Part III.
EXPERIMENTAL DESIGN, ANALYSIS, and CONCLUSION

Part III of this dissertation describes the design and then the results of the experiments that were developed to compare the core model and human performance. For each of the two studies, the design process consisted of the specification of a series of images that varied with respect to the important physical variables chosen for study as well as the development of instructions and an interface for conducting the human observer experiment. In Chapter 7 statistical and graphical methods are used to analyze and draw conclusions about the conformance of the model results with the human data. The conclusion of this work restates the philosophy motivating this research and talks about its implications in light of the experimental conclusions. The discussion furthermore chronicles what was learned and suggests how that might influence future development.

6. EXPERIMENTAL DESIGN

The purpose of this research was to compare the performance of a model of human vision with that of its human counterpart for two important medical image estimation tasks. This entire work strives to construct two relevant, meaningful experimental bases for making that comparison. An initial step, described in the preceding chapters, involved a choice about imaging systems and their accompanying parameters and estimation tasks. The task protocols and image production methods were carefully chosen to lend any results external validity. Following those commitments, what remained was to devise an experimental design to test the essential hypothesis. This chapter details the particular choices of values for image property parameters and describes and attempts to justify the many aspects of the human observer experiment.

For both experiments a series of images were required for which the humans and the model could provide their respective estimates. As described previously, the angiography images were to vary with respect to stenosis depth, blur, and noise. The portal images possessed a natural variation in the distance between the treatment beam and the vertebral body edges and were processed by varying two SHAHE parameters. One would like to believe and have tested that the model might be predictive of the human for *many* physical parameters of the system or processing method. That two were chosen was simply a matter of limiting the scope of the human observer study; more parameters would have subjected human observers and their experimenter to a lengthy study. In each case, three levels of each of the parameters were chosen: it was necessary to choose just enough levels to capture non-linear trends. Finally, because there were many human subjects but only one model, it was important to incorporate into the experimental design sufficient variation in backgrounds such that there were enough observations from the model at each of experimental conditions.

Medical students were used as the observers. It is recognized that the greatest validity might have been achieved by employing certified radiologists and radiation oncologists. However, academic radiologists are busy and difficult to recruit. It was not feasible to use these specialists because the number of observers required to demonstrate an effect, be it positive or negative, with sufficient statistical power was too large. There are several reasons why student observers were acceptable. The two tasks are essentially visual, relying upon little or no learned interpretation. Medical students know human anatomy and are at least familiar with the principles of radiographic imaging. Furthermore, the model tested is a model about fundamental human vision. It does not attempt to embody any of the cognitive abilities that radiologists might somehow bring to bear on these tasks. The comparison was never intended to be one based on interpretive skills but instead upon fundamental visual tasks.

The parameter level decisions were based on extensive pilot research. The levels had to be ones that were sufficiently "spaced" to allow observer variation across them. It would not have been interesting to choose levels that were so similar that they caused little variation in performance. Conversely, levels could not be chosen that were at the extreme end of the range of possibilities. Images acquired or processed with those parameter values would be considered clinically unrealistic, and there would be no question about their lack of quality. Also, such degraded conditions could cause the model to fail to compute an estimate as described by the criteria in Sections 4.4 and 5.4. It can be the case that the ridge flow process (Section 3.3) that proceeds from an initial guess may fail to converge to a position in that region of scale space that satisfies the conditions for the ridge

definition. In the absence of a core, the protocols for determining a stenosis or distance estimate by the model can not be conducted. Section 4.4 attempted to characterize those failures for the angiography task in order to qualify the applicability of the model. Section 6.1 below mentions how these estimate failures were treated during the image generation process. In the portal imaging experiment, in the rare cases where core formation failed an alternative measure of gap object width, the width corresponding to the scale of maximal medialness at the gap object center (see Section 5.4) was produced. For the portal images, it was not so much a matter of choosing parameters within a range in which the model could compute but of remaining in a clinically-useful subset of the parameters where essential anatomical structures could be recognized visually. The pilot studies were important in determining parameter values for both experiments that reflected these computational and subjective considerations.

Observers were from first to fourth-year medical students and radiology technology students at the University of North Carolina-Chapel Hill. Each observer possessed (possibly corrected) normal eyesight. All observers signed a consent form and were paid $10.00 for the angiography study or $15.00 for the portal imaging study for what was roughly a two-hour experiment.

Observers viewed the monitor from roughly one-half meter, but that distance was not controlled. The ambient luminance of the moderately darkened room was 20 lux, and the mean luminance of the display was 16 fL. The monitor used in the experiments was perceptually linearized based upon the monitor luminance measurements together with the CIELUV luminance model, an international standard defining the relationship between displayed and perceived intensities.[1]

After the parameter levels, described next, were chosen and the images were generated as described in Sections 4.3 and 5.3, the human observers and the model both "operated" upon the same image sets. The majority of this chapter pertains to the psychophysics of the delivery of these images and collection of responses in the human experiment.

6.1 Angiography Experiment Conditions

The intention for the angiography design was to choose a set of orthogonal parameters so that each experimental condition could be described by one of three levels of each of three parameters. In an attempt to do that, images were generated according to the following specifications (Figure 6.1). Stenosis depths of 25, 50, or 75 percent were used in the creation of simulated vessels. Images were Gaussian blurred with standard deviations of 1.0, 2.5, or 4.0 pixel units on vessels that were 24 pixels at normal width. Finally, images were scaled in three different ways prior to the addition of Poisson noise: pixel greyscale intensity ranges from 64 to 200, 350, or 500 were used to create successively less noisy angiograms.

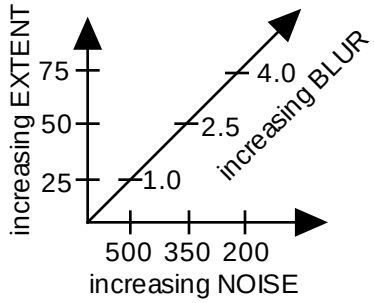
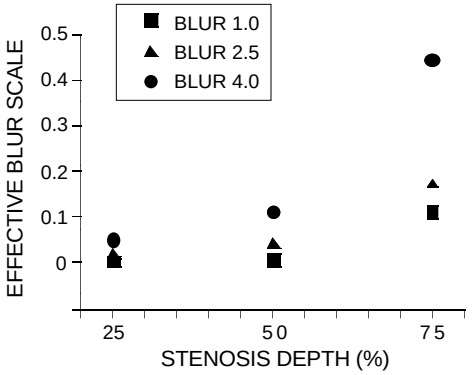


Figure 6.1. Angiography experiment parameter values. Simulated angiograms were produced according to 27 experimental conditions. The three parameters studied (depth, noise, and blur) were varied along three levels as shown.

In light of the recent development of principles for describing blur and noise levels, it becomes apparent that the design of this experiment is anything but orthogonal. Sections 4.3.3 and 4.3.4 describe how our hypothesis of the zoom invariance in the visual system's perception of objects has motivated object-relevant measures of blur and noise. The result is that if one measures the blur and noise conditions with the metrics proposed in those sections, there turn out to be many different blur and noise conditions present in the angiography design. While the parameter levels used to create the experimental images defined an orthogonal relationship, the resulting descriptive measures do not.

[1] B.M. Hemminger, R.E. Johnston, J.P. Rolland, K.E. Muller, "Perceptual Linearization of Video Display Monitors for Medical Image Presentation," Medical Imaging 1994: Image Capture, Formatting, and Display SPIE 2164 (1994): 222-240.

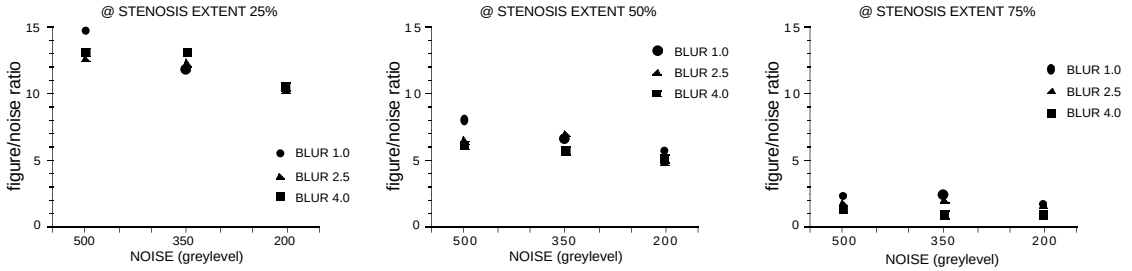
In Figure 6.2, the effective blur scales (EBS's) calculated via Equation 4.8 are plotted as a function of stenosis depth with parameter labels that are the original Gaussian standard deviations used to blur the images. The EBS values are not based on measurements on images but can simply be calculated from the known stenotic half-widths of the vessels and the three levels of the standard deviation parameter used to vary the acquisition blur. Larger values of the EBS parameter correspond to more blur.



	DEPTH 25%	DEPTH 50%	DEPTH 75%
BLUR 1.0	0.0031	0.0069	0.1111
BLUR 2.5	0.0193	0.0434	0.1736
BLUR 4.0	0.0494	0.1111	0.4444

Figure 6.2. Effective blur scales for the angiography conditions.

The actual figure-to-noise ratios (FNR's), from Equation 4.10, are presented in Figure 6.3 using the original labels for the 27 experiment conditions. The estimates are means from the FNR's computed for the six simulated angiograms at each condition. Smaller values of FNR should correspond to the noisiest conditions.



	DEPTH 25%			DEPTH 50%			DEPTH 75%		
	B 1.0	B 2.5	B 4.0	B 1.0	B 2.5	B 4.0	B 1.0	B 2.5	B 4.0
NOISE 200	10.34	10.26	10.45	5.70	4.89	5.06	1.71	1.55	0.96
NOISE 350	11.80	12.30	13.02	6.60	6.98	5.60	2.35	1.95	0.78
NOISE 500	14.67	12.64	13.05	7.97	6.51	6.09	2.26	1.81	1.35

Figure 6.3. Figure-to-noise ratios for the angiography conditions.

Both measurements reflect very strongly the effect of figure width: the conditions with most constricted widths have higher effective blur scales and lower figure-to-noise ratios. Naturally, the EBS values are correlated with the original Gaussian standard deviations. Likewise, the FNR's as expected decrease as the value for the intensity scalings used to generate the noise decreases.

CONDITION	FNR	EBS	BACKGROUND					
d25, b1.0, n200	10.34	0.0031	11	26	25	15	23	16
d25, b1.0, n350	11.80	0.0031	6	7	12	5	4	21
d25, b1.0, n500	14.67	0.0031	27	3	24	22	26	6
d25, b2.5, n200	10.26	0.0193	13	6	14	19	24	11
d25, b2.5, n350	12.30	0.0193	14	21	10	3	13	23
d25, b2.5, n500	12.64	0.0193	5	4	3	13	27	8
d25, b4.0, n200	10.45	0.0494	15	5	1	12	22	2
d25, b4.0, n350	13.02	0.0494	1	16	19	4	5	13
d25, b4.0, n500	13.05	0.0494	4	8	18	20	11	17
d50, b1.0, n200	5.70	0.0069	17	19	22	18	25	12
d50, b1.0, n350	6.60	0.0069	3	25	7	16	15	5
d50, b1.0, n500	7.97	0.0069	19	27	9	23	3	25
d50, b2.5, n200	4.89	0.0434	10	2	6	21	19	3
d50, b2.5, n350	6.98	0.0434	23	18	4	7	16	14
d50, b2.5, n500	6.51	0.0434	12	22	17	9	18	20
d50, b4.0, n200	5.06	0.1111	22	10	13	6	14	15
d50, b4.0, n350	5.60	0.1111	21	9	20	14	12	27
d50, b4.0, n500	6.09	0.1111	8	15	16	24	1	10
d75, b1.0, n200	1.71	0.1111	20	23	26	27	6	22
d75, b1.0, n350	2.35	0.1111	25	12	23	2	9	4
d75, b1.0, n500	2.26	0.1111	24	17	21	1	8	7
d75, b2.5, n200	1.55	0.1736	7	24	15	10	21	1
d75, b2.5, n350	1.95	0.1736	16	14	8	11	20	26
d75, b2.5, n500	1.81	0.1736	2	11	27	26	10	24
d75, b4.0, n200	0.96	0.4444	26	20	11	8	7	19
d75, b4.0, n350	0.78	0.4444	9	13	2	25	17	18
d75, b4.0, n500	1.35	0.4444	18	1	5	17	2	9

Table 6.1. Conditions for angiographic image sequence. Each of the 27 experimental conditions (leftmost column, "d"=stenosis depth, "b"=blur, and "n"=noise, or equivalently FNR and EBS) was randomly assigned six of the possible 27 backgrounds (a patient anatomy and simulated vessel).

To generate the sequence of experimental images, a set of 27 digitized and diffused scout angiograms were each assigned a different simulated vessel path. Six images were obtained at each of the 27 (depth * blur * noise) experimental conditions by applying each of those conditions to *six* (from among the 27 possible) angiographic backgrounds. That number of background images at each experimental condition resulted in a set of 162 different images that in turn allowed a reasonable human observer experiment duration. The images (see Table 6.1) were a subset of all of the (729) possible combinations of the 27 backgrounds and the 27 conditions. It was these 162 images that were judged by the humans and for which core model computations were simultaneously performed.

The images utilized in the experiment were ones for which a legitimate stenosis estimate derived from the vessel core was obtained at each of the 27 conditions applied to the image. Prior to the generation process, separate pools of patient backgrounds and simulated vessels were created. During the generation process, an unused background/vessel pair was chosen randomly from the pools for which stenosis estimates had not yet been computed. In the event that stenosis estimates for that particular background were successfully computed at the 27 depth/noise/blur conditions, it was used as an experimental image. Otherwise, if an estimation failure occurred at any point during the processing of the 27 conditions, the background and vessel were returned to the pools, that background/vessel pair was marked as a failure, and some other combination was subsequently attempted. The process was halted when 27 backgrounds were obtained for which estimates were successful at all 27 conditions. Section 4.4 discusses the potential causes of the computation failures. Most of them occurred at the highest noise and blur conditions. The generation process created 729 images--the random selection of the subset of 162 was not performed until after all the images were produced.

6.2 Angiography Human Observer Experiment

The width judgments that contribute to a stenosis estimate are described in Section 4.2. In clinical practice, the radiologist may use a ruler to determine those measurements, but it is just as common that he/she will simply make an "eyeball" estimate.[2] Putting a ruler down upon the vessel and marking, visually or physically, the intersection of the vessel boundary with the gradations on the ruler is strictly a one-dimensional task. The core model that was tested in this dissertation is a model for the mechanisms of human shape perception. There is little

[2] M. Mauro, personal communication, April 1993

hope that it might be predictive of human performance for the task where the ruler is employed and the shape of the vessel is never contemplated. Therefore the human estimates examined in this dissertation were entirely visual, and no superimposable measuring device was allowed.

A means of providing an estimate about stenosis depth without the benefit of the scale or basis that the ruler provides was thus needed. Many observers, particularly non-radiologists, might not have possessed a visual knowledge or recognition of percent stenosis magnitudes. Furthermore, even if an observer claimed to know a, for example 50 percent, constriction when he/she saw it and might repeatedly recognize such with consistency, that observer could be biased, i.e., incorrect by some consistent amount. A measurement tool was developed that could allow the observer to make an estimate of the stenosis without any prior conceptions about, or potential biases for, numerical concepts of percent stenosis (Figure 6.5). A prototype vessel was provided whose depth of stenosis could be *adjusted* until it matched, to the user's perception, that of the angiogram in question. The percent stenosis estimate was in turn calculated from the normal and constricted width of the prototype. No attempt was made to match the vessel path of the prototype to the path of the corresponding vessel to be judged: the prototype vessels were perfectly straight. The prototype vessels were generated by the same means as the vessels in the angiographic simulations (Section 4.3.2), so they possessed the same intensity fall-off and constriction profile. They were displayed in white on a grey background.

An important characteristic of the estimation tool was that the width along the normal portions of the prototype vessel was varied from trial to trial (Figure 6.4). It was the case that the width-at-normal of the vessels in the simulated images remained constant. If the normal width of the prototype vessels were also constant on every trial, then the observer could simply adjust the prototype stenosis such that width at the most constricted portion of the prototype stenosis matched that in the most constricted portion of the vessel in question. This situation is not only not clinically realistic, but again reduces the task to one that is not dependent upon shape properties of the vessel. The varying width of the prototype vessel forces the observer to take into account both constricted and normal width in both vessels and thereby produce a relative, ratio judgment. Five prototype vessel widths were used: one from -25, -12.5, 0, 12.5, and 25 percent of the vessel width (24 pixels) in the simulated images was chosen randomly as the prototype vessel width on each trial. The initial depth of the stenosis in the prototype vessel was chosen randomly each time. The observer simply had to adjust the prototype from whatever its initial position happened to be.

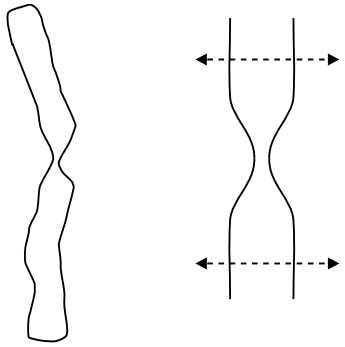


Figure 6.4. Prototype vessel width variation. The prototype vessel (right) was to be adjusted by the observer until its relative stenosis depth matched that of the vessel on the left. The width of the normal portions of the prototype vessel was varied from trial to trial in order to force the observer to compare the shapes of the vessels and not just their widths at the constrictions.

The human observer experiment was conducted in a window on a computer screen (Figure 6.5). On each trial, the simulated angiogram of interest appeared in the window on the left. Observers were asked to use left and right arrow keys on the keypad to adjust the depth of stenosis on the right such that its percent stenosis matched that of the vessel in question. The space bar was used to advance to the next trial but could not be depressed until some adjustment had been done to the prototype vessel stenosis. The position of the stenosis was indicated with an arrow.

Thirteen observers participated in the study. The observers were provided with a lengthy explanation of angiography and the task of interest, along with specific directions about indicating their responses. They participated in 36 practice trials without feedback with images that were not in the experimental set: the levels of noise and blur were from the same levels as those examined in the experiment, but stenosis depths and backgrounds were different.

The observers viewed a total of 182 images. Because only three levels of actual stenosis were displayed throughout the experiment, twenty "ringers" with stenosis depths of 20, 40, 60, or 80 percent were inserted among

the 162 experimental images. Each observer viewed the 182 images in a randomized order. Data were collected by the computer program throughout the experiment and were output in a textfile for subsequent analysis.

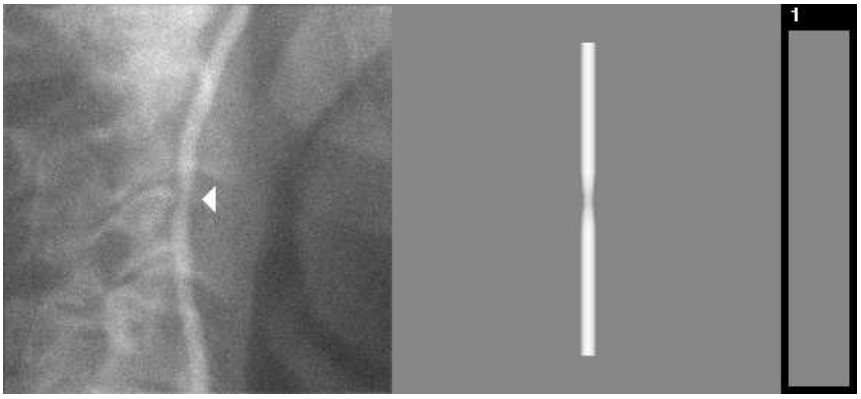


Figure 6.5. Angiography experiment window. Observers used the keypad to adjust the depth of stenosis in the prototype vessel in the image on the right until it matched that in the simulated angiogram on the left.

No feedback was provided. For images that were especially degraded by blur or noise and the vessel, at the constriction, seemed to practically "disappear" or "fade out," observers were instructed simply to make their best inference about the underlying depth of the vessel.

Observers reported that they were comfortable with the relative ratio task. A number of them reported making a "global" or "overall" shape comparison between the two vessels. In those cases they reflected that they regarded the prototype vessel as a magnified or minified replica of the vessel in question when making the comparison, as opposed to adjusting the local stenotic width by the factor needed to reconcile the normal widths.

6.3 Portal Imaging Experiment Conditions

The portal imaging experiment studied the variation of two SHAHE parameters along three levels for a total of nine experimental conditions. Unsharp masking gain parameter values of 1, 3, or 5 were studied. The resulting preprocessed images were processed with contrast limitation values of 2, 7, or 12 (Figure 6.6).

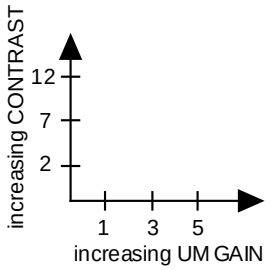


Figure 6.6. Portal Imaging experiment parameter values. Digitized portal images were processed according to these nine different SHAHE parameter settings.

Eighteen backgrounds at each condition were chosen randomly from among 27 possible backgrounds. The backgrounds were processed with SHAHE (Section 5.1) according to the parameter values specified by the condition (Table 6.2). A full factorial of the 27 backgrounds with the nine SHAHE conditions would have created a lengthy experiment for the humans, so a random subset of 18 of the 27 backgrounds were used at each condition. The set of 162 images were randomly presented to each of the subjects in the human observer experiment, and model calculations were performed for each of them as well.

COND.	BACKGROUND																	
g1, c2	25	4	22	26	24	15	23	10	14	9	16	3	8	17	1	7	12	11
g1, c7	14	2	10	6	8	22	21	15	26	24	5	12	19	9	17	25	27	1
g1, c12	24	6	20	19	10	16	18	9	23	22	3	21	4	13	14	2	5	26
g3, c2	10	15	18	8	3	4	20	6	25	21	22	26	9	16	19	13	17	7
g3, c7	6	9	3	12	27	24	13	14	20	7	15	18	2	11	16	10	22	17
g3, c12	21	16	25	11	20	23	4	19	2	12	7	15	3	24	13	22	14	18
g5, c2	1	25	15	21	13	5	12	20	6	16	18	8	23	27	7	4	26	2
g5, c7	5	21	27	2	19	18	1	23	17	13	4	24	20	14	6	9	11	25
g5, c12	12	19	1	20	26	14	15	2	8	18	25	11	22	6	4	5	13	3

Table 6.2. Conditions for portal image sequence. Each of the nine experimental conditions ("g"=gain, "c"=contrast limitation) were assigned 18 of the possible 27 backgrounds.

6.4 Portal Imaging Human Observer Experiment

The portal imaging experiment in which the human observers participated consisted of the presentation of each of the images in the set and the collection of a distance estimate from the observer on each trial. As described more thoroughly in the previous chapter, the observer's task for this experiment was to measure the distance between the treatment field edge and posterior edge of the cervical vertebral bodies. As was the case with the angiography experiment, a ruler or other measuring device placed on top of the image was not allowed. Aligning the gradations on a ruler with the two edges and calculating the distance between them is not a shape judgment. Instead, a measuring tool was provided for the user adjacent to the image to be judged that could be adjusted to indicate the distance. The tool consisted of a horizontal line that could be adjusted until it was as long as the distance between the edges. The observer used the computer mouse to grab either end of the line and drag it to lengthen or shorten it. The ends of the line (relative to the border of the background in which the line was placed) need not have had any positional correspondence to the location of the edges in the real image; it was simply necessary for the observer to make the length of the line match the edge-to-edge distance. The initial length and position of the line on each trial were random--the observer adjusted its length from however it first appeared. The experiment was conducted in the window shown in Figure 6.7.

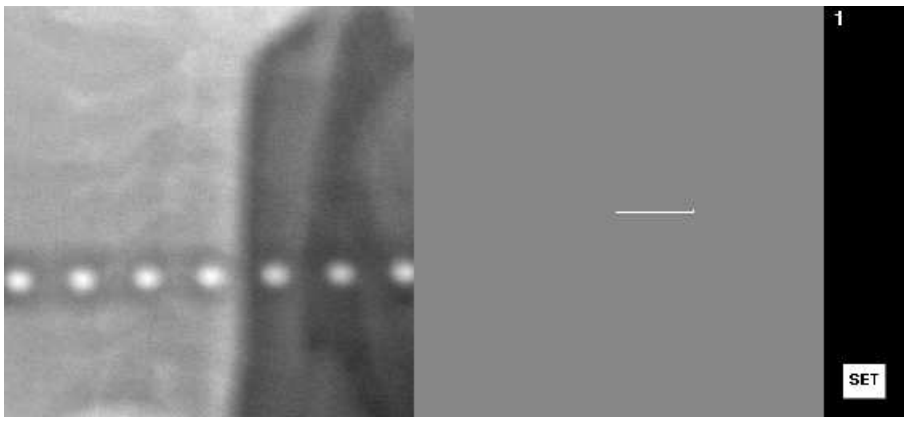


Figure 6.7. Portal imaging experiment window. Observers used the mouse to adjust the length of the line in the window on the right until its length matched the distance between the treatment field edge and the posterior edge of the vertebral body in the image on the left at a height corresponding to the tool.

The vertical position of the distance tool, which was always precisely vertically in the middle of its background, served to indicate to the user the corresponding vertical position in the adjacent image. The patient's anterior anatomy was always positioned to the right in the image, and the treatment field was always located anterior to the vertebral bodies, so the judgment on each trial was between the treatment field edge on the right and the vertebral body edge on the left.

Very specific instructions had to be provided as to which aspects of the two edges to use in the distance decision. The treatment field edge is very broad. Furthermore, its characteristics were inevitably changed by the image processing (see Figure 6.8). The edge can be made to appear broader under certain conditions that cause the CLAHE artifact referred to in Section 5.1. Alternatively, SHAHE can sometimes, if the unsharp masking is done properly, improve the sharpness of the edge. Observers were instructed and subsequently trained to identify the sharpest break in the beam edge region. Identifying the treatment field edge in each image was never a

problem; it was always quite prominent. The real task involved determining where in its broad profile to use as its precise location.

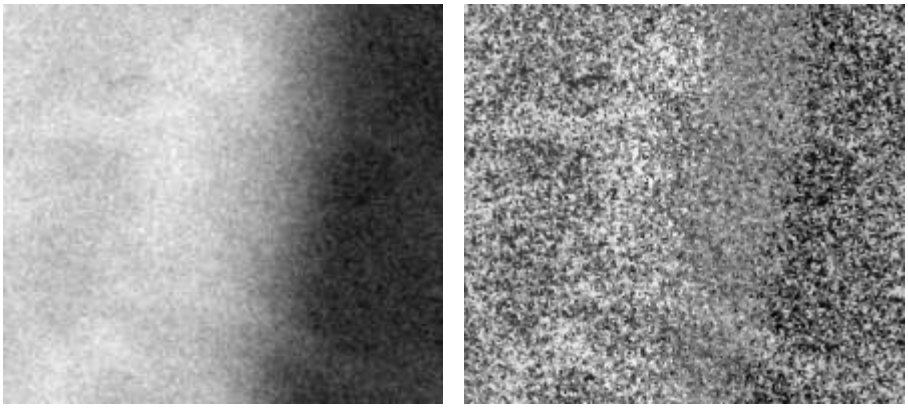


Figure 6.8. Treatment field edge appearances. The parameter settings to SHAHE could significantly influence the shape and intensity characteristics of the field edge. The images above are the same image processed with two different SHAHE parameter adjustments. Observers were instructed to use the exact position of the sharpest break between the dark beam interior on the right and the lighter beam edge band on the left.

In the lateral view, the posterior edge of the vertebral body appears as a thin white vertical line. The treatment field edge was always nearly parallel to the vertebral body edge, so finding the more prominent beam edge was often helpful in finding the vertebral body edge. In the case that the vertebral body edge possessed much of a width, the observers were instructed to use its posterior (left) edge.

Eleven observers participated in the experiments. They first worked through 39 practice trials. On each trial, before adjusting the measuring tool, the observer was required to use the cursor to point to the two edges he/she was using for the distance judgment. All edge decisions were verified and explained by the experimenter. Each of the backgrounds that were processed differently throughout the experiment were viewed at least once during the practice so that the essential structures for each background could be identified in the presence of the experimenter. To add some additional clinical variability, 24 "ringers" that contained different patient backgrounds, were inserted randomly throughout the image sequence.

The observer was instructed to call upon the experimenter at any time if he/she had any doubt as to which edges in the image to use in the distance judgment. In those cases, the experimenter indicated in a general but useful way the edges in question. The mouse cursor was disabled (after the practice trials) in the image window so that it could not be used to point to the edges or use it in any way in making the judgment. No feedback was given regarding the correctness of the distance estimates in the actual study; observers were simply instructed to make their best estimate.

Observers had the unforeseen opportunity to use the borders of the measuring tool background to aid in the estimation of the distance. That is, the observer could position the ends of the line at the same relative distances from borders of the background as those of the vertebral body and treatment field edges in the actual image. However, observers rarely reported using this strategy. They instead preferred the intended approach of matching the length of the line to the distance between the anatomical and field edges. This is further evidence for the human representation of gap objects and the plausibility that the core model may be used to determine treatment field clearance by computing gap object width.

6.5 Summary

There were two important considerations in the design of the experiments in this research: experimental conditions and human psychophysics. Exploratory pilot research, which studied both human and model performance over a broad range of potential parameter value possibilities, provided an estimate of a subrange of the values in which the human was systematically variable and at the same time in which the model could provide usable estimates. Any additional leeway in choices was influenced by the clinical relevance of the settings. Together, this information was used to decide upon three levels of each of the parameters designated for investigation.

The understanding that blur and noise ought to be quantified via object-relevant measures followed the design and execution of the angiography experiments. Were it to be done again, a more straightforward design and analysis would be achieved by choosing discrete levels of the EBS and FNR parameters. The relevant production parameters could be adjusted to produce images that resulted in one of the levels of these *descriptive*

parameters. For example, the intensity scalings could be done in such a way as to obtain an FNR that closely matched one of three chosen levels. That way an orthogonal design is created where each condition truly can be described by one of a few levels of each parameter. While the principles reflected in the object-relevant blur and figure-to-noise ratio measures have only recently been espoused, this research can only be strengthened by adopting, however late in the process, these appropriate metrics.

The experiments incorporate a great deal of variability, for the sake of clinical validity, in the conditions that were not controlled for. Specifically, many patient anatomies and simulated vessel paths were chosen in order to lend any results about the efficacy of the visual model optimal generalizability.

Most importantly, it was imperative to develop a clean psychophysical experiment for the human observers so that the study could produce meaningful results that could in turn be the standard against which the model was tested. A great deal of consideration was given to the design of the measurement tools that were used by the observers to make an estimate. Those tools allowed visual estimation of the quantities of interest by observers who were not familiar with the underlying numerical representation of the estimates. At the same time, the tools required the observer to perform a shape judgment, a task that was comparable to the core model's assessment. Other aspects of the experimental design were dictated by standard and sensible psychophysical practices. The conclusions drawn from the statistical analyses in the following chapter can hopefully be made with confidence that the data upon which they were based were collected from a clean and unconfounded experimental design.