to the cellular origin of thalamocortical structure.

- *Case 1:* **Probability of release (PR) Model:** We assume the $e(\vec{r})$ is constant for both ON/OFF channels, $p_r(\vec{r})$ is spatially modulated to form the RF.
- *Case 2:* **Synaptic efficacy (SE) Model:** We assume $p_r(\vec{r})$ is constant for both channels and $e(\vec{r})$ is spatially modulated to form the RF.

Responses of cells to **flashed bars** can reveal more aspects of the temporal dynamics of synaptic conductance. Orientation tuning curves[1] as shown in figure (1) are trial averages of the responses over different time scales. At the onset of the stimulus a rapid increase in the firing rate is followed by a decrease due to synaptic depression. The PR model shows changing orientation selectivity in time. At non-preferred orientations the increase in the firing rate at the onset of the stimulus is small but sustained firing rates are higher than for the preferred orientation. The SE model however, shows consistent selectivity in time due to the fact that the effect of depression is constant regardless of orientation. The reason for the different temporal dynamics, of these models, is that synaptic depression is enhanced by high P_r and by presynaptic firing rate. In the PR model both of these factors are compounded. Thus, when a bar at an optimal orientation is presented, synapses excited by that bar would depress faster in the PR model than in the SE model.

We simulated and analyzed the effect of dynamic synapses on two simple models of orientation selectivity in simple cells in V1. Receptive fields composed of a synaptic efficacy structure show properties similar to those displayed by cortical receptive fields [for review see Orban, 1984]. The orientation tuning curve is unimodal and retains the same prefered orientation over time. Increasing the contrast, causes an increase in firing rates at low contrasts and tends to saturate at high contrasts. For flashed bars we can see that the response has a quick transient peak, the magnitude and slope of the peak increases with the contrast of the stimuli. This peak then decays, it decay

---

[1] An interactive demo program (written with Matlab 5 for unix platforms) can be obtained from http://www.physics.brown.edu/people/artun/publications/. This program contains all source code for interactively generating tuning curves for flashed bars.

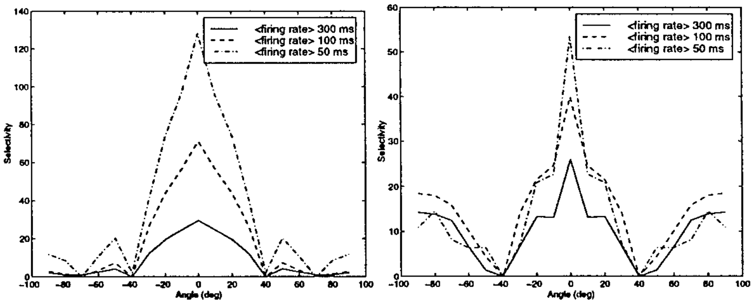

Figure 1: Tuning curves for SE (left) and PR (right) model. Firing rates are averaged over 50, 100, 300 ms for both models. The SE model shows unimodal tuning curves at all times whereas PR model shows bimodal tuning curves in which the angle that produces maximal response changes over time.

faster for high contrast stimuli than for low contrast stimuli. The firing rate decay in our model is indicative of the synaptic dynamics, and depends on the parameters of synapses such as their recovery time. This also resembles the behavior of cortical neurons although in cortical neurons the decay of the response can depend on other factors as well, such as depression of LGN responses for prolonged inputs and on network effects. In contrast, the structured p_r model has bimodal tuning curves that change their orientation in time. We have not found evidence in the literature that such cells exist. It is possible, however, that models in which both p_r and e have spatial structure would exhibit biologically plausible tuning curves.

We have shown evidence that the temporal structure of the response can be used to distinguish between features that can not be distinguished by firing rate alone. Qualitatively similar results have been obtained experimentally (Gawne et al., 1996) and have been shown to enhance the capacity of the neural code (Shouval and Artun, 1997).

2 Analysis of Visual Scenes

Recognition of complex objects involves segmentation, recognition and context. Associated is the recognition/segmentation dilemma: An object is made of features; in order to recognize the object it sometimes helps to recognize the features. In order to recognize features we have to segment an object into parts and identify each part as some feature. But how do we know how/where to segment the object if we don't know what the object is?

Also, recognition often depends on context.

Similarly, an ambiguous word (Figure 3 - left) can be equally well segmented into "l r t e" and "b i t e" in an out of contex situation. In extreme cases, like reading the foreign language script, without the contex we are not able even to start segmentation, Figure 3 - right.

There seem to be problems that can not be resolved at the low level (preprocessing) or identification, no matter how good the preprocessing or identification (of constituent parts - features) is. A possible solution is in interaction of preprocessing with cognitive level (feed back).

The question that we want to answer in this research is how can we employ cognitive information in feed back - feed forward networks to aid in segmentation and identification, and how can BCM and other learning rules (RCE, backpropagation, *etc.*) be combined to construct the various networks.

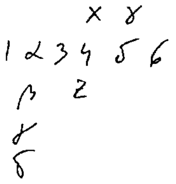

Figure 2: In identification of single characters context plays a role.

Figure 3: Ambiguous word - left, and a word from Cyrillic - right.

In the following sections we will present some results from an experiment with recognition of online cursive script. Some of the advantages of online handwriting recognition are: it is a one dimensional problem; it has wide applications, and the basic problem is the same as in speech recognition. We use a large data set which consists of 1000 words written by 100 different writers which we obtained from David Rumelhart. We next plan to use the developed system to segment and identify visual images.

Conventional feed forward systems, currently in wide use, can be schematically presented in the Figure 4 - left.

Our focus is on interactive, feed forward - feed back, architecture and influence of cognitive level on lower level processing Figure 4 - right.

Our goal is to build a system that uses many average or even below average preprocessing modules and networks but one that *knows* how to use them. If the system is not sure about what the correct answer is, it should be able to locate an area where the error or ambiguity is, examine it more closely or just differently, and then correct the error if possible.

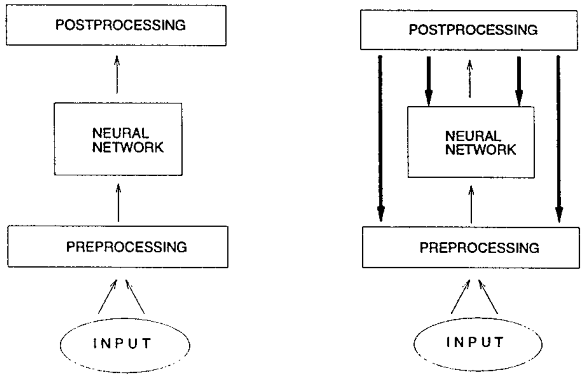

Figure 4: Conventional feed forward and interactive feed forward - feed-back architecture.

2.1 Overview of the system

The system works as follows: First the preprocessor transforms the pattern into a set of strokes. Then, the network that is trained on words, but recognizes individual letters is swept over the preprocessed pattern and the output is a matrix that gives us the confidence of the presence of any letter from the alphabet at any position within the pattern. Each row corresponds to one letter, and each column to the position of a letter within the pattern. This matrix we call "activation matrix". Although the activation matrix provides the information about the position of a letter within the pattern, what we really need is an estimate of the position of the letter with respect to every dictionary *word*. In order to obtain that estimate we have constructed a new network called "positioning network" discussed later.

The next stage is similar to saccadic jumps. The idea is that we don't see all the letters at once, but discover them one by one, depending on their saliency, and in the process of discovering them we recognize a word or build the confidence of recognizing a word. We first choose the number from the activation matrix with the highest activation. From the coordinates of that number we know which letter it is and where it is within the pattern. The value of that number tells us how likely it is that the letter is correctly recognized without the influence of the context information. This part of the algorithm - choosing the highest activation from the activation matrix - is implemented as an WTA algorithm. Then, we focus our attention to the next letter with the next highest activation, *etc.* When ever we discover a new letter it can activate any word from a dictionary.

After all of the letters or group of letters have been discovered from the scene, the following two outcomes are possible: If one of the word-neurons becomes much more excited than other word-neurons we classify it as a correct word. If there are more than one word-neurons with high activations, then it is a signal of possible confussion. In that case we plan to use the feedback module and go back and investigate some regions of the pattern more closely.

2.2 Recognition network

The recognition network that we have built is a multi-layer feed-forward network based on the weight sharing technique, Figure 5. All the neurons of the hidden layer or the output matrix that are in the same row have restricted receptive fields and share the same weights. The consequence of local connections between the neurons and weight sharing is that the number of free parameters is greatly reduced which increases the generalization properties of the network. The other important feature of weight sharing technique is that it builds shift invariance into the system. After the network is trained, the neurons can capture the pattern to which they have became selective no matter where it appears in space since the weights for all the neurons in the row are the same. One advantage of this architecture is that we do not need to segment words into letters in order to train the network on letters.

2.3 Positioning network

In order to build an interactive system, one has to have transparent representation of the word that enables easy detection of errors and possibility to correct them. The current, prevailing method, in recognition of cursive words and sequence analysis, the Hidden Markov Model (HMM), couldn't meet our needs. An error is so much "hidden" in the model, that is hard to locate and correct it. Another drawback of the HMM that we want to overcome, is that it can not be extended to

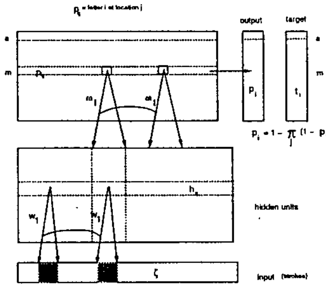

Figure 5: Architecture of the weight sharing network.

recognition of two dimensional objects and our goal is to apply the system to more general problems of scene analysis. Therefore, in addition to recognition network, we had designed a new "positioning network" for estimating accurate position of a letter. The positioning network consists of "position detection neurons". These neurons have receptive fields of different sizes and fire whenever an object is within their receptive field, regardless of the nature of the object. The result of the positioning network is an ordered set of letters. With each letter we associate a number that estimates the "goodness" of its position within the word.

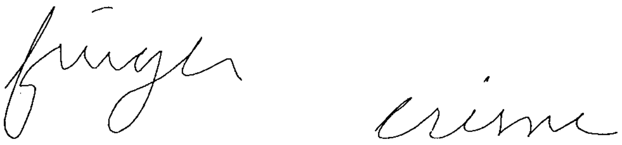

Figure 6: Example of correctly classified word: finger, and incorrectly classified word: crime.

2.4 Some preliminary results and conclusions

For good writers the correct word is in the top 5 words over 97% of the time, and the correct word is the top word over 93% of the time.

For bad writers the correct word is in the top 5 words over 90% of the time, and the correct word is the top word over 70% of the time, which presents an improvement of about 8% since our last report.

Still to be implemented is a system in which postprocessing module combines networks in search for a letter at certain place or gives instructions to the preprocessor to do a different type of preprocessing.

The weight sharing network that we have implemented was convenient in terms of training but, in principle, any network that is trained on isolated characters can be used *e.g.* RCE, BCM *etc.*

We are continuing this work.

3 Denoising of sonar imagery

As part of our effort to construct a integrated system for mine detection, we started investigating various image enhancement techniques which are geared towards a specific classifier[2] that has proven to have excellent results on this data. We started by investigating various wavelet and wavelet packet denoising methods.

3.1 Description of data

The data base we used consists of a 60-image set from a side-scan sonar (SSS0) collected at the Naval Surface Warfare Center (NSWC). They are encoded as 8-bit gray scale images, 1024 range cells by 511 cross-range cells. The 60 images contain 33 targets; some contain more than one target while others contain no targets. Target-like non-targets appear throughout the images. A typical mine-like target consists of a strong highlight on its left side and a long shadow down range on its right side. Unfortunately the presence of clutter can mask this structure.

Real sonar image data is preferred over simulated sonar data because sonar simulations are expensive and do not capture all the critical dynamics associated with actual sonar images.

3.2 Methods

We considered two different denoising methods: a low-pass filter and wavelets. As a low-pass filter we chose a Gaussian filter with $\sigma = 2$. Its dimension has been chosen to be approximately the same as that of a typical mine-like object.

The wavelet denoising we adopted is a combination of two ideas: the more common shrinkage (Donoho, 1995) and the adapted waveform analysis (Coifman and Majid,). It consists in shrinking the wavelet transform coefficients at different scales, each scale corresponding to a different mother wavelet. In the present work we first used a Coiflet-5 mother wavelet, shrinked the coefficients on the finest scale, then a Symmlet-8 mother wavelet, shrinking the coefficients on the subsequent scale. These wavelets have been chosen so that the finest scales correspond exactly to the dimension of the mine-like targets.

3.3 Frequency response

To get further understanding of what the different denoising really do to the images we analyzed their frequency response. Figure 7 depicts the Fourier transform of an original image (top), the wavelet denoised image (center) and of the Gaussian filtered image (bottom) respectively.

We note the presence of very high values in the low frequency domain in the original images. A possible interpretation is the presence of regular periodic structures (sand waves on the sea bottom, trails created by fish nets), or a correlation between pixels due to the data acquisition process. Neither of the two denoising we used have effect on these low frequencies. Both of them act as low pass filters, reducing the values of high frequency coefficients where the noise is supposed to be.

---

[2] Constructed by Dr. Gerry Dobeck from NSWC.

3.4 Testing

In order to test the different denoising techniques we used the Advanced Mine Detection and Classification (AMDAC) algorithm developed at the Naval Surface Warfare Center (NSWC) by Gerald Dobeck (Dobeck and Hyland, 1995). It consists of an improved detection density algorithm, a classification feature extractor, a k-nearest neighbor attractor-based neural network (KNN) classifier, and an optimal discriminatory filter classifier (ODFC).

We chose to concentrate on the detection stage since it is considered to be the most critical. Its purpose is to scan the entire image and identify candidate mine-like regions that will be more thoroughly analyzed by the subsequent classification stages. If a mine-like region is not detected at this stage it won't be possible to recover it afterwards.

Performance summary

	PdPc (%)	FA/Image
Original	91	1.17
Wavelets	97	1.37
Gaussian	91	1.57

Table 1: Performance of the detection stage of the AMDAC algorithm for different denoising techniques.

3.5 Results

Table 1 shows the performance of the detection stage of the AMDAC algorithm for the two different denoising we adopted. It appears that wavelet denoising can increase the number of correct detections, keeping the number of false alarms per image reasonably low. The improvement is around 6% which corresponds to the detection of two mine-like targets formerly missed by the detection algorithm. The Gaussian filter could not improve the performance of the detection algorithm. On the contrary, it increased the number of false alarms per image.

The frequency response of the two denoising methods we tested is qualitatively the same. Both of them act on the image reducing the values of high frequency coefficients. Thus, the difference in their performance is not directly linked to their frequency response but to their ability to retain higher order structure. This topic will be further studied in the near future.

I would be happy to answer any questions you might have concerning this report.

Sincerely,

Leon N Cooper
Thomas J. Watson, Sr.
Professor of Science
Director, Institute for Brain and Neural Science
Enclosure: Publications, Reports and Abstracts.

References

Abbott, L. F., Varela, J. A., Sen, K., and Nelson, S. B. (1997). Synaptic depression and cortical gain control. *Science*, pages 220–223.

Ben-Yishai, R., Bar-Or, R. L., and Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proc. Natal. Acad. Sci. USA*, 92:3844–3848.

Bienenstock, E. L., Cooper, L. N., and Munro, P. W. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience*, 2:32–48.

Coifman, R. R. and Majid, F. Adapted waveform analysis and denoising. In Meyer, Y. and Roques, S., editors, *Progress in Wavelet Analysis and Applications (Proceedings of the International Conference "Wavelets and Applications", Toulouse, France, June, 1992)*, pages 63–76. Editions Frontieres, B.P. 33, 91192 Gif-sur-Yvette, Cedex, France, 1993.

Dobeck, G. J. and Hyland, J. (1995). Sea mine detection and classification using side-looking sonars. *Proceedings of the SPIE Annual International Symposium on Aerospace/Defense Sensing, Simulation and Control*, 2496.

Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627.

Ferster, D., Chung, S., and Wheat, H. (1996). Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature*, 380:249–252.

Gawne, T. J., Kjaer, T. W., and Richmond, B. J. (1996). Latency: Another potential code for feature binding in striate cortex. *Journal of Neurophysiology*, 76:448–453.

Gil, Z., Amitai, Y., Castro, M. A., and Connors, B. W. (1996). Different frequency modulation and $GABA_B$ involvment at thalamocortical and intracortical synapses. *Society for Neuroscience abstracts.*

Isaac, J. T., Crair, M. C., Nicoll, R. A., and Malenka, R. C. (1997). Silent synapses during development of thalamocortical inputs. *Neuron*, 18:269–280.

Isaac, J. T., Nicoll, R. A., and Malenka, R. C. (1995). Evidence for silent synapses: Implications for the expression of LTP. *Neuron*, 15:427–434.

J.Castro-Almanacos, M. and Connors, B. W. (1996). Cellular mechanisms of augmenting response: Short-term plasticity in a thalamocortical pathway. *The Journal of Neuroscience*, 16(23):7742–7756.

Katz, L. C. and Shatz, C. J. (1996). Synaptic activity and the construction of cortical circuits. *Science*, 274:1133–1138.

Liao, D., Hessler, N. A., and Malinow, R. (1995). Activation of postsynaptically silent synapses during pairing-induced LTP. *Nature*, 375:400–404.

Markram, H. and Tsodyks, M. V. (1996). Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature*, 382:807–810.

Orban., G. A. (1984). *Neuronal Operations in the Visual Cortex.* Springer Verlag.

Shouval, H. and Artun, O. B. (1997). Can stochastic neurons support spatio-temporal codes. In *To Appear: CNS*97.*

Somers, D., Nelson, S. B., and Sur, M. (1995). An emergent model of orientation selectivity in cat visual cortical simple cells. *Journal of Neuroscience*, 15:5448–5465.

Stevens, C. F. and Wang, Y. (1994). Changes in reliability of synaptic transmission as a mechnism for plasticity. *Nature*, 371:704–707.

Stratford, K. J., Tarczy-Hornoch, K., Martin, K. A. C., Bannister, N. J., and Jack, J. J. B. (1996). Excitatory synaptic inputs to spiny stellate cells in cat visual cortex. *Nature*, 382:258–261.

Tsodyks, M. V. and Markram, H. (1997). The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc. Natl. Acad. Sci.*, 94:719–723.

Zuker, R. S. (1989). Short-term synaptic plasticity. In Cowan, W. M., Shooter, E. M., Stevens, C. F., and Thompson, R. F., editors, *Annual Review of Neuroscience*, volume 12, pages 13–32. Annual Reviewes, Palo Alto, CA.

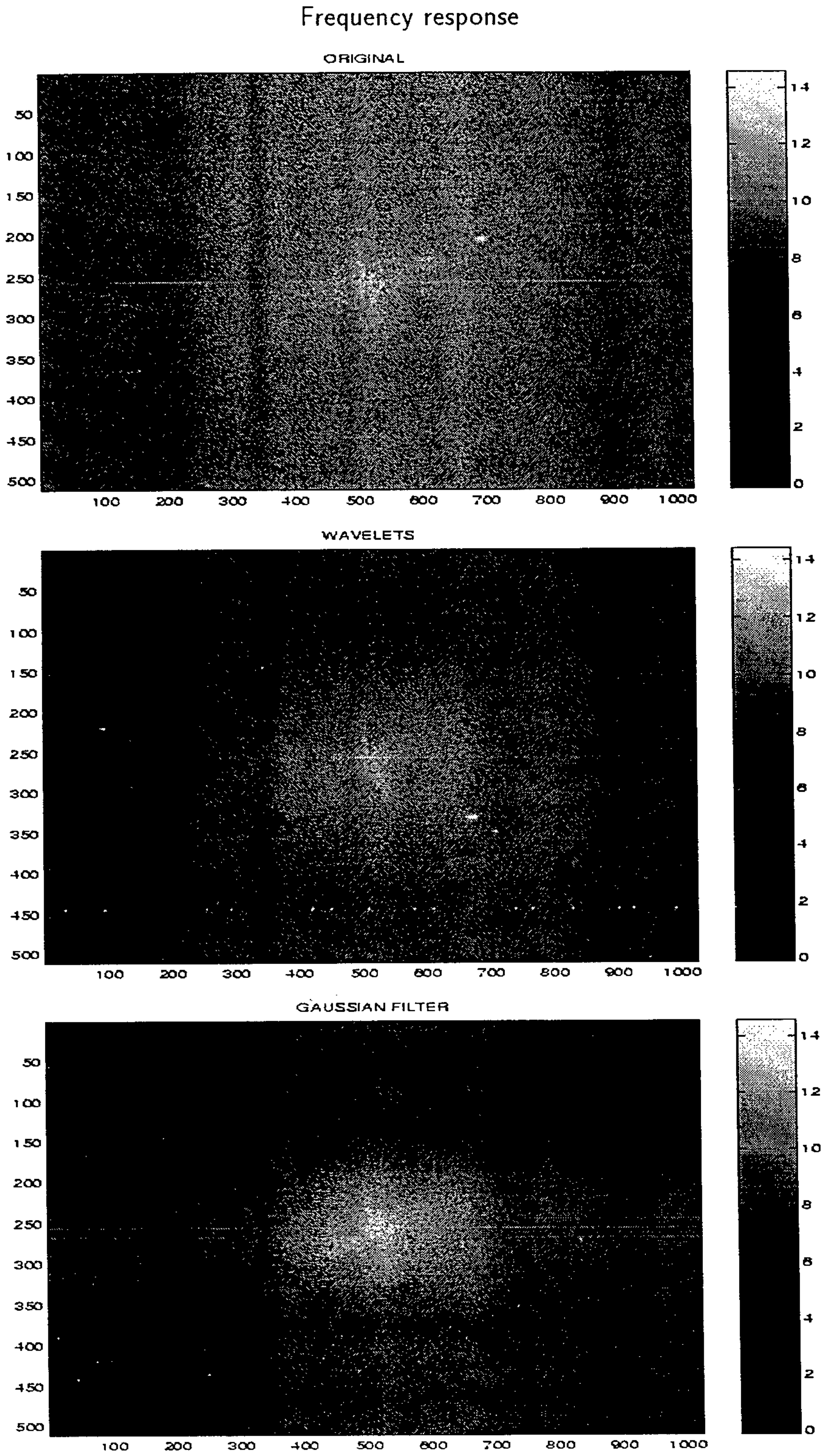

Figure 7: Fourier transform of an original image (top), the wavelet denoised image (center) and of the Gaussian filtered image (bottom).