DOCUMENT MANIPULATION AND TYPOGRAPHY

Proceedings of the International Conference on Electronic Publishing, Document Manipulation and Typography, Nice (France) April 20-22 1988

Edited by J.C. van Vliet

EP88

INRIA

THE CAMBRIDGE SERIES ON ELECTRONIC PUBLISHING

IMAGE PROCESSING ASPECTS OF TYPE

ROBERT A. MORRIS

The University of Massachusetts at Boston and *Interleaf, Inc.*

ABSTRACT

Classical image processing models can be applied to individual letters and to text as a whole. By combining these models with contemporary models of human vision, some aspects of type design can be considered to reflect needs of the visual system.

1. Spectra

Type, both black-and-white and gray-scale, can be regarded as a function mapping points in the plane into intensities, which are numbers between 0 and 1. Since (western) type is made up largely of vertical strokes, we can sometimes conveniently represent it as a one-dimensional signal, with the intensity at horizontal position *x*, considered as the average intensity along the vertical line at *x*. In either the one- or two-dimensional case, it is possible to talk about the *spatial frequencies* present in the signal and about the amplitude of the image at each frequency. These spatial frequencies characterize the image and can reflect features of the type, detailed below. In addition, the human visual system has different responses to different spatial frequencies, and this affects the way we see type (see Section 2.).

1.1. *Generalities*

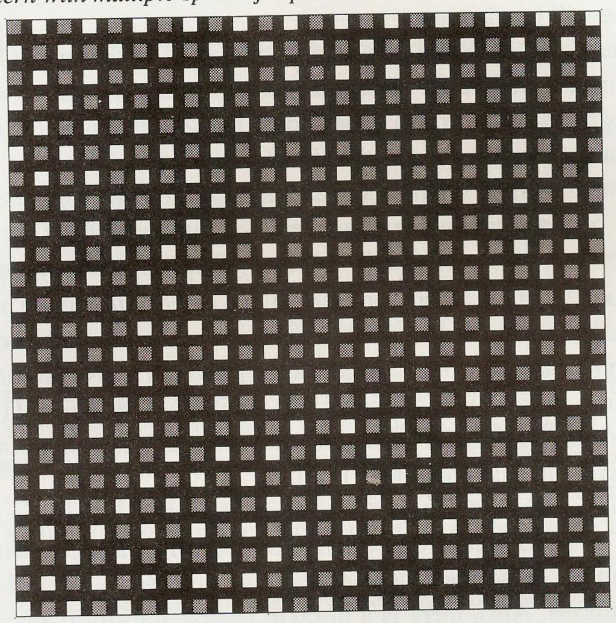
In the one-dimensional black-and-white case, spatial frequency can be conveniently thought of as the rate of alternation of the strokes between black and white, per unit distance. This distance is most appropriately measured not in a fixed linear measure, but rather in the amount of visual angle subtended by the image under study. Intuitively, this is reasonable: 100-point type viewed at 400 cm. will subtend the same visual angle as 10-point type at the normal 40 cm. reading distance. If we had no visual cues about the viewing distances, we would judge these to be the same size type (ignoring for now the important typographic principle that type should not be linearly scaled). For example, 10-point type at 40 cm. alternates between black and white at the rate of about 7–8 cycles per degree (*cpd*) of visual angle. (This can be confirmed with a ruler and simple trigonometry on the reader's favorite 10-point type.) This spatial subtense measure is also appropriate for the two-dimensional case, which we discuss later.

Even when we are viewing essentially analog type, such as might be produced by ink which flows on the page (as opposed to the discrete marking produced by laser printers or digital photo-typesetters on their original film), the visual system is *sampling* the image we see. This is because the image is perceived early in the process by discrete photoreceptors on the retina. The distance between these receptors is a limiting factor in the resolution with which we can perceive an image. It is known to be about 1 minute of visual angle, corresponding to a maximum perceivable frequency of 60 cpd. Some of the implications for digital typography of these limits were observed in [Bigelow83]. Omitting considerations of half-toning, in which multiple one bit pixels make one gray cell, this limit implies that pixels about 1/600-th inch apart can not be distinguished even at the minimum distances at which one can focus. Thus, a resolution of 1200 dots per inch (dpi) permits the finest distinguishable alternation of white with black, and, indeed, digital typesetters traditionally are manufactured at about this resolution.

Sampling continuous signals, either in the marking process or in the vision system, creates several well-known problems for whatever system is reconstructing the image. In the computer graphics literature, all of these are coloquially known as *aliasing*, but the major one is simply the roundoff error inherent in making a discrete choice from a continuous signal. This is most evident in the familiar ''jaggies,'' or staircase effect, seen in curves and some diagonal lines. Aliasing results when two signals become indistinguishable due to undersampling. Both artifacts can be dealt with by increasing the sample rate, i.e., the resolution, but, in theory, never completely (see Ch. 12 of [Castleman79] for a treatment of the tradeoffs in dealing with sampled signals). Any signal with sharp edges is guaranteed to have components of arbitrarily high spatial frequency.

A famous theorem of signal processing theory, the Shannon–Nyquist sampling theorem, asserts that aliasing is not possible *for a band-limited signal* (one with an upper bound on frequencies present) when the signal is sampled at a rate at least twice that of the highest frequency. Unfortunately, any signal with sharp edges is never band limited. Increasing the sample rate of any signal can move the aliasing artifacts to frequencies not perceivable, but the ''anti–aliased'' gray-scale fonts in some use are smoothing the jaggies principally by reducing the roundoff error in the image's representation. In addition, some recent vision research ([Yellott84]) cpd has shown that the visual system is in fact not subject to as much aliasing as

Figure 1: *Pattern with multiple spatial frequencies*



might be expected from the 60 cpd limit on its sampling rate. This apparently is due to the fact that the sampling is not uniform, a ''technique'' also applied in sampling oscilloscopes, whereby periodic signals of higher frequency than the instrument's sample rate can be reconstructed by sampling at different points in successive cycles. We will not touch much on aliasing issues in type, nor directly on gray-scale fonts, which only recently have been modeled with sufficient rigor to advance beyond elementary stages ([Naiman88]).

It is easy to understand that the visual system has differing responses to different spatial frequencies by consideration of an image such as in Figure 1, due to Kirkham ([Sekuler85], p. 169). There are actually patterns at two spatial frequencies present, but the lower one, three extra clusters of dark dots, will become visible to most viewers only if they squint, which has the effect of filtering out the higher spatial frequency, making visible the pattern at the low frequency.

It is important to note that no enhancement of the low frequency has taken place, but rather only an attenuation of the higher one. An even more dramatic illustration of these issues can be had by viewing the figure from a distance of several feet instead of at arms length. This has the effect of raising both spatial frequencies, to the point where the response to the higher one is negligible compared to the response to the lower. The higher frequency pattern will then be a uniform gray and the lower one will be visible.

A signal can be decomposed into a weighted sum of all the spatial frequencies it contains by means of Fourier analysis. These frequencies can be computed by Fourier Transforms in one or two dimensions. The resulting collection of complex numbers is called the *spectrum* of the signal

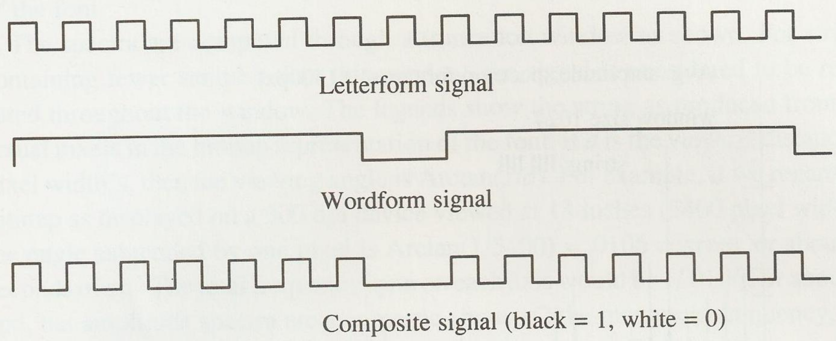
Since our interest is in sampled signals, namely the digital representation on a discrete grid of black-and-white characters we can study the *Discrete Fourier Transform* (DFT) whose k–th sample frequency $F(k)$ is given in one-dimension by $F(k) = \Sigma f(n)\exp(2\pi ikn/N)$, where $f(n)$ is the signal at the n–*th* sample point, N is the number of samples, and the sum is taken over all N samples. In the corresponding two-dimensional case we would consider a signal which is given a value at integer points (u,v). In both cases we assume that f is 0 outside some interval (rectangle, in the two-dimensional case). For black-and-white characters, f will always be 0 or 1, namely the pixel value at coordinate (u,v).

The DFT has several advantages. The main one is that it enables us to use bit-mapped fonts as the samples from which to compute the transform with no further sampling required. In addition, it is efficiently computable by well-known Fast Fourier Transforms. Also, there is a discrete inverse transform analogous to the inverse Fourier Transform from which one can perfectly reconstruct the original discrete signal given its DFT. Finally, under suitable conditions it gives a good approximation to the complete spectrum. We do not dwell here on the conditions which make this true, nor on techniques for improving this approximation when it fails, but we turn next to some specific spectra which we have computed with FFT's, so that some of these points can be illustrated.

1.2. *Average amplitude spectra for lines of text*

In general the spectrum is a complex valued function. The magnitude of the spectrum is the *amplitude spectrum* and its square is called the *power spectrum*, for adequate physical reasons in simple cases (cf. Sec. 4.3 [Oppenheim83]). It is thus a measure of how much signal is present at each frequency. Earlier we suggested that the spatial frequency due to the strokes of letters, which we call the *letterform frequency*, is an obvious component of a line of text, and we shall see next that it is a major one. A second component is the *wordform frequency* due to alternation between black and white of the words of text in a line such as might be perceived from a distance too great to distinguish letters (i.e., at which the spatial frequencies presented by the strokes are too high to be perceived). Without more detail, suffice it to say that we might expect to see fequencies in the spectrum corresponding at least to

Figure 2: *Idealized text signals.*

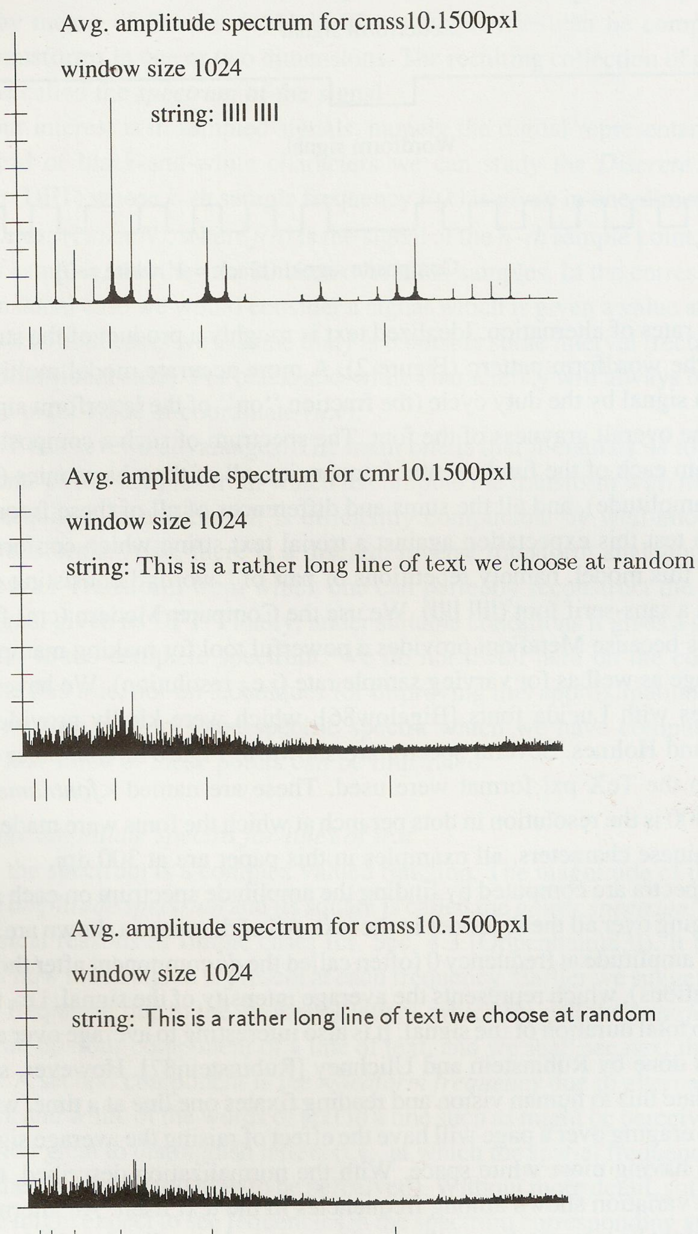


these two rates of alternation. Idealized text is roughly a product of the stroke pattern and the wordform pattern (Figure 2). A more accurate model multiplies the wordform signal by the duty cycle (the fraction ''on'' of the letterform signal), reflecting the overall grayness of the font. The spectrum of such a composite signal will contain each of the fundamental frequencies, all of their harmonics (with decreasing amplitude), and all the sums and differences of all of these frequencies.

We can test this expectation against a trivial text string which conforms quite closely to this model, namely repetitions of pair of ''words'' consisting of upper case I's in a sans-serif font (IIII IIII). We use the Computer Modern (cm) family in our studies because MetaFont provides a powerful tool for making manipulations of the image as well as for varying sample rate (i.e., resolution). We have also begun studies with Lucida fonts [Bigelow86], which were kindly provided us by Bigelow and Holmes. Several spectra are shown in Figure 3. Computer Modern bitmaps in the TeX pxl format were used. These are named *<fontname>.Npxl*, where N/500 is the resolution in dots per inch at which the fonts were made. Except for the Chinese characters, all examples in this paper are at 300 dpi.

These spectra are computed by finding the amplitude spectrum on each scanline and averaging over all the scanlines in the text. All of the spectra shown are normalized to the amplitude at frequency 0 (often called the *dc* component after the electrical applications), which represents the average intensity of the signal, i.e., the ratio of ''on'' to total duration of the signal. It is also interesting to average over an entire page, as is done by Rubinstein and Ulichney [Rubinstein87]. However, since we want to relate this to human vision and reading fixates one line at a time, we do not do this. Averaging over a page will have the effect of raising the average signal (the dc) due to having more white space. With the normalization described, this will reduce the variation shown among frequencies in the text itself. An alternative in this case might be to normalize to the maximum non-dc amplitude, which will gen-

Figure 3: *One-dimensional spectra*



erally be the letterform frequency. The dc gives a measure of the overall grayness of the font.

The spectra are computed through a truncation window as shown. For strings containing fewer samples than this window, the signal is considered to be replicated throughout the window. The legends show the string as produced from the actual pixels in the bitmap representation of the font. If d is the viewing distance in pixel width's, then the viewing angle is Arctan($1/d$). For example, if we regard the bitmap as displayed on a 300 dpi device viewed at 18 inches (5400 pixel widths), the angle subtended by one pixel is Arctan(1/5400) = .0106 degrees, or about 38 seconds of arc. The total frequency span on each axis would be 1/.0106 , or about 94 cpd, but amplitude spectra are symmetric about 1/2 the maximum frequency, and these graphs are plotted only to that frequency, i.e., about 47 cpd with the viewing hypotheses above. Below each graph are tick marks at one octave intervals, beginning at 1 cpd. In the trivial string, the principal component is close to 8 cpd, and this could have also been deduced by trigonometry from the bitmap. (The upper case I in cmss10 is 4 pixels wide with 11 pixels of setwidth, i.e., 7 white pixels between characters. The space between the ''words'' is 13 pixels.) Real spectra are, of course, more complex, as shown in the spectra for cmr10 and its san-serif cousin cmss10. The serif face generally has thinner strokes, which puts relatively more energy in the principal frequency; both faces have peaks a little above 8 cpd. The over-boldness of the sans-serif face corresponds to raising the value of the wordform signal (which, recall, is multiplied by the duty cycle of the letterform signal); this is what leads to more energy in the lower frequencies. This is visible in the spectra, with cmss having slightly more energy around 2 cpd.

It is not necessary to compute FFT's to ascertain the principal frequencies present in these average amplitude spectra. It suffices to count strokes in a line, calling each one cycle, and to divide by the visual subtense of the measured line. This gives the letterform frequency. Similarly, counting words produces the wordform frequency. This method produces results that are quite accurate for Computer Modern and the heavily tuned bitmapped Interleaf Classic font, as well as for Bigelow and Holmes' Lucida fonts, and there is no reason it should not be for others. In Table 1 we show some of the computed frequencies obtained by this kind of measure applied to a large text (one hundred lines of Bronte's *Wuthering Heights*). All letters were assumed to have two strokes, except i, j, l, and t (one each) and m, and w (three each). Visual subtense was computed from the font metrics assuming an 18-inch viewing distance.

A few things are worth noting. First, these frequencies are generally in agreement with those given by the DFT methods, although of course this counting scheme does not reveal amplitude. Second, apparent from these measurements is

Table 1: *Predominant Spatial Frequencies of Several Fonts*

Font	fl	fw	Font	fl	fw
cmr7	10.39	1.31			
cmr8	9.85	1.25	cmss8	10.29	1.31
cmr10	8.20	1.04	cmss10	8.33	1.12
cmr12	6.79	.89	cmss12	7.46	.95
cmr17	5.20	.66	cmss17	5.37	.69
cl06	10.57	1.27			
cl08	9.16	1.10			
cl10	7.83	.93			
cl12	6.71	.80			
cl14	5.80	.70			
cl 18	4.53	.55			

cmr = Computer Modern Roman
cmss = Computer Modern Sans-Serif
cl = Interleaf Classic
Type size in points is indicated in font name.
fl is letterform , fw is wordform frequency.

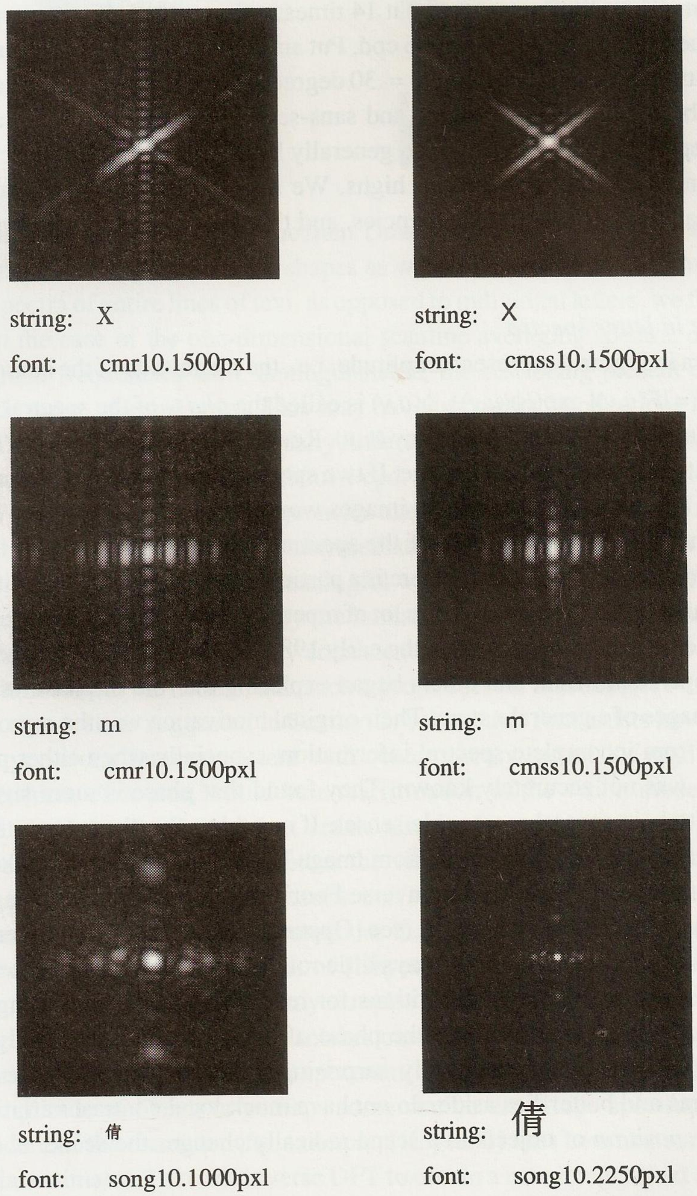
that the designers of all of these fonts have not scaled the fonts linearly down to small sizes. If they had, cmr7 would have a letterform frequency of 10/7, that of cmr10, i.e., 11.7 cpd. One study of human vision [Legge87], described below, suggests that readability drops off rapidly at high spatial frequencies, and this would seem to argue in favor of these expanded fonts.

1.3. *Two-dimensional spectra*

Two-dimensional amplitude spectra reveal features of individual letters which are not discernible by the scanline averaging described above, and also suggest ways to quantify certain differences among typefaces (serif vs. san-serif faces, for example). Figure 4 shows the spectra of several letters in various faces. The spectra are centered at the origin, with a distance of about 47 cpd on each axis. In each of these discrete amplitude spectra, the frequency samples represented are at about 94/128 = .74 cpd because these spectra were computed through a window of 128 samples. Note also that the characters are horizontally centered in a white space, which adds a horizontal low frequency artifact to the spectra.

Some features in the spatial domain can be deduced from these two-dimensional spectra. For example, in the spectrum of the cmr10 upper case X, one can see that there is more low frequency energy in the spectral feature at 45 degrees than in that at –45 degrees. Since the amplitude spectra represent *variation* in intensity, these directions will generally be at 90 degrees to the letter features. Thus the spectrum reflects that one stroke is thicker than the other, namely the one at angle –45 degrees in the character.

Figure 4: *Two–dimensional spectra of characters*



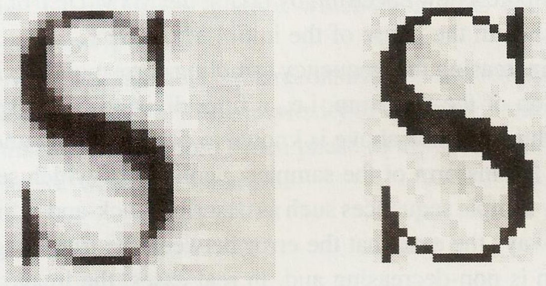
Of particular interest is the difference between serif and sans-serif characters. In the Computer Modern fonts the serifs are largely horizontal, which means they represent variation in the *vertical* direction. For example, the serif upper case X shows a repeating vertical pattern near the center line which is not manifest in the sans-serif spectrum. The pattern repeats about 14 times in the upper half, which means its base frequency is about $47/14 = 3.36$ cpd. Put another way, the serif presents a pattern repeating at a distance of $1/3.36 = .30$ degrees of visual angle. A similar difference can be seen between the serif and sans-serif lower case m. A character the same except for a thinner serif would generally have similar low frequency amplitude, but more, or higher intensity, highs. We suggest below that this additional energy is at visually relevant frequencies, and this may lend weight to arguments that serifs add to legibility.

1.4. *Phase in letter spectra*

The spectra in Figure 4 represent amplitude, i.e., the magnitude of the complex values $F(u,v) = |F(u,v)| \exp(i\phi(u,v))$. $\phi(u,v)$ is called the *phase* of the spectral value at the two-dimensional spatial frequency (u,v). Roughly speaking, phase is a measure of the position of patterns in an image. If two spectra differed only in phase, and by a constant amount, then the original images would be translations of one another. On the other hand, the amplitude of the spectral values is a measure of the local contrast variation of the image. If there is a particularly high spectral amplitude at a given frequency, then the image has a lot of repetition of a pattern at this frequency.

In a series of papers beginning in the early 1980's ([Hayes80], [Oppenheim81]), Hayes, Oppenheim, Lim, and others began exploring the role of phase vs. amplitude for images of a general nature. Their original motivation was the reproduction of signals from incomplete spectral information, especially when either phase or magnitude was not accurately known. They found that phase is more significant than amplitude in several reasonable senses. If one takes the phase from the spectrum of image A and the amplitude from image B, combining them to make a synthetic spectrum, and then takes the inverse Fourier Transform, the resulting image is generally recognizable as image A (see [Oppenheim83] Figure 4.35 for an example). This suggests that amplitude plays little role, and, in fact, the aforementioned series of papers give iterative algorithms for reproducing an image using essentially constant amplitude and only the phase of the spectrum under study. Upon reflection, this might not be entirely surprising, given that most images in the world, zebras and butterflies aside, do *not* have much local contrast variation, but varying the *position* of objects in a scene radically changes the scene.

Figure 5: *Reconstruction from phase only (left), one bit of phase (right)*



In work with my colleague Guozhen Duan (Duan88), we wondered whether these generalities applied to letter shapes as well. Indeed, when we examined amplitude spectra of entire lines of text, as opposed to individual letters, we found that, just as in the case of the one-dimensional scanline averaging spectra, only a few predominant frequencies were distinguishable, the rest being present over a not very wide range of amplitudes. This may have important implications for human vision. The visual system has radically different responses to different spatial frequencies, with an overall peak at about 4 cpd. The example of Figure 1 shows that very low, as well as very high, frequencies may need substantially more amplitude to be distinguishable when among intermediate frequencies. Although we have not done so, we suggest that a more meaningful way to normalize spectra of images is to weight the spatial frequencies by the human visual system's response. The human vision literature is sufficiently consistent about the gross structure of this response, if not about the details, that such a project seems reasonable. For type, this is likely to reveal, as suggested in Section 2. below, that font artists have been designing type along lines consistent with contemporary models of high visual response (which, of course, would be a reaffirmation of the vision models, not the designs! Undoubtedly skilled font artists already have the ''correct'' notion of what is easy to see, no matter whether the vision scientists do or not.)

Duan proposed, for technical reasons, that we consider the amplitude spectra of each character to have the same features as that of a square black box of about the same size. The iterative algorithms of [Hayes80] amount to this: Take a discrete spectrum with at least twice as many samples as there are points in the support of the image (that is, if the character vanishes outside an NxN rectangle, take an MxM DFT with Mx2N). In this enlarged spectrum, replace the amplitude with the ''virtual'' amplitude of Duan, or perhaps some similar amplitude spectrum which is independent of the character (for example, [Hayes80] uses constant amplitude for this initial estimate). Take the inverse DFT to obtain a new image. Next construct a restricted image which coincides with the new one on the NxN rectangle, but is 0 outside it. Finally, compute a new spectrum by taking the MxM DFT of this image and replacing its phase with the phase of the initial MxM DFT.

Roughly speaking, increasing the frequency sampling corresponds to increasing the frequency resolution of the spectrum, i.e., a finer division of the spectrum is computed. The algorithm sketched above is known to converge only under conditions on the so-called z-transform of the samples – conditions which seem not to hold for general binary sample sequences such as describe black-and-white characters. However, it is always the case that the error between the true image and the iterated approximation is non-decreasing and, in real cases, the image seems always to be quite recoverable. Garcia and Calero ([Garcia84]) suggest that initial amplitude estimates should depend somewhat on the image content, and this is the direction we are exploring. Note that the inverse transform may not return to an image with only two levels in it, nor even necessarily look like the original character. But after only a few iterations of this procedure, the original character is easily recognized, especially if some threshold is selected and all gray values below that are set to black and all above it set to white. We can take this as an indication that the amplitude of the spectrum is relatively unimportant in distinguishing characters from one another. This is illustrated in Figure 5, which represents a reconstructed upper case S from Computer Modern Roman digitized at 300dpi, after 30 iterations of the Hayes algorithm as implemented by Duan. In related investigations, we found that using all the amplitude information and only one bit of phase information also recovered the character quickly with similar iterative algorithms. [Curtis84] gives some abstract conditions for an image to be reconstructed from 1 bit of phase data, and shows that as the image size gets larger, the probability approaches 1 that a randomly chosen image meets these conditions.

1.5. *Chinese*

We have begun studies with the the Song font of Chinese characters produced with MetaFont by Hobby and Gu [Hobby83]. Chinese characters have substantially more horizontal strokes than do western fonts. This is visible in the amplitude spectra as 90 degree rotational symmetries not common in western spectra. The character represented in Figure 4 also shows the presence of a strong diagonal stroke at right angles to the corresponding spectral feature. This character is shown at two resolutions, with more high frequency visible in the high resolution spectrum. The low resolution picture can be regarded as an enlargement of the center of the high resolution one (but note that they are not entirely comparable because each character is in a 128x128 window with white side bearings extending to the window edge.

Since the higher resolution character fills this window more, it does not show as much of the low frequency energy which represents the artificial replication through the window.)

At this writing we have not begun reconstruction studies with Chinese characters of the sort described above. The complexity of Chinese characters would lead to the conjecture that stroke position, hence phase in the spectrum, is even more critical for them than for western characters.

2. Vision

2.1. *Visual models and type models*

In recent years, a substantial amount has been learned about spatial vision (Sekuler85, pp. 145–178). In particular, it has been known since 1968 that monochromatic vision can be modeled by independent spatial frequency tuned channels, in much the same way that color vision can be modeled by three channels tuned with peaks at the frequencies of red, green, and blue light. The fundamental evidence for this consists of measuring a subject's response to two different spatial frequencies, then adapting the subject to one of these frequencies, causing the visual system to reduce its response to that pattern. If the frequencies are sufficiently widely separated, the response will not be reduced at the non-adapted frequency.

There is much literature on the details of these channels. Recently, work of Wilson [Wilson84] suggests that there may be as few as six channels, each with a response which is given by the difference of two Gaussian functions, and having peaks approximately at .8, 1.7, 2.8, 4, 8, and 16 cpd. Generally, these channels overlap in pairs, but not otherwise. In essence, the channels behave as filters, each responding to signals within its passband. The information of all of them is summed in a way that makes the entire pattern present an image to the visual system. This kind of model is confirmed by psychophysical experiments with humans, as well as by probes inserted into *single cells* of the visual cortex of cats and monkeys. Cells can be isolated that respond to patterns of some frequency and orientation but not to others. There is intuitive appeal to multi-channel models, in that they can easily distinguish signals from (uniformly distributed) noise. (The noise is that which has equal output in all channels. All else is signal. In fact, this observation has been used to reduce noise in medical images [Baker80].)

We will speculate in this section how some of the spectral properties of type, which we have described above, fit in with these models of human vision. Roughly, we want to argue that type designers have implicitly designed with parameters which cause the type to evoke high response from the human visual system. It is difficult to separate our cognitive processes, in particular the recognition of words, from the reading process. Type designers attempt to do this with their initial design color studies, in which conventional collections of letters are studied. In the final analysis, though, type is about words, and the design of a typeface is not complete without the study of the appearance of real text.

Very few psychophysical studies have been made about reading with attention to spatial frequency models of vision. The most careful of those few is a series of works by Legge et al., culminating in [Legge87]. That work relates reading rate to text contrast and character width. It makes an experimental conclusion which is consistent with a well-known typographic phenomenon, for which we describe a somewhat theoretical foundation below. Legge's conclusion is that at the high contrast typical of reading, the reading rate is highest for characters subtending about .25 degrees. Given that most characters have approximately two strokes, this means that the spatial frequency presented by the strokes is about 8 cpd. As our table of measurements above shows, this is about the frequency of most 10-point type, and it is well known that type gets *harder*, not easier to read as its size increases well above 12-point. Most reading authorities speculate that this decreased legibility comes from the requirement to make longer saccadic eye motions, the motions between fixations which we make when reading, with the attendant reduced information content per line. We will suggest below that the multi-channel models imply that bigger type actually presents information at frequencies less discernible to the early part of the visual system.

The Legge study mentioned above also relates reading rate to psychophysical contrast, but this is not the same thing as the contrast which type designers manipulate when they vary stem weights. (Note that Legge's experiments were done on CRT's, not paper, possibly a consequential difference.) In type design, what is changed is the duty cycle (the fraction ''on'') of the letterform signal idealized in Figure 2. Indeed, even that example is unreal in that the white space between strokes is typically several times the stroke width. The duty cycle of the wordform signal, however, is an artifact of the mean word width, which is, of course, a function not only of the type but of the language as well. For example, German text is often printed in 9-point type, which then tends to have wordform frequencies similar to larger English type, even though higher letterform frequencies (however, the response curves for human visual response are broader at their peaks than on their shoulders, which means that we might expect more tolerance in the letterform than the wordform frequencies). In addition, a more accurate form of this idealized model multiplies the wordform signal by the duty cycle of the letterform signal to account for the overall average grayness of text without regard to the details of letters. That notion of contrast – sometimes called the color of the text – then becomes built into the compound signal model and can have filters, such as proposed in [Wilson84] applied to the signal.

It is beyond the scope of this paper to detail the application of multi-channel filtering models to type models, but the central feature is this: The overall peak of visual response is not near 8 cpd, but rather near 4 cpd. If the visual system were responding only to the strokes in letters, we would expect best response not from 10-point type but rather from something about twice as large. What is known is that recognition of individual words or even letters *does* increase as size increases above 12-point, although reading rate for text decreases. The overall response to a compound signal, such as presented by text (namely, the major frequencies presented by strokes and by words), can indeed be shown to decrease as the stroke frequency decreases below 8 cpd (i.e., the type size increases above 10-point). Thus, we can speculate that 10-point type is actually easier for the early visual system to process than much bigger type, and we need not resort to eye motion explanations.

2.2. *Two–dimensional vision*

The one-dimensional model of Wilson can be extended to two-dimensions (Wilson83), but we have not attempted much to combine it with type models, as our two-dimensional work has largely consisted of the reconstruction studies described in Section 1.4. There is some evidence in the vision literature that monochromatic vision is more sensitive to phase than to spectral amplitude changes and this would be expected if natural images have most of their information in their phase. Since this seems to be the case for letters also, once again it would suggest that type designers have implicitly chosen ''optimal'' variables to manipulate. Roughly speaking, radical changes in spectral amplitude throughout a font could be accomplished by radical changes in stroke width. It is probably difficult for the unskilled eye to detect these changes as much as mis-positioning of strokes, which are phase changes. For example, a lower case m with its third stroke too close to the second might be more easily mistaken for an n than would one whose strokes were uniformly a little too thick. Indeed, the amplitude spectra of an m and an n are nearly identical because repeated m's and repeated n's present little difference from one another in their rate of alternation of black and white.

On the other hand, serifs, especially horizontal ones (as would be the case in most western letters), do present additional information both in phase and amplitude, and our preliminary indication is that they are at visually relevant frequencies, i.e., their presence enhances the signal in the Wilson model.

Acknowledgments. The FFT's were coded by Richard Ratta, YuBing Chen and Guozhen Duan and converted to generate PostScript by the author. Most of the two-dimensional data described were taken by Duan, who also implemented the reconstruction algorithms. The manuscript was prepared with the Interleaf Technical Publishing system in Times-Roman type producing PostScript output. The spectra and reconstruction images were made from encapsulated PostScript embedded in the document. Thanks to Bigelow and Holmes and to Interleaf, Inc. for the use of their font data. PostScript is a trademark of Adobe Systems. Lucida is a trademark of Bigelow and Holmes. The research was supported by NSF Grant DCR–85–03154 at UMASS/Boston.

References

[1] K. D. Baker, and F. D. Sullivan (1980), ''Multiple Band Pass Filters in Image Processing,'' *IEEE Proc.* **127(E)**, No. 5, Sept. 1980, pp. 173–184.

[2] C. Bigelow and D. Day (1983), ''Digital Typography,'' *Scientific American*, **249**, No. 2, pp. 106–119.

[3] C. Bigelow and K. Holmes (1986), ''The design of Lucida: an integrated family of types for electronic literacy,'' *Text Processing and Document Manipulations*, J.C. vanVliet, editor, Cambridge University Press, Cambridge. (This is the proceedings of the Nottingham Conference, EP86.)

[4] S. R. Curtis, J. S. Lim, and A. V. Oppenheim (1984), ''Signal Reconstruction from One Bit of Fourier Transform Phase,'' *IEEE Conference on Acoustics, Speech, and Signal Processing*, pp. 12A.5.1–12A–5.4.

[5] G. Duan and R. A. Morris (1988), ''Phase information in characters,'' work in progress, Dept. of Math. and C.S., Univ. of Massachusetts at Boston, MA.

[6] K. R. Castleman (1979), *Digital Image Processing*, Prentice–Hall, Englewood Cliffs, N.J.

[7] N. Garcia and A. Calero (1984), ''Faster Phase Only Image Reconstruction,'' *IEEE Conference on Acoustics, Speech, and Signal Processing*, pp. 12A.3.1–12A–3.4.

[8] M. H. Hayes, J. S. Lim, and A. V. Oppenheim (1980), ''Signal Reconstruction from Phase or Magnitude,'' *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **ASSP–28**, No. 6., pp. 672–680.

[9] J. Hobby and G. Gu (1983), ''A Chinese Meta-Font,'' Stanford Technical Report, Stan–CS–83–974, Stanford Department of Computer Science, Stanford, CA.

[10] A. Naiman (1988) *The Use of Grayscale for Improved Character Presentation*, Ph.D. dissertation, in progress, Department of Computer Science, University of Toronto.

[11] A. V. Oppenheim and J. S. Lim (1981), ''The importance of phase in signals,'' *Proc. IEEE*, **69**, pp. 528–541.

[12] A. V. Oppenheim and A. S. Willsky with I. T. Young (1983), *Signals and Systems*, Prentice Hall, Englewood Cliffs, N.J.

[13] G. C. Phillips and H. R. Wilson (1984), ''Orientation bandwidths of spatial mechanisms measured by masking,'' *J. Opt. Soc. Am. Series A*, **1**, pp. 226–232.

[14] R. Rubinstein (1988), *Digital Typography*, Addison–Wesley, Reading, MA, to be published.

[15] R. Rubinstein and R. Ulichney (1987), personal communication. See also, [Rubinstein88].

[16] R. Sekuler and R. Blake (1985), *Perception*, Alfred A. Knopf, New York, N.Y.

[17] H. R. Wilson and D. J. Gelb (1984), ''Modified line-element theory for spatial frequency discrimination,'' *J. Opt. Soc. Am. Series A*, **1(1)**, pp. 124–131.

[18] J. I. Yellott, Jr., B. A. Wandell, and T. N. Cornsweet (1984), ''The beginnings of visual perception: the retinal image and its initial encoding,'' Chapter 7 in *Handbook of Physiology – The Nervous System III*. The American Physiological Society, Bethesda, MD.