Gabor Representations of Spatiotemporal Visual Images

Bruce MacLennan*

Technical Report CS-91-144

September 13, 1991

Abstract

We review Gabor's Uncertainty Principle and the limits it places on the representation of any signal. Representations in terms of Gabor elementary functions (Gaussian-modulated sinusoids), which are optimal in terms of this uncertainty principle, are compared with Fourier and wavelet representations. We also review Daugman's evidence for representations based on two-dimensional Gabor functions in mammalian visual cortex. We suggest three-dimensional Gabor elementary functions as a model for motion selectivity in complex and hypercomplex cells in visual cortex. This model also suggests a computational role for low frequency oscillations (such as the alpha rhythm) in visual cortex.

1 Introduction

Dennis Gabor is best known as the father of holography, in recognition of his development of its theory in 1947. Here, however, we are concerned with his theory of communication, published in 1946 (Gabor 1946), two years before Claude Shannon's more famous theory (Shannon 1948). Gabor's theory was not simply an anticipation of Shannon's (as was Hartley's, for example); rather it addresses a completely different aspect of the nature of communication. It also provides a basis for the representation and processing of information in vision and perhaps other sensory modalities. This aspect will be our concern here.

First we review Gabor's Uncertainty Principle, which defines limits on the representation of any signal. Next we discuss the representation of signals in terms of Gabor elementary functions (Gaussian-modulated sinusoids), which is optimal in terms of the uncertainty principle and has several advantages over representations based

*Computer Science Department, University of Tennessee, Knoxville, TN 37996; maclennan@cs.utk.edu.

on Fourier series and the Sampling Theorem. After reviewing John G. Daugman's research supporting the presence of Gabor's representations in mammalian vision, we discuss its pros and cons compared to wavelet-based representations. Finally we present an extension of the Gabor representation and apply it to the representation and processing of spatiotemporal patterns in the visual cortex.

2 The Gabor Uncertainty Principle

Gabor proved his uncertainty principle by applying to arbitrary signals the same mathematical apparatus as used in the Heisenberg-Weyl derivation of the uncertainty principle in quantum mechanics. We give our own version of this proof in Appendix A; here our intent is to build intuition, so we present several informal derivations.

Suppose we are trying to measure the frequency of a tone. Intuitively, the longer the sample we take, the more accurate will be our measurement (Fig. 1), which suggests that the error in measuring the frequency, Δf, is inversely related to the duration of the measurement, Δt. This intuition can be made a little more precise by considering a very basic kind of frequency measurement. Suppose we have a device that counts every time our signal reaches a maximum; then the number of maxima in an interval of time Δt will be the average frequency during that interval (Fig. 2). How long must Δt be in order to guarantee we can distinguish frequencies differing by Δf? This will occur when the counts for f and $f + \Delta f$ are guaranteed to differ by at least one (Fig. 3). That is,

$$(f + \Delta f)\Delta t - f\Delta t \geq 1,$$

or,

$$\Delta f \Delta t \geq 1. \tag{1}$$

This is the basic *Gabor Uncertainty Principle*; it means that the product of the uncertainties in frequency and time must exceed a fixed constant, and so the accuracy with which one of them can be measured limits the best possible accuracy with which the other can be measured.[1]

Heisenberg's Uncertainty Principle is a simple corollary of Eq. 1, since according to quantum mechanics the energy of a photon is proportional to its frequency, $E = hf$.

[1] Note that we have shown a duration $\Delta t \geq 1/\Delta f$ is necessary to *discriminate* frequencies differing by Δf. On the other hand, if we *measure* a frequency f during an interval Δt, then the actual frequency could be as low as $f - 1/\Delta t$ or as high as $f + 1/\Delta t$. Therefore, the uncertainty around f is $\Delta f \geq 2/\Delta t$, giving the uncertainty principle $\Delta f \Delta t \geq 2$. Furthermore, there are other methods of measuring the frequency, such as counting sign changes (zero crossings), which would give $\Delta f \Delta t \geq 1/2$ for the discrimination case and $\Delta f \Delta t \geq 1$ for the measurement case. Thus although the exact constant depends on what we are measuring and how we are measuring it, its value doesn't much matter, since the conclusion is the same: there is a lower limit on the product of the uncertainties in the time and frequency domains. For the sake of simplicity we use the constant 1.

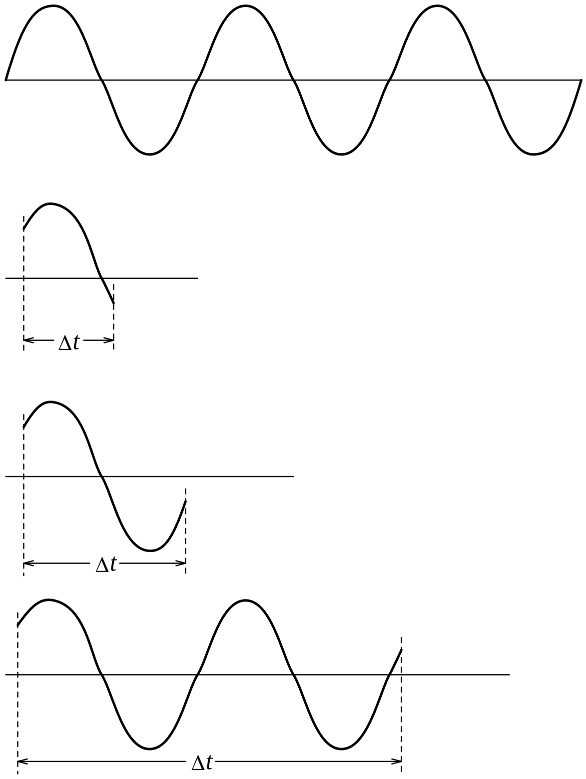

Figure 1: Improved frequency measurement over longer time intervals. The uncertainty in the frequency Δf decreases as the measurement interval Δt increases, and vice versa.

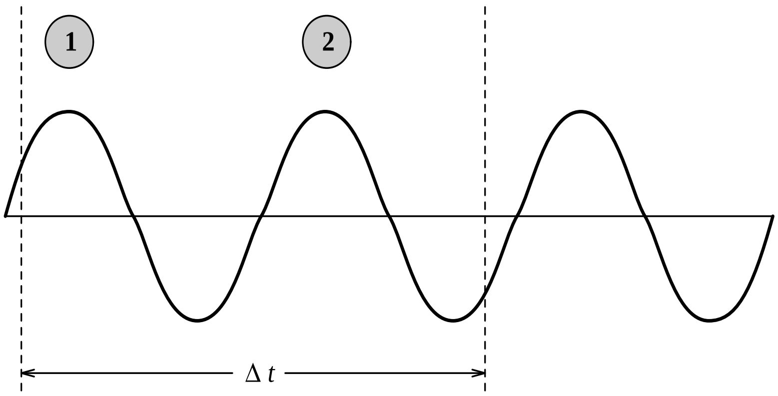

Figure 2: Measuring frequency by counting maxima in a given time interval. The circled numbers indicate the maxima counted during the measurement interval Δt. Since signals of other frequencies could also have the same number of maxima in that interval, there is an uncertainty Δf in the frequency.

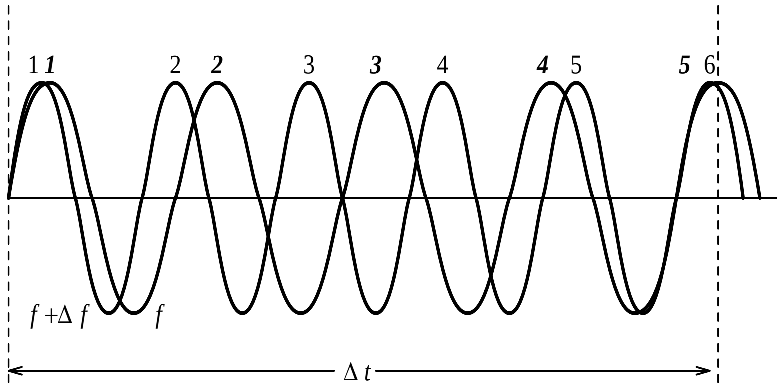

Figure 3: Minimum time interval Δt to detect frequency difference Δf. If two signals differ in frequency by Δf, then a measurement of duration $\Delta t \geq 1/\Delta f$ is required to guarantee a difference in counts of maxima. Italic numbers indicate maxima of signal of frequency f; roman numbers indicate maxima of signal of higher frequency $f + \Delta f$.

Multiplying both sides of Eq. 1 by h (Planck's constant) yields

$$\Delta E \Delta t \geq h,$$

which is one form of Heisenberg's principle.[2] Of course, Heisenberg derived his principle first; Gabor's accomplishment was to show that the same mathematical derivation applied to communication systems.

A more formal derivation of Gabor's Uncertainty Principle is based on the observation that the "spread" of a signal and its Fourier transform are inversely proportional (Fig. 4).[3] To accomplish this we must first specify a way of measuring the spread of functions, especially when they are not *strictly local* (i.e., have noncompact support). For suppose we measure a frequency f over an interval of time Δt; this does not imply that the frequency during that interval was always in the range $f \pm \Delta f$; it means only that the *average* frequency over that interval was in $f \pm \Delta f$. The instantaneous frequency could have varied widely, and so its spectrum might look like that in Fig. 5 (assumed to be centered on f). Nevertheless, we can assign a *nominal bandwidth* to the spectrum that measures its spread around the measured frequency f. Alternately we can imagine that Fig. 5 represents the transfer function of a band-pass filter; the nominal bandwidth is a measure of the width of the band compared with that of an ideal band-pass filter. We say that nominal bandwidth measures the spectrum's *localization in the frequency domain*. Similarly, to a signal that may not actually be

[2]Different methods of measuring ΔE and Δt yield different constants on the right-hand side, such as $\hbar = h/2\pi$ or $\hbar/2$. Again, the exact constant doesn't matter. Also, since $E = pv/2 = px/2t$, we have the other common form of the Heisenberg principle, $\Delta p \Delta x \geq 2h$.

[3]The derivation follows Yu (1976, pp. 44–45).

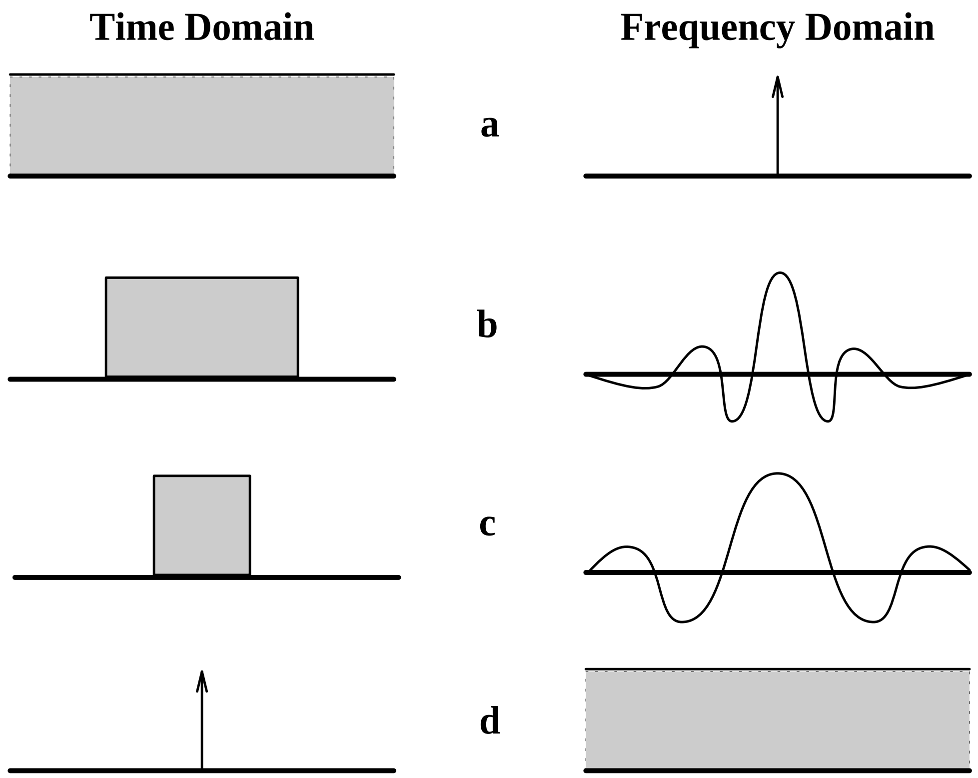

Figure 4: The "spread" of a signal and its Fourier transform are inversely proportional. **(a)** A constant function in the time domain corresponds to a unit impulse (Dirac delta function) in the frequency domain. **(b, c)** As the width of a pulse in the time domain decreases, its spectrum in the frequency domain spreads (spectrum shown is schematic). **(d)** A unit impulse in the time domain has a spectrum which is a constant function.

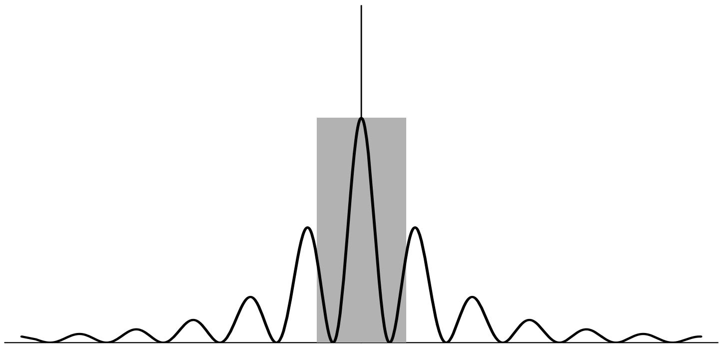

Figure 5: Nominal bandwidth in frequency domain of nonnegative spectrum. The nominal bandwidth is the width of a rectangular pulse (shaded) that has the same area as the continuous spectrum and has a height equal to its amplitude at the origin.

localized in a particular interval of time, we assign a *nominal duration* that measures its spread in time, and thus its localization in the time domain.[4]

Although there are many ways to define these measures, we define the *nominal duration* of a nonnegative signal ϕ to be the duration of a rectangular pulse of the same area and amplitude at the origin as the signal (Fig. 6).[5] Thus the nominal duration Δt is defined by the equation

$$\Delta t\,\phi(0) = \int_{-\infty}^{\infty} \phi(t)\mathrm{d}t \qquad (\phi(t) \geq 0). \tag{2}$$

Similarly, the *nominal bandwidth* of the Fourier transform of ϕ, $\Phi = \mathcal{F}(\phi)$, is defined

$$\Delta f\,\Phi(0) = \int_{-\infty}^{\infty} \Phi(f)\mathrm{d}f \qquad (\Phi(f) \geq 0). \tag{3}$$

Next write $\Phi(0)$ as the Fourier transform of ϕ evaluated at $f = 0$:

$$\Phi(0) = \left.\int_{-\infty}^{\infty} \phi(t)e^{-2\pi i f t}\mathrm{d}t\right|_{f=0} = \int_{-\infty}^{\infty} \phi(t)\mathrm{d}t = \Delta t\,\phi(0).$$

Therefore,

$$\Delta t = \frac{\Phi(0)}{\phi(0)}. \tag{4}$$

[4]We call a function *local* if most of its area is concentrated in a compact region; we call it *strictly local* if it has compact support (roughly, it is zero outside of a compact region). For example, the normal distribution is local, but a finite pulse is strictly local. Note that we can have a local function that is in fact more localized than a given strictly local function. We call a function *nonlocal* if its area is spread more or less uniformly over its (noncompact) domain; sine and cosine are good examples.

[5]General (possibly negative) signals are considered later. Obviously there are many ways to measure the spread of a function, for example, Gabor (1946) uses the variance, as does Hamming (1989, pp. 181–184); in Appendix A we use the standard deviation. The choice of measure affects only the constant on the right-hand side of the uncertainty relation.

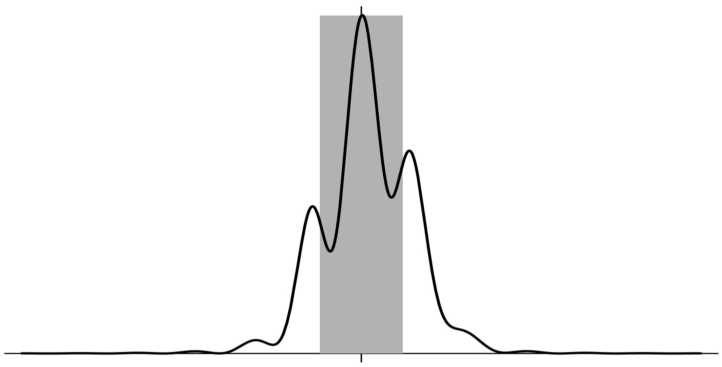

Figure 6: Nominal duration in time domain of nonnegative signal. The nominal duration is the duration of a rectangular pulse (shaded) that has the same area as the signal and has a height equal to its amplitude at the origin.

Similarly, applying the inverse Fourier transform,

$$\phi(0) = \int_{-\infty}^{\infty} \Phi(f) e^{2\pi i f t} \mathrm{d}f \Big|_{t=0} = \int_{-\infty}^{\infty} \Phi(f) \mathrm{d}f = \Delta f \, \Phi(0).$$

Therefore,

$$\Delta f = \frac{\phi(0)}{\Phi(0)}. \tag{5}$$

Multiplying Eq. 4 by Eq. 5 yields

$$\Delta f \Delta t = \frac{\phi(0)}{\Phi(0)} \frac{\Phi(0)}{\phi(0)} = 1. \tag{6}$$

Thus we see that the nominal duration and nominal bandwidth are reciprocals of each other, provided the signal and its Fourier transform are both nonnegative. In other words, there is a minimum possible simultaneous localization of the signal in the time and frequency domains.

Now we consider the general case, in which the signal and its Fourier transform may take on negative values. This is accomplished by defining the nominal spreads in terms of the absolute values of the functions (Fig. 7):

$$\Delta t |\phi(0)| = \int_{-\infty}^{\infty} |\phi(t)| \mathrm{d}t, \tag{7}$$

$$\Delta f |\Phi(0)| = \int_{-\infty}^{\infty} |\Phi(f)| \mathrm{d}f. \tag{8}$$

The absolute value weakens our previous equality to an inequality:

$$\begin{aligned} \Delta t |\phi(0)| &= \int |\phi(t)| \mathrm{d}t \;\geq\; \left| \int \phi(t) \mathrm{d}t \right| \;=\; |\Phi(0)|, \\ \Delta f |\Phi(0)| &= \int |\Phi(f)| \mathrm{d}f \;\geq\; \left| \int \Phi(f) \mathrm{d}f \right| \;=\; |\phi(0)|. \end{aligned}$$

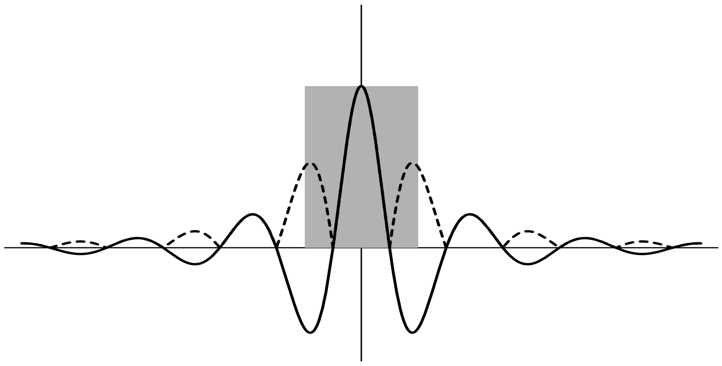

Figure 7: Nominal duration in time domain of arbitrary signal. Signal shown as solid line, absolute value of signal shown as dashed line. The nominal bandwidth of a spectrum is the width of a rectangular pulse (shaded) that has a height equal to the spectrum's amplitude at the origin, and that has the same area as the absolute value of the spectrum.

These equations give bounds on the nominal spreads in terms of the signal and its transform at the origin:

$$\Delta f \geq \frac{|\phi(0)|}{|\Phi(0)|}, \quad \Delta t \geq \frac{|\Phi(0)|}{|\phi(0)|}.$$

From these equations we get the general Gabor Uncertainty Principle:

$$\Delta f \Delta t \geq 1. \tag{9}$$

It should be noted that such an uncertainty principle applies whenever we make simultaneous measurements of a function and its Fourier transform.[6]

The implications of Gabor's principle are easier to understand by looking at it in "Fourier space," where the abscissa reflects the time domain and the ordinate the frequency domain (Fig. 8). Then Gabor's principle says that the spreads or uncertainties in the time and frequency measurements must define a rectangle in Fourier space whose area is at least 1. Thus we can decrease Δt, and so localize the signal better in the time domain, or decrease Δf, and so localize it better in the frequency domain, but we cannot localize it arbitrarily well in both domains simultaneously. The most we can localize signals in the Fourier domain is into rectangles of size $\Delta f \Delta t = 1$.

[6] In the language of quantum mechanics, f and t are *conjugate variables*.

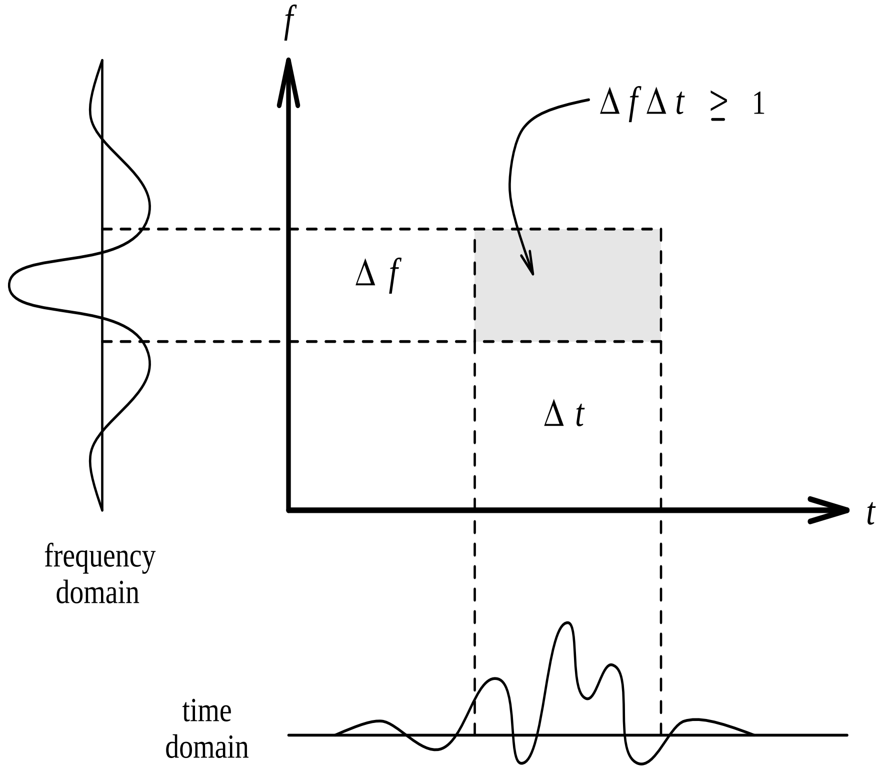

Figure 8: Minimum possible localization of signal in Fourier space. The product of the nominal duration Δt and nominal bandwidth Δf of a signal must be at least 1.

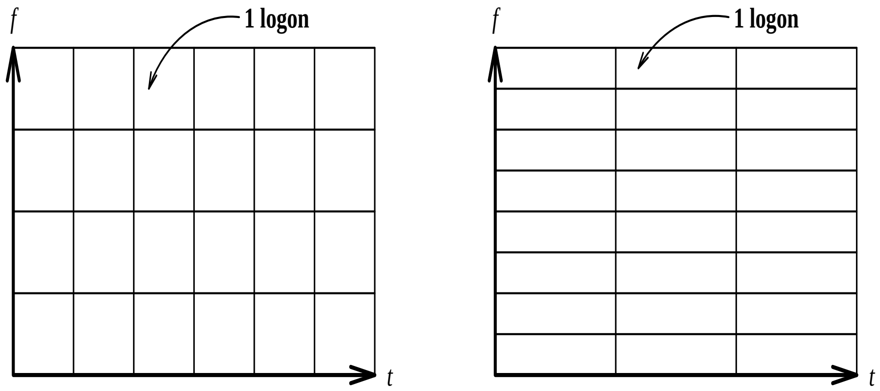

Figure 9: A band-limited signal is defined by a fixed number of elementary information cells. The most efficient possible (noiseless) channel divides Fourier space into information cells of size $\Delta f \Delta t = 1$. Each cell contains one *logon* of information.

3 Gabor Representation of One-Dimensional Signals

Suppose we are transmitting information by sending signals of various frequencies of bandwidth F during an interval of time T.[7] Suppose we sample the signal over intervals of length Δt to determine the signal strength at various frequencies (say, through a bank of band-pass filters). Then the closest frequencies we will be able to distinguish will be given by $\Delta f = 1/\Delta t$; that is, any frequencies differing by less than this Δf will be operationally indistinguishable. Thus our measuring apparatus divides Fourier space into *information cells* of size $\Delta f \Delta t \geq 1$ (Fig. 9). Since the most efficient possible (noiseless) channel will have $\Delta f \Delta t = 1$, the number of such *elementary information cells* determines the maximum amount of information that can be transmitted. No matter how we divide up Fourier space, its area gives the number of elementary information cells, and thus the number of independent quantities that can be transmitted. For example, in the simple case where $T = M\Delta t$, $F = N\Delta f$ and $\Delta f \Delta t = 1$, we are able to transmit MN independent quantities. For this reason Gabor defined a $\Delta f \Delta t = 1$ rectangle in Fourier space to be the basic *quantum* of information and called it a *logon*. Thus any device (of the given bandwidth) can transmit at most MN logons of information (in the given time interval).[8]

[7] As Gabor notes, all real, physical signals have finite bandwidth and duration.

[8] The reader will wonder how Gabor's measure of information relates to Shannon's; in fact they are orthogonal. Gabor's measure, which may be called *structural* information, quantifies the number of possible degrees of freedom. Shannon's measure, which may be called *metrical* or *selective* information, quantifies the decrease in a priori uncertainty in a single one of these degrees of freedom. For example, in an optical device, the resolving power is equivalent to the logon content or structural information, whereas the logarithm of the number of discriminable brightness levels is equivalent to

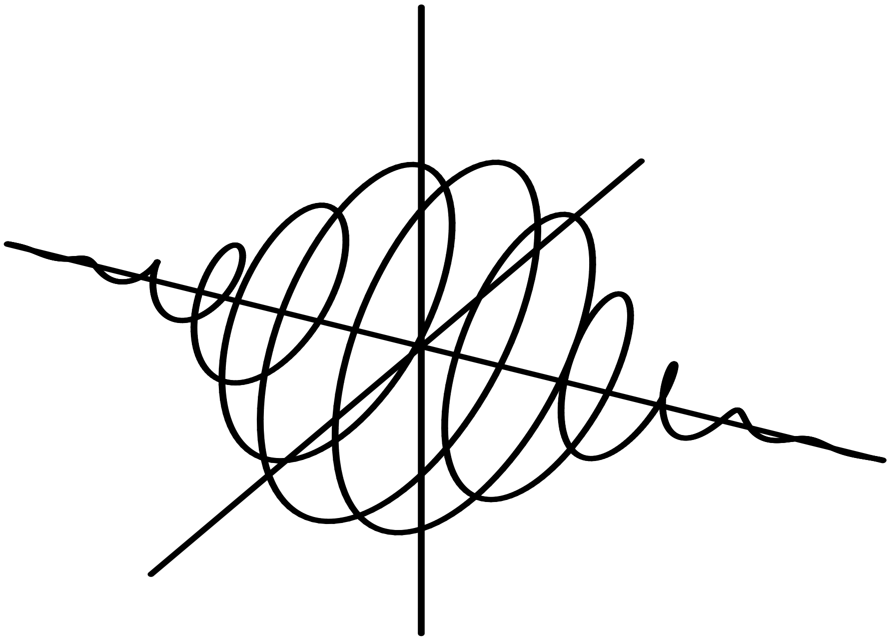

Figure 10: The Gaussian-modulated complex exponential, or Gabor elementary function ϕ_{jk}. The t axis goes from left to right through the center of the spiral. The imaginary axis is vertical; the real axis is horizontal, perpendicular to the other two axes. In this case $j = 0$ (no displacement from origin), $k = 1$, $\Delta f = 1$ and $\alpha^2 = 20$. The function is plotted from $t = -6$ to $t = 6$.

Gabor also showed that the minimum area in Fourier space is achieved by Gaussian-modulated complex exponential functions of the form (Fig. 10):

$$\phi_{jk}(t) = \exp[-\pi(t - j\Delta t)^2/\alpha^2] \exp[2\pi ik\,\Delta f\,(t - j\,\Delta t)], \tag{10}$$

where $\Delta f \Delta t = 1$.[9] Notice that the first factor leads to a Gaussian envelope centered on $j\Delta t$, and the second factor is the conjugate exponential form of the trigonometric functions of frequency $k\Delta f$. The parameter α determines the locality (spread) of the Gaussian envelope; it is proportional to its standard deviation. So we have a periodic function modulated by a Gaussian envelope, a *coherent state* or *wave packet* in the terminology of quantum mechanics. This can be seen more clearly by using Euler's formula to rewrite Eq. 10 in terms of the cis (cosine + i sine) function and then in

selective or Shannon information. Both notions of information are necessary for a complete theory of communication (MacKay 1969, pp. 178–180, 186–189; Cherry 1978, pp. 47–49). Remarkably, by 1928 Hartley had anticipated Gabor and Shannon by suggesting that the information transmittable over a channel is proportional to $MN \log S$, where S is the number of discriminable power levels (Cherry 1978, p. 47).

[9] Although Gabor showed that these functions occupy minimum area in terms of his (variance-based) definition of nominal spread, they also do so in terms of the definitions in Eqs. 7 and 8.

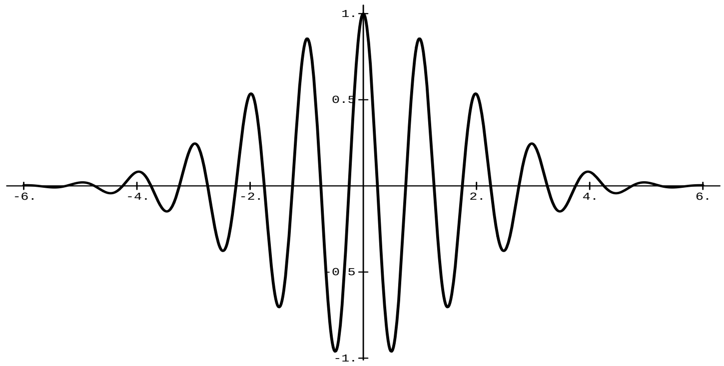

Figure 11: The Gaussian cosine function C_{jk}, or even-symmetric component of ϕ_{jk}. In this case $j = 0$ (no displacement from origin), $k = 1$, $\Delta f = 1$ and $\alpha^2 = 20$.

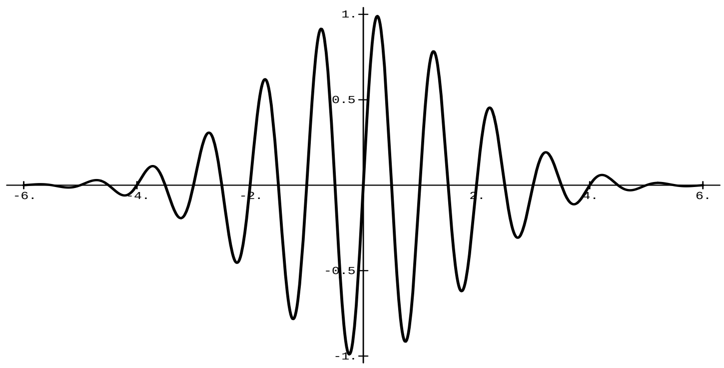

Figure 12: The Gaussian sine function S_{jk}, or odd-symmetric component of ϕ_{jk}. In this case $j = 0$ (no displacement from origin), $k = 1$, $\Delta f = 1$ and $\alpha^2 = 20$.

terms of the sine and cosine functions:

$$\begin{aligned} \phi_{jk}(t) &= \exp[-\pi(t - j\Delta t)^2/\alpha^2]\mathrm{cis}[2\pi k\Delta f\,(t - j\,\Delta t)], \\ &= \exp[-\pi(t - j\Delta t)^2/\alpha^2]\cos[2\pi k\Delta f\,(t - j\,\Delta t)] + \\ &\quad\ i\exp[-\pi(t - j\Delta t)^2/\alpha^2]\sin[2\pi k\Delta f\,(t - j\,\Delta t)], \end{aligned}$$

Thus the Gabor elementary function is the sum of the *Gaussian cosine* and *Gaussian sine* functions (Figs. 11 and 12). If we let C_{jk} and S_{jk} represent the Gaussian cosines and sines:

$$C_{jk}(t) = \exp[-\pi(t - j\Delta t)^2/\alpha^2]\cos[2\pi k\Delta f\,(t - j\,\Delta t)], \tag{11}$$

$$S_{jk}(t) = \exp[-\pi(t - j\Delta t)^2/\alpha^2]\sin[2\pi k\Delta f\,(t - j\,\Delta t)], \tag{12}$$

then $\phi_{jk} = C_{jk} + iS_{jk}$.

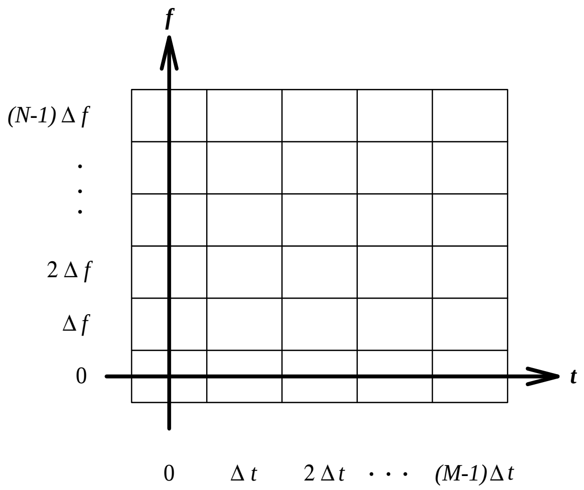

Figure 13: Representation of band-limited, finite-length signal by elementary information cells centered on frequencies $f = 0$, Δf, $2\,\Delta f$,..., $(N-1)\Delta f$, and on times $t = 0$, Δt, $2\,\Delta t$,..., $(M-1)\Delta t$. Cells are indexed by frequency-band quantum numbers $k = 0, 1, \ldots, N-1$ and time-interval quantum numbers $j = 0, 1, \ldots, M-1$.

So far we have had little to say about the coefficients associated with the elementary information cells; this is the topic we now address. Suppose a rectangular region of Fourier space is divided into MN elementary information cells, and that ψ is a signal whose duration and bandwidth are confined to that region. For simplicity we assume that the cells are centered on frequencies $f = 0$, Δf, $2\,\Delta f$,..., $(N-1)\Delta f$, and on times $t = 0$, Δt, $2\,\Delta t$,..., $(M-1)\Delta t$ (Fig. 13). Gabor showed that any such (finite energy) ψ can be represented as a linear superposition of Gaussian sinusoids:

$$\psi = \sum_{j=0}^{M-1} \sum_{k=0}^{N-1} a_{jk} C_{jk} + b_{jk} S_{jk}. \tag{13}$$

Each Gaussian cosine C_{jk} or sine S_{jk} is localized in the cell centered on time $j\,\Delta t$ and frequency $k\,\Delta f$; we call j the cell's *time-interval quantum number* and k its *frequency-band quantum number*. The real coefficients a_{jk} and b_{jk} show the amplitudes of Gaussian cosines and sines in each cell.

It would appear that ψ is determined by $2MN$ real coefficients, but since $S_{j0} = 0$, only $2MN - M$ of the coefficients are independent, as we can see by writing Eq. 13 in the form:

$$\psi = \sum_{j=0}^{M-1} \left\{ a_{j0} C_{j0} + \sum_{k=1}^{N-1} a_{jk} C_{jk} + b_{jk} S_{jk} \right\}.$$

The parameters a_{j0} determine the DC value of ψ in each of the M time intervals.

Just as is done in Fourier series, we can express ψ as a complex series,

$$\psi = \sum_{j=0}^{M-1} \sum_{k=-N+1}^{N-1} c_{jk}\phi_{jk}, \tag{14}$$

where the complex coefficients are given by:

$$\begin{aligned} c_{j0} &= a_{j0}, \\ c_{jk} &= (a_{jk} - ib_{jk})/2, \quad k > 0, \\ c_{jk} &= (a_{jk} + ib_{jk})/2, \quad k < 0. \end{aligned}$$

Notice that $c_{j,-k} = c^*_{j,k}$, the complex conjugate of $c_{j,k}$. We omit the derivation of the complex series as it can be found in any standard textbook on Fourier series.

Although there are $M(2N-1)$ complex coefficients c_{jk}, we have seen that $M(N-1)$ of them are complex conjugates of the others, and thus are not independent. Out of the remaining MN complex coefficients (one corresponding to each elementary information cell), the M coefficients c_{j0} are real, so once again we find that the signal is determined by $2MN - M$ real values. Thus Gabor has shown that a signal of duration T and bandwidth F has $T(2F - \Delta f)$ (real) degrees of freedom, and is thus capable of conveying $M(2N - 1)$ independent real values.[10]

Gabor's measure of information is consistent with the number of degrees of freedom given by the Sampling Theorem (Shannon 1948). To see this, observe that the highest frequency elementary information cells are *centered* at frequency $(N - 1)\Delta f$; therefore their maximum frequency (as defined by their nominal spread) is $f_{\mathrm{m}} = (N - 1/2)\Delta f$. The Sampling Theorem says that to reconstruct ψ we must take equally spaced samples at a minimum of the *Nyquist frequency*, which is twice the maximum frequency. Therefore in time T the number of samples we must take is:

$$2f_{\mathrm{m}}T = 2(N - 1/2)\Delta f\, T = (2N - 1)\Delta f\, \Delta t\, M = (2N - 1)M.$$

So Gabor's analysis and Shannon's Sampling Theorem both show that $(2N - 1)M$ real parameters determine a signal of duration T and bandwidth F.

We can also compare these results with the representation of the signal by a finite Fourier series. To do this we treat the signal ψ as periodic with period T; then its highest frequency relative to this period is

$$H = f_{\mathrm{m}}T = (N - 1/2)\Delta f(M\, \Delta t) = (2N - 1)M/2.$$

The signal can be represented exactly by an $H + 1$ term Fourier series:

$$\psi(t) = \sum_{n=0}^{H} d_n \cos(2\pi nt/T) + e_n \sin(2\pi nt/T).$$

[10]Note that $FT = (M\,\Delta f)(N\,\Delta t) = MN(\Delta f\,\Delta t) = MN$. Another way to interpret the formula $M(2N - 1)$ is that for each time interval and frequency band we have two real parameters — an amplitude and a phase — except for the DC band, which has only an amplitude.

There appear to be $2(H+1)$ parameters d_n, e_n, but e_0 is irrelevant since the corresponding sine term is identically zero. It also can be shown that e_n is irrelevant, since (by the Sampling Theorem) the signal is determined by $2H$ points, and over these points the last sine term is linearly dependent on the other terms. Thus ψ can be represented by a Fourier series determined by $2H = (2N-1)M$ real parameters:

$$\psi(t) = d_0 + \sum_{n=1}^{H-1} \{d_n \cos(2\pi nt/T) + e_n \sin(2\pi nt/T)\} + d_H \cos(2\pi Ht/T).$$

Again, the band-limited, finite-length signal is seen to have $(2N-1)M$ real degrees of freedom.[11]

We have seen that any band-limited signal of finite duration can be represented by a finite superposition of Gabor elementary functions. This raises the question of whether arbitrary functions can be represented as (possibly infinite) superpositions of the Gabor functions. In fact it can be shown (Heil & Walnut 1989, pp. 656–657) that the set of Gabor functions ϕ_{jk} is *complete* in $L_2(\mathbf{R})$, the set of square-integrable functions. That is, any signal ψ of finite energy can be written as an infinite sum

$$\psi = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} c_{jk}\phi_{jk}.$$

Equivalently,

$$\psi = \sum_{j,k=0}^{\infty} a_{jk}C_{jk} + b_{jk}S_{jk}.$$

(Note $S_{j0} = 0$, so the b_{j0} are irrelevant.) On the other hand, the Gabor elementary functions do not form a basis for the $L_2(\mathbf{R})$ functions, an issue addressed later (Sections 6 and 9).

There is another way to understand the relation between representations based on Gabor elementary functions, Fourier series, and the Sampling Theorem (Gabor 1946, p. 435). Notice that as $\alpha \to \infty$ the Gabor functions become

$$\begin{aligned} \phi_{jk}(t) &= \exp[2\pi i k \Delta f (t - j\,\Delta t)], \\ C_{jk}(t) &= \cos[2\pi k \Delta f (t - j\,\Delta t)], \\ S_{jk}(t) &= \sin[2\pi k \Delta f (t - j\,\Delta t)]. \end{aligned}$$

That is, in the $\alpha = \infty$ limit the wave packets have *no* locality, and the Gabor representation reduces to the Fourier representation, sinusoids at a spacing Δf. Conversely,

[11] Some authors argue that the logon content is $MN+1$ complex parameters or $(2N-1)M+1$ real parameters; it is also possible to argue an extra degree of freedom in the Fourier and Sampling representations. The practical difference is slight, since typically $MN \gg 1$, but the issue is important for information theory (MacKay 1969, pp. 185–186; Brillouin 1956, p. 97). We leave it unresolved.

as $\alpha \to 0$ the wave packets become more and more localized, and in the limit pass over into Dirac delta functions (impulses) at a spacing of Δt:

$$\begin{aligned}\phi_{jk}(t) &= \delta(t - j\Delta t) + i\delta(t - j\Delta t),\\ C_{jk}(t) &= S_{jk}(t) \;=\; \delta(t - j\Delta t).\end{aligned}$$

We see that the $\alpha = 0$ limit represents two samples (a_{jk} and b_{jk}) for each Δt interval, as required by the Sampling Theorem.

The value of the Gabor representation lies in the locality of the elementary functions. That is, although they are not *strictly* local (of compact support), their sensitivity is concentrated in a small interval of time (measured by the nominal duration). Because of the locality property, the Gabor representation is physically more realistic than the Fourier representation, since it represents a band-limited signal of finite duration (i.e., a physically realistic signal) by a finite superposition of temporally local elementary signals. In contrast, the Fourier representation of such a signal requires an infinite superposition of nonlocal signals, and so depends on enormous cancellation in order to result in a local superposition (Strang 1989, p. 614). This and the fact that the Gabor elementary functions correspond to a quantum of information are good theoretical reasons for choosing them as representational primitives.

4 Gabor Representation of Two-Dimensional Signals

We have seen that any (finite energy) one-dimensional function $\psi : \mathbf{R} \to \mathbf{R}$ can be represented as a linear superposition of Gabor elementary functions, each of which represents one logon or quantum of information about the function. Although we thought of these functions as time-varying signals $\psi(t)$, it should be clear that this is not essential to the theory. $\psi(x)$ could also represent a spatial pattern, in which case the Gabor elementary functions represent information cells localized in *space* and *spatial* frequency. We must make the change to the spatial domain when we come to problems in vision, where it is necessary to consider two-dimensional functions $\psi : \mathbf{R}^2 \to \mathbf{R}$, where $\psi(x, y)$ represents the intensity at spatial location (x, y).

It might be expected that two-dimensional signals could be represented in terms of two-dimensional analogues of Gabor elementary functions, and in the early 1980s a number of researchers suggested Gaussian-modulated sinusoids as models of the receptive fields of simple cells in visual cortex (Marčelja 1980; Daugman 1980; Watson 1982; Pribram & Carlton 1986). Our presentation is based on Daugman (1985a, 1988).

Daugman proved[12] two-dimensional analogues of Gabor's Uncertainty Principle,

$$\Delta x \Delta u \geq 1/4\pi, \quad \Delta y \Delta v \geq 1/4\pi$$

[12]Our own proofs can be found in Appendices A and B.

(where Δu and Δv are the uncertainties in the x and y spatial frequencies), and showed that the elementary information cells are occupied by *Gabor elementary functions* of the form:

$$\phi_{pquv}(x,y) = \exp\left\{-\pi\left[\frac{(x-p)^2}{\alpha^2} + \frac{(y-q)^2}{\beta^2}\right]\right\} \exp\{2\pi i[u(x-p)+v(y-q)]\}. \quad (15)$$

The first factor is a two-dimensional Gaussian distribution centered on the point (p,q); the second factor is the conjugate exponential form of the trigonometric functions, also centered on (p,q). The parameters (u,v) determine the wave packet's location in the frequency domain just as (p,q) determine its location in the spatial domain. The 2D Gabor function's nominal x-spread and y-spread are α and β, and so these parameters determine its two-dimensional shape and spread.[13]

As we did for the Gabor representation of 1D signals, we will make use of Gabor elementary functions located on a regular grid in the spatial and spectral domains. In this case we index the functions by the quantum numbers j, k, l, m:

$$\begin{aligned}\phi_{jklm}(x,y) &= \exp\left\{-\pi\left[\frac{(x-j\,\Delta x)^2}{\alpha^2} + \frac{(y-k\,\Delta y)^2}{\beta^2}\right]\right\} \times \\ &\quad \exp\{2\pi i[l\,\Delta u(x-j\,\Delta x)+m\,\Delta v(y-k\,\Delta y)]\},\end{aligned}$$

where the spacing is determined by $\Delta x\,\Delta u = 1$ and $\Delta y\,\Delta v = 1$.

The spatial frequency of the function in Eq. 15 is $f = \sqrt{u^2+v^2}$ and its orientation is $\theta = \arctan(v/u)$. Conversely, $u = f\cos\theta$ and $v = f\sin\theta$. This gives an alternate form for the elementary functions:

$$\phi_{pqf\theta}(x,y) = \exp\left\{-\pi\left[\frac{(x-p)^2}{\alpha^2} + \frac{(y-q)^2}{\beta^2}\right]\right\} \exp\{2\pi i f(\cos\theta+\sin\theta)\}. \quad (16)$$

The structure of Eq. 15 may be easier to understand by writing it in vector form; let $\mathbf{x} = (x,y)$ be an arbitrary point in the plane, let $\mathbf{p} = (p,q)$ be the center of the function, let $\mathbf{u} = (u,v)$ be the *wave vector* (which represents the packet's frequency along each axis). Finally, let the diagonal matrix

$$S = \begin{pmatrix} \alpha^{-1} & 0 \\ 0 & \beta^{-1} \end{pmatrix}$$

represent the function's shape. Then,

$$\phi_{\mathbf{pu}}(\mathbf{x}) = \exp\{-\pi\|S(\mathbf{x}-\mathbf{p})\|^2\} \exp[2\pi i\mathbf{u}\cdot(\mathbf{x}-\mathbf{p})]. \quad (17)$$

Now it is clear that the 2D Gaussian envelope falls off with the square of the distance from $\mathbf{p}$ scaled in accord with S. Similarly, the periodic part has its origin at $\mathbf{p}$. Since

[13] The standard deviation of the Gaussian on the x-axis is proportional to α, and on the y-axis to β; see Appendix B.

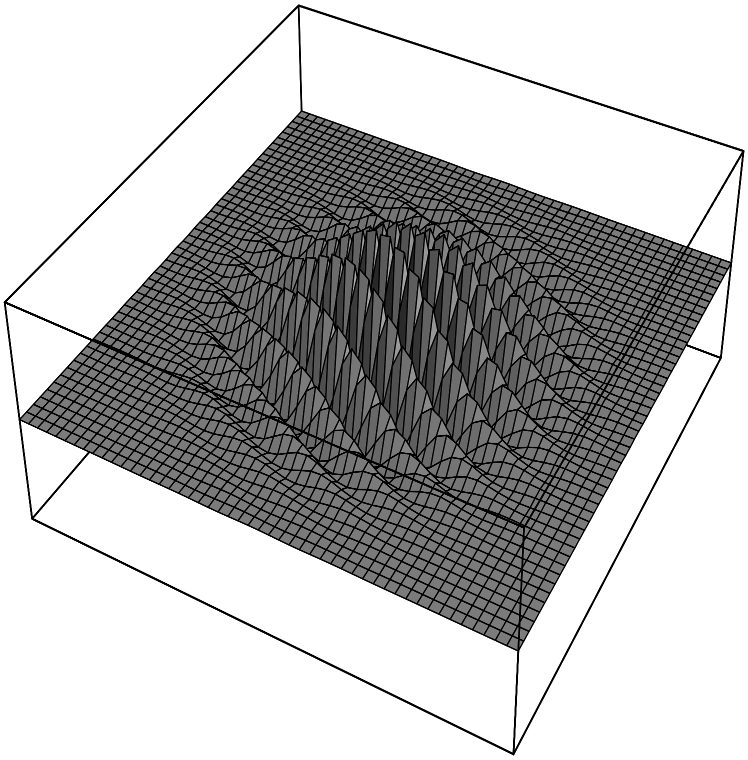

Figure 14: The even (cosine) component of a 2D Gabor elementary function. The function shown has $\alpha^2 = \beta^2 = 20$, $u = 1/2, v = 1$, and $p = q = 0$. It is plotted for all $x, y \in [-6, 6]$.

$\mathbf{u} \cdot (\mathbf{x} - \mathbf{p})$ projects $\mathbf{x} - \mathbf{p}$ onto $\mathbf{u}$, the phase of the periodic function is constant in a direction perpendicular to the wave-vector $\mathbf{u}$. Thus the orientation of the periodic part is given by $\mathbf{u}$ and its frequency is given by $\|\mathbf{u}\| = f$.

The overall shape of the 2D Gabor elementary function is easiest to understand in terms of its even-symmetric (cosine) and odd-symmetric (sine) components, so we separate the periodic part of Eq. 17 to get $\phi_{\mathbf{pu}}(\mathbf{x}) = C_{\mathbf{pu}} + iS_{\mathbf{pu}}$, where

$$\begin{aligned} C_{\mathbf{pu}}(\mathbf{x}) &= \exp\{-\pi\|S(\mathbf{x} - \mathbf{p})\|^2\} \cos[2\pi\mathbf{u} \cdot (\mathbf{x} - \mathbf{p})], \\ S_{\mathbf{pu}}(\mathbf{x}) &= \exp\{-\pi\|S(\mathbf{x} - \mathbf{p})\|^2\} \sin[2\pi\mathbf{u} \cdot (\mathbf{x} - \mathbf{p})]. \end{aligned}$$

One of these 2D Gaussian sinusoids is shown in Fig. 14; it can be described as an *oriented grating patch*.

The 2D Gabor Uncertainty Principle can be understood by looking at Fig. 15. On the left we see a schematic representation of the even component of a 2D Gabor elementary function; on the right we see a schematic representation of its Fourier transform. Now consider Fig. 16, which shows a Gabor function like that in Fig. 15, but wider in the x direction. Looking at the figure we can see that its increased width

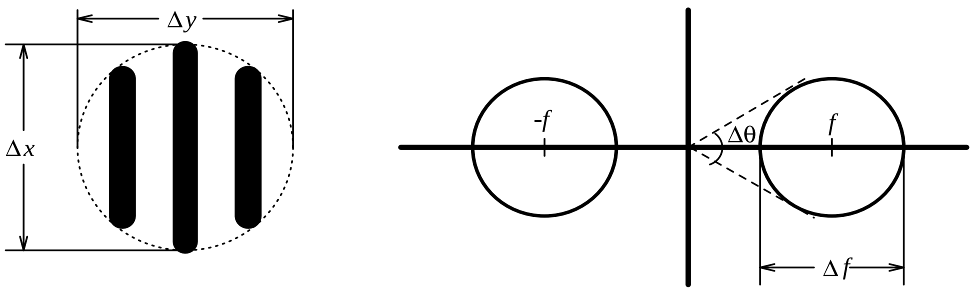

Figure 15: Schematic representation of even component of 2D Gabor elementary function in space and frequency domains (adapted from Daugman (1985b)).

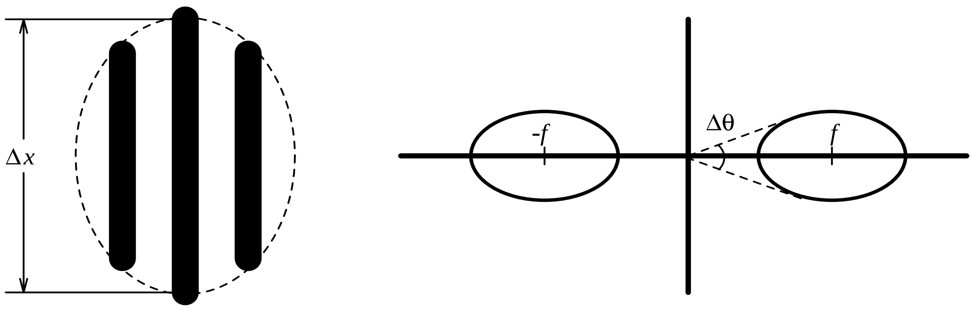

Figure 16: Schematic representation of even component of 2D Gabor elementary function in space and frequency domains showing Δx vs. $\Delta\theta$ tradeoff (adapted from Daugman (1985b)). Thus x and θ are conjugate variables.

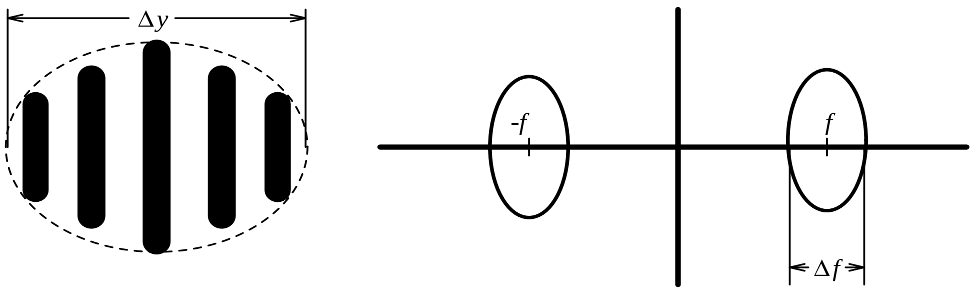

Figure 17: Schematic representation of even component of 2D Gabor elementary function in space and frequency domains showing Δy vs. Δf tradeoff (adapted from Daugman (1985b)). Thus y and f are conjugate variables.

will provide greater sensitivity to orientation, and this can be seen in the frequency domain, where $\Delta\theta \approx 1/(f\Delta x)$ is smaller.[14] Thus there is a tradeoff between localization in the *conjugate variables* x and θ, since $\Delta x \Delta\theta \approx 1/f$. Figure 17 shows the effect of stretching the Gabor function in the y dimension. Just as for one-dimensional signals, the increased number of samples allows a more accurate determination of the frequency, and so decreases $\Delta f = 1/\Delta y$. Thus there is a tradeoff between localization in the conjugate variables y and f, since $\Delta y \Delta f = 1$.

5 Evidence for 2D Gabor Representation in Vision

Daugman (1984, 1985a, 1988) summarizes the considerable physiological evidence that 2D Gabor elementary functions are fundamental to visual processing in several mammalian species. Since proponents of alternate hypotheses, such as those based on wavelets or Laplacian edge detectors, will have to demonstrate that they can account as well for these data, we briefly review the results.

First, measurements of the receptive fields of simple cells in cat visual cortex have shown them to be like Gaussian-modulated sinusoids (Jones & Palmer 1987); Daugman has shown that 97% of them are statistically indistinguishable from the odd- or even-symmetric parts of a 2D Gabor elementary function.

Pollen and Ronner (1981) found a quadrature phase relation between pairs of simple cells in the same cortical column; that is, adjacent simple cells have grating patches that are 90° out of phase, but matched in preferred orientation and frequency. These cells could be computing the odd- and even-symmetric parts of the complex 2D Gabor function, in accord with Euler's formula:

$$e^{2\pi i f x} = \cos 2\pi f x + i \sin 2\pi f x.$$

[14]The relationship $\Delta\theta \approx 1 / f\Delta x$ holds for small $\Delta\theta$.

Alternately, a simple cell could represent the sum of two Gabor elementary functions, in accord with the formulas:

$$\begin{aligned} 2\cos 2\pi f x &= e^{-2\pi i f x} + e^{2\pi i f x}, \\ 2\sin 2\pi f x &= i e^{-2\pi i f x} - i e^{2\pi i f x}. \end{aligned}$$

Daugman (1988) presents an additional argument in favor of the 2D Gabor elementary functions, which is based on their efficiency. An optimal image coding scheme is given by principal components analysis via the Karhunen-Loève transform. However, this method is dependent on the particular image to be encoded. To get an image-independent encoding scheme, we can make the reasonable assumption that the image statistics are locally stationary, in which case the Karhunen-Loève transform is equivalent to a windowed Fourier analysis in each of the local regions. The 2D Gabor representation is a good approximation to this scheme.

Daugman (1984) also conducted a series of *psychophysical* experiments, which allowed him to infer the tuning sensitivities of the entire visual channel (in humans). These were *masking* experiments in which a fixed grating was presented together with a grating that was variable in orientation and spatial frequency. The experiments determined how much the fixed grating interfered with the perception of the variable grating; this was determined by measuring how much the fixed grating raised the threshold at which the variable grating became visible. The underlying assumption is that the neurons in the visual channels that are involved in perceiving the fixed grating become fatigued or saturated, and so the variable grating is difficult to perceive to the extent it shares the same neural channels. Hence the degree of masking measures the response sensitivity of that neural channel to gratings of other frequencies and orientations.

The results of these psychophysical experiments were consistent with the neurophysiological data from cat visual cortex: visual channels have a frequency bandwidth of 1–2 octaves and an orientation half-bandwidth $\pm 15°$ (i.e., $30°$ total angular bandwidth). Furthermore, the regions of sensitivity in the *spectral* domain were elliptical and twice as large in the orientation dimension as in the frequency dimension (i.e., corresponding to Fig. 17). Such a sensitivity profile corresponds to a width/length ratio in the *spatial* domain of $\lambda = \alpha/\beta = 1/2$, in good agreement with neurophysiological data (Jones & Palmer 1987; Movshon 1979).

For optimal (minimum uncertainty) 2D Gabor filters, a relationship can be calculated between the aspect ratio λ, the orientation half-bandwidth $\Delta\theta_{1/2}$ and the spatial frequency bandwidth r in octaves (Daugman 1985b):[15]

$$\Delta\theta_{1/2} = \arcsin\left(\lambda \frac{2^r - 1}{2^r + 1}\right).$$

[15]That is, $2^r = f'/f$, where f and f' are focal frequencies of two filters; we are measuring bandwidth by a ratio rather than a difference. In this case $r = 3/2$.

For the observed $\lambda = 1/2$ and $r = 1.5$ octaves, the formula gives $\Delta\theta_{1/2} = 13.8°$, in good agreement with the observed 15°, and supporting the hypothesis that receptive field profiles are close to Gabor elementary functions. This is reinforced by calculating the area in Fourier space of the inferred filters, which is about 2.5 times the Gabor minimum, whereas other idealized 2D filters have areas of at least 6.5 times the minimum.

6 Problems with the Gabor Representation

One argument against the hypothesis that the vision system uses a 2D Gabor representation is that the Gabor elementary functions are not strictly local; that is, along with their Gaussian envelope, they stretch out to infinity. In mathematical terms, they have *noncompact support*. This is biologically implausible, since receptive fields are at least limited to the area of the retina, and in fact more limited than that. (Daugman notes that receptive fields die out after five or six extrema.) One answer to this argument is that the Gabor representation is intended as an idealized mathematical model, and that we shouldn't expect it to be exactly realized in the biology. In any case, the Gaussian envelope is well localized: 99.7% of its area is within 3 standard deviations of the mean, 99.994% within 4 (see Figs. 11, 12, 14). A receptive field that is statistically indistinguishable from a Gabor function in 97% of the cases is surely a good enough approximation to the mathematical ideal.

A second argument against the Gabor representation is that it is *nonorthogonal.*[16] One result of this is that it is comparatively difficult to compute the coefficients of a 2D Gabor representation. For orthogonal representations, such as the Fourier representation and orthogonal wavelet representations (see the following section), the coefficients are computed by a simple inner product. In contrast, an algorithm for computing the coefficients of a 1D Gabor representation was not published before 1980, and Daugman uses an iterative relaxation algorithm to compute the coefficients for the 2D Gabor representation (Daugman 1988).

Further, Daugman claims that simulation studies have shown that nonorthogonal representations can lead to various artifacts, including edge echos and spurious zero-crossings (Daugman 1988). The paradox is that, as Daugman observes, nonorthogonal representations are ubiquitous in biological sensory and motor systems. Thus, whatever the disadvantages of nonorthogonality, nature seems to have found ways around them; we consider some of the possibilities in Section 9 and in MacLennan & Pribram (1991).

[16] Indeed, the 2D Gabor elementary functions (Eq. 15) do not even generate a *frame*, when $\Delta x\,\Delta u = 1$ or $\Delta y\,\Delta v = 1$ (Heil & Walnut 1989, pp. 656–657; Daubechies & al. 1986, p. 1274). They do generate a frame for certain values of $\Delta u < 1/\Delta x$ and $\Delta v < 1/\Delta y$, in particular, for $\Delta u = 1/m\Delta x$, $\Delta v = 1/n\Delta y$ where $m, n = 2, 3, 4, \ldots$ (Daubechies & al. 1986, p. 1275). See also Section 9.

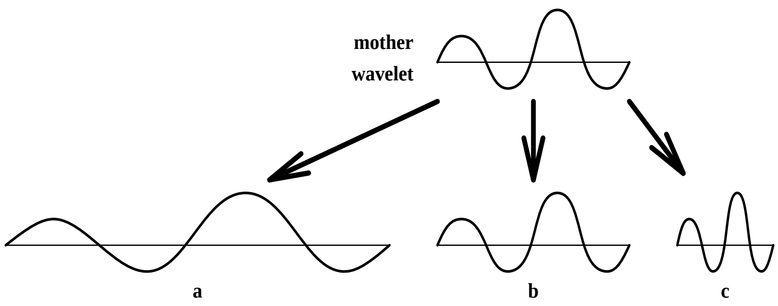

Figure 18: Examples of dyadic dilates of a mother wavelet. The dyadic dilates of a mother wavelet $\phi(x)$ have the form $\phi(2^k x)$ for positive and negative integers k. This figure shows the $k = -1, 0, 1$ dilates (a, b, c, respectively).

7 Gabor versus Wavelet Representations

Wavelets have been proposed as an alternative to Gabor elementary functions as a basis for representation in the visual system. We limit ourselves to a brief introduction.[17]

A family of wavelets is a complete set of functions, all generated from a *mother wavelet* by the operations of *dilation* and *translation*. Most commonly we are concerned with *dyadic* dilations and translations: A *dyadic dilate* of a function $\phi : \mathbf{R} \rightarrow \mathbf{R}$ is a function of the form $\phi(2^k x)$, for some integer k. Thus $\phi(2^{-1}x)$ is ϕ dilated (stretched) by a factor of two (around the origin), and $\phi(2x)$ is ϕ contracted (shrunk) by a factor of two (around the origin); see Fig. 18.[18]

A *dyadic translate* of a dilated function has the form $\phi(2^k x - j)$ for some integer j. The effect of the translation is clearer if we write the function in the equivalent form $\phi[2^k(x - j/2^k)]$, since then we see that the dilate $\phi(2^k x)$ has been translated to all the dyadic points $j/2^k$ (Fig. 19). Thus the general form of the wavelets generated from mother wavelet ϕ by dyadic dilation and translation is:

$$\phi_{jk}(x) = \phi(2^k x - j), \quad j, k \in \mathbf{Z}. \tag{18}$$

Figure 20 shows a well-known mother wavelet, the *Haar wavelet.*

Next we make some observations about wavelets:

1. Although we have discussed wavelets in terms of a one-dimensional mother wavelet, it should be clear that wavelets of higher dimension can be defined in the same way:
$$\phi_{\mathbf{p}k}(\mathbf{x}) = \phi(2^k \mathbf{x} - \mathbf{p}), \tag{19}$$

[17] Several good overviews of wavelets have been published, including Daubechies (1988), Strang (1989) and Heil & Walnut (1989). Our exposition is based mostly on Strang (1989).

[18] Some authors define the dilation by $2^{k/2}\phi(2^k x)$ so that its L_2 norm is the same as that of the mother wavelet. This is convenient if one is trying to construct an orthonormal wavelet basis, but is unnecessary for our purposes here.

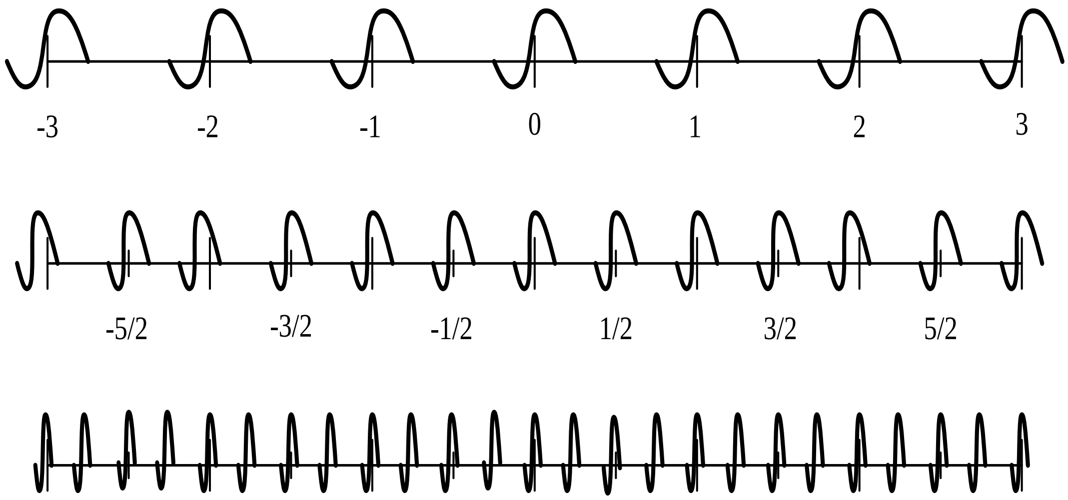

Figure 19: Examples of dyadic translates of a dyadic dilate of a mother wavelet. The mother wavelet $\phi(x)$ generates the family of wavelets $\phi(2^k - j)$ for all integers j and k. The first row of the figure depicts the $k = 0$ wavelets; they are centered on the integers $j = 0, \pm 1, \pm 2, \ldots$ The second row shows the $k = 1$ wavelets, centered on the half-integers $j = 0, \pm 1/2, \pm 2/2, \pm 3/2, \ldots$ The third row shows the $k = 2$ wavelets, centered on $j = 0, \pm 1/4, \pm 2/4, \pm 3/4, \ldots$

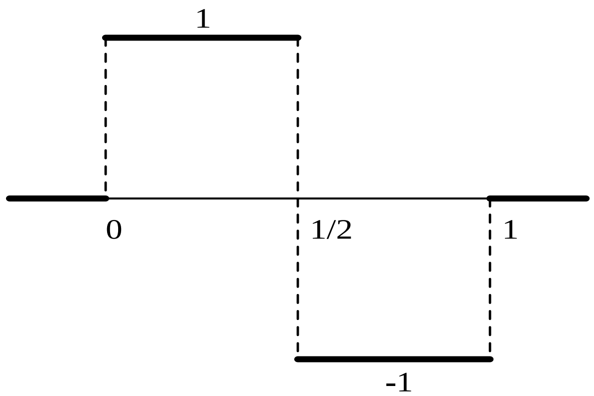

Figure 20: The Haar mother wavelet, which generates an orthogonal family of wavelets. A principal disadvantage of this wavelet is its discontinuity.

for $\phi : \mathbf{R}^n \rightarrow \mathbf{R}$, $\mathbf{p} \in \mathbf{Z}^n$ and $k \in \mathbf{Z}$. Higher-dimensional wavelets are necessary to model vision.

2. Since a wavelet family is by definition *complete*, any (finite energy) function can be represented by a (possibly infinite) linear superposition of wavelets:

$$\psi = \sum_{j,k} c_{jk}\phi_{jk}.$$

   This immediately raises the question of how the wavelet coefficients c_{jk} can be computed; we take it up later.

3. Families of wavelets need not be orthogonal. Although the original definition of 'wavelet' implied orthogonality, the term is now generally used for any complete family generated by dilation and translation. We use the term *orthogonal wavelet* for the mother wavelet of an orthogonal family. (The Haar wavelets, which are based on the mother wavelet in Fig. 20, are orthogonal.)

4. Although our pictures have suggested that wavelets are *strictly local* (i.e., of compact support), this is not necessarily the case. In fact, it is generally difficult to construct families of strictly local wavelets that are orthogonal, and the resulting basis functions tend to be irregular (Strang 1989, p. 615).

Why use wavelets instead of other representations, such as the Fourier or Gabor transforms? One reason is that wavelets permit functions to be represented as linear superpositions of *strictly local* elementary functions. This is especially important when the function to be represented is itself strictly local, since in this case a representation in terms of nonlocal elementary functions depends on enormous cancellation; think of the Fourier representation of a pulse.

Representing strictly local functions in terms of strictly local functions makes sense, and wavelets are well suited to a strictly local representation. If a mother wavelet ϕ is strictly local to an interval $[-L/2, L/2]$ around the origin (i.e., its support is in this interval), then we can see that a wavelet representation of ψ is a *multiresolution*[19] decomposition of the function: the wavelet coefficient c_{jk} gives information about ψ at a scale of $L/2^k$ and in the region $j/2^k$.

Coefficients are easy to compute if the wavelets are orthogonal. The simplest and most familiar way is via the inner product:

$$c_{jk} = \langle \psi, \phi_{jk} \rangle / \|\phi_{jk}\|.$$

There are also more efficient methods, such as *Mallat's Tree Algorithm* (Mallat 1989a, 1989b). Here we note only that the wavelet coefficients can be computed by a simple multilayer linear neural network (Fig. 21). In this algorithm, matrix H is a high-pass filter that computes the wavelet coefficients $\mathbf{b}_j$ at resolution level 2^{-j}, and matrix L is a low-pass filter that passes the "blurred" image on to the next stage.

[19]For a review, see Daubechies (1988).

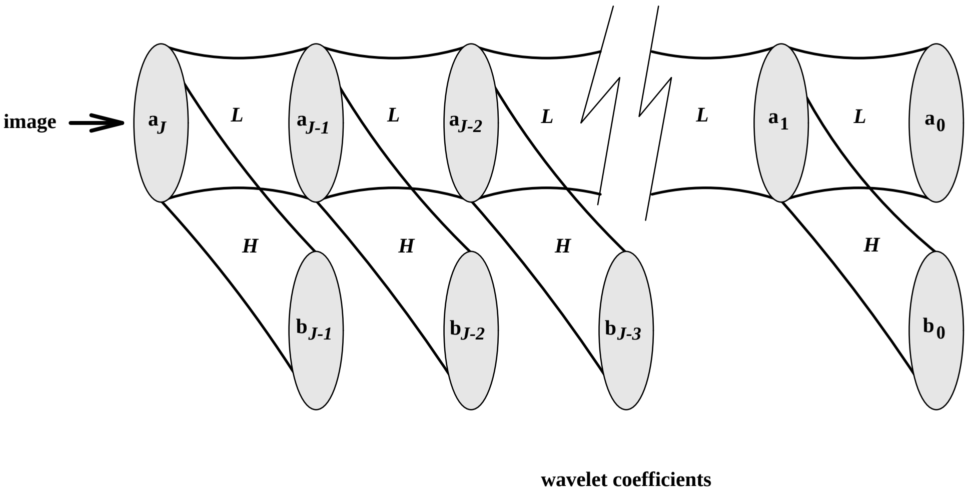

Figure 21: Mallat's Tree Algorithm for wavelet expansion implemented as multilayer linear neural net. Matrix H represents a linear neural network implementing a high-pass filter to extract the wavelet coefficients $\mathbf{b}_j$ at each level of resolution. Matrix L represents a linear neural network implementing a low-pass filter that passes a "blurred" image $\mathbf{a}_j$ on to the next stage for processing.

8 Gabor Wavelets

There are some obvious similarities between wavelets and Gabor elementary functions: they are both complete families of functions, the members of which are predominantly sensitive to variations at a particular scale and at a particular location (in space or time).[20] Indeed, the Gabor functions can be generated from a Gaussian mother function by translation and *periodic modulation*, in the same way that wavelets are generated from a mother wavelet by translation and *dilation* (Heil & Walnut 1989).

Daugman unifies Gabor elementary functions and wavelets by defining *Gabor wavelets* (Daugman 1988). These anisotropic (oriented) wavelets are generated from a fixed Gabor elementary function (Eq. 17) by dilation, translation and rotation.[21] Dilation, of course, also has the ancillary effect of changing the frequency of the Gabor function. This fits well with neurophysiological and psychophysical data indicating a *log-polar* distribution of response selectivity in cells in the visual cortex, which show an orientation half-bandwidth of $\pm 15°$ and a frequency bandwidth of 1.5 octaves (Daugman 1988). That is, a space in which polar angle represents orientation and radial distance represents spatial frequency is efficiently covered by Gabor filters with

[20]Although the Gabor functions are not strictly local, their Gaussian envelope causes their greatest sensitivity to be concentrated near the center of that envelope.

[21]These are not true 2D wavelets, in the usual sense, which are generated from a mother wavelet by dilation and translation (Eq. 19). Thus true 2D Gabor wavelets would have the same orientation as the mother wavelet.

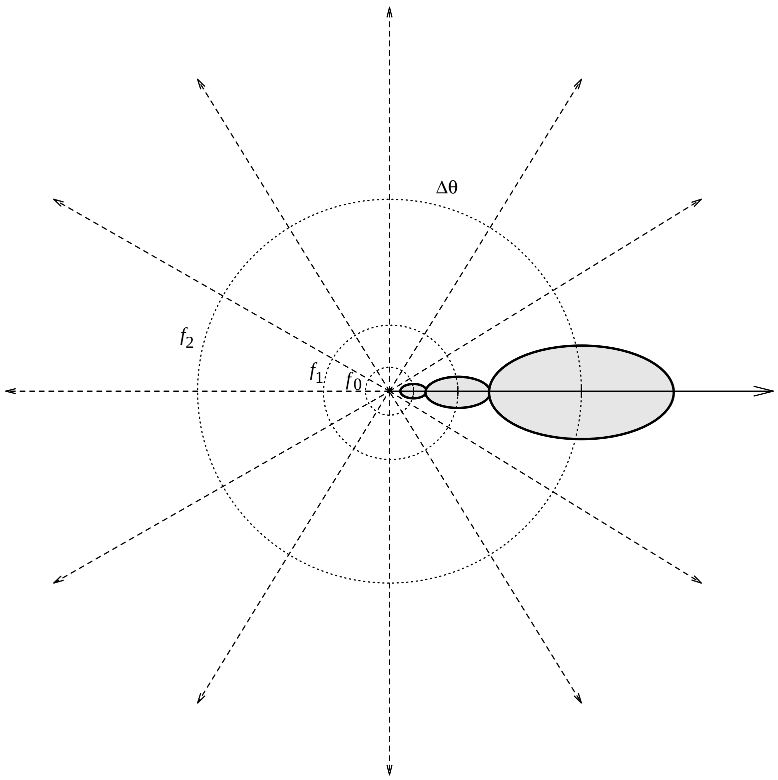

Figure 22: Log-polar distribution of 2D Gabor filters. Shaded ellipses represent envelopes of 2D Gabor filters with aspect ratio $\lambda = 1/2$. The filters are oriented in multiples of $\Delta\theta = 30°$ and have focal frequencies $f_0, f_1, f_2, \ldots$, where $f_k = d^k f_0$ and $d = 2^{3/2}$. Notice how effectively ellipses of these orientations, sizes and aspect ratios cover the space.

aspect ratio $\lambda = 1/2$, orientation a multiple of $30°$, and central frequency at radii in the ratio $2^{3/2}$ (Fig. 22).

9 The Orthogonality Issue

Of course, Gabor wavelets are *not* orthogonal, so their attractive match to the data is coupled with mathematical difficulties. But, orthogonality is a rather delicate property — functions either are or aren't orthogonal; there are no degrees of orthogonality — and so it is probably too fragile for biology to be able to depend on it. Perhaps we should not be surprised that nonorthogonality is ubiquitous in biological systems; rather we should learn how nature lives with it and even exploits it.

The principal difficulty with a nonorthogonal set of elementary functions is in computing the wavelet coefficients. In other words, although we know that any finite-energy ψ can be represented by a linear superposition of the Gabor elementary functions, $\psi = \sum_{jklm} c_{jklm}\phi_{jklm}$, their nonorthogonality means that the coefficients are not defined by a simple inner product, $c_{jklm} = \langle\psi, \phi_{jklm}\rangle / \|\phi_{jklm}\|$. Daugman (1988) has described an iterative relaxation algorithm for expanding an image in terms of Gabor wavelets or other nonorthogonal codes; it operates by gradient descent in the squared L_2 error of representation:

$$\|\psi - \sum_{jklm} c_{jklm}\phi_{jklm}\|^2.$$

It is unlikely that such iterative algorithms are implemented in biological neural networks, since their speed is limited by the neuron impulse rate (say 1 msec. per impulse, with many impulses required to represent an analog quantity).[22] On the other hand, iteration in local circuits in the dendritic net does not depend on impulse generation, and so could proceed much faster, a possibility we explore elsewhere (MacLennan & Pribram 1991).

Although the Gabor coefficients cannot be computed by inner products, the evidence from receptive field studies is that the primary visual system does compute inner products, so we must question their functional role.[23] One possibility is that the inner products may be good estimates of the Gabor coefficients, and so a good place to start the relaxation process. (Daugman's algorithm does this.)

A second possibility we consider is that although the Gabor wavelets are not orthogonal, they may nevertheless be a *frame* (Heil & Walnut 1989), which is a generalization of a basis. If the functions ϕ_{jklm} are a frame, then there is a bounded linear operator S such that

$$\psi = \sum_{jklm} \langle\psi, S^{-1}\phi_{jklm}\rangle\phi_{jklm} = \sum_{jklm} \langle\psi, \phi_{jklm}\rangle S^{-1}\phi_{jklm}.$$

In other words, the inner products $\langle\psi, \phi_{jklm}\rangle$, which are apparantly computed by the primary visual system, give the representation of ψ in terms of the *dual frame* $\{S^{-1}\phi_{jklm}\}$.

Now we must address the question of whether the Gabor wavelets are a frame. It has been known for some time that the 1D Gabor wavelets are *not* a frame for $\Delta f\, \Delta t = 1$, but Daubechies & al. (1986, p. 1275) show that they *are* a frame for $\Delta t = 1 / m\Delta f$ where $m = 2, 3, 4, \ldots$[24]

[22] There is at most 1 KHz. of available bandwidth since that is the maximum spike rate. Therefore, to distinguish N discrete values, we need frequencies separated by at least $\Delta f = 1000/N$. Applying the Gabor Uncertainty Principle gives $\Delta t = N/1000$ seconds, or N milliseconds to reliably transmit the value. Thus it takes at least 10 msec. to transmit an analog value with one digit of precision and at least 100 msec. to transmit it with two digits of precision.

[23] Here, "primary visual system" refers to the retina, lateral geniculate nucleus and primary visual cortex (VI). Since the representational primitives of the retina + LGN system seem to be either radially symmetric differences of Gaussians or radially-symmetric Gaussian sinusoids (Pribram 1991, p. 74), the Gabor coefficients must be computed from the coefficients of these radial basis functions.

[24] These conditions are sufficient, but perhaps not necessary. Also note that for larger m the frame is *tighter*, which means S^{-1} is more nearly a scalar.

Since a 2D Gabor wavelet is the outer product of two 1D Gabor wavelets, $\phi_{jklm}(x, y) = \phi_{jl}(x)\phi_{km}(y)$, it is straight-forward to show that the 2D wavelets are a frame when $\Delta x\,\Delta u = 1/m$ and $\Delta y\,\Delta v = 1/n$, where $m, n = 2, 3, 4, \ldots$ These conditions are compatible with the constraints imposed by the Gabor Uncertainty Principle. For example, $m = n = 13$ gives $\Delta x\,\Delta u = \Delta y\,\Delta v = 1/13$, which is slightly larger than the minimum $1 / 4\pi \approx 1/12.6$. Further, they are consistent with Daugman's (1984) observation that receptive fields occupy about 2.5 times the theoretical minimum area, since in the case $m = n = 8$ the functions occupy $16\pi^2/64 \approx 2.47$ times the minimum area.

Finally, we observe that there is really no *a priori* reason for the visual cortex to compute the Gabor coefficients, because there is no need for it to reconstitute the input image ψ from the coefficients:

$$\psi = \sum_{jklm} c_{jklm}\phi_{jklm}.$$

It must be remembered that this equation is only a mathematically convenient way of guaranteeing that no information is lost in computing the coefficients. Since the visual cortex harbors no homunculus, it does not need to reconstruct the image, and it may work directly in terms of the inner products.

10 3D Gabor Representation of Spatiotemporal Signals

Gabor's research was motivated in part by the observation that our perception of sound is simultaneously of duration and pitch, and therefore that an analysis of sound should be in terms of elements localized in both duration and frequency (Gabor 1946, pp. 431–432). Exactly the same argument may be made for vision. In the spatial domain we see simultaneously both extent and texture (spatial frequency). Likewise, in the temporal domain we perceive simultaneously duration and motion (temporal frequency).

Thus we see that the use of 2D Gabor elementary functions to model visual image representation is unrealistic in a significant way: it ignores the *temporal structure* of images. It is as though vision were merely a succession of separate images, each independent of the next. On the other hand, if we applied to visual images the 1D Gabor functions (Section 3), we would capture their temporal structure, but not their spatial structure, which Daugman and others have shown to be central to understanding vision.

An obvious solution to this problem is to combine the two analyses and consider the evolution in time of two-dimensional spatial signals. Thus we will take the input to the visual cortex to be a three-dimensional signal $\psi(x, y, t)$, $\psi : \mathbf{R}^3 \rightarrow \mathbf{R}$. Sometimes it will be more convenient to write $\psi(\mathbf{x}, t)$ where the vector $\mathbf{x} = (x, y)$ represents spatial position. Such a signal has a Fourier transform $\Psi(\zeta, \eta, \nu)$, where ζ and η are spatial frequencies and ν is a temporal frequency.

Having seen the 1D and 2D Gabor Uncertainty Principles, it is perhaps hardly surprising that there is a 3D Uncertainty Principle holding between pairs of conjugate variables:

$$\begin{aligned}
\Delta x\,\Delta\zeta &\geq 1\,/\,4\pi, \\
\Delta y\,\Delta\eta &\geq 1\,/\,4\pi, \\
\Delta t\,\Delta\nu &\geq 1\,/\,4\pi,
\end{aligned}$$

where we have defined the nominal spreads in terms of the standard deviations of the functions. For a proof of this uncertainty principle, see Appendix A.

It is also straight-forward to show (Appendix B) that the Gabor inequalities become equalities for the 3D Gabor elementary functions, which have the form:

$$\begin{aligned}
\phi_{pqruvw}(x, y, t) &= \exp\left\{-\pi\left[\frac{(x-p)^2}{\alpha^2} + \frac{(y-q)^2}{\beta^2} + \frac{(t-r)^2}{\gamma^2}\right]\right\} \times \\
& \quad \exp\left\{2\pi i[u(x-p) + v(y-q) + w(t-r)]\right\}. \qquad (20)
\end{aligned}$$

The wave packet is localized around space-time coordinates (p, q, r); that is, it is centered at $x = p$, $y = q$ in space and $t = r$ in time. Its location in the corresponding frequency domain is given by (u, v, w), its two spatial frequencies and one temporal frequency. This is apparent from the Fourier transform of ϕ:

$$\begin{aligned}
\Phi_{pqruvw}(\zeta, \eta, \nu) &= \exp\left\{-\pi\left[(\zeta-u)^2\alpha^2 + (\eta-v)^2\beta^2 + (\nu-w)^2\gamma^2\right]\right\} \times \\
& \quad \exp\left\{2\pi i[x(\zeta-u) + y(\eta-v) + t(\nu-w)]\right\}. \qquad (21)
\end{aligned}$$

It can be shown (Appendix B) that the standard deviations of ϕ around the x, y and t axes are proportional to α, β and γ, respectively; thus α, β and γ determine the wave packet's shape. Conversely, the standard deviations of Φ are proportional to α^{-1}, β^{-1} and γ^{-1}.

As before, the Gabor uncertainty relations permit signals to be localized in Fourier space no more accurately than a cell of size

$$\Delta x\,\Delta y\,\Delta t\,\Delta\zeta\,\Delta\eta\,\Delta\nu \geq 1\,/\,64\pi^3.$$

Indeed, the information cells can be no smaller than $1\,/\,4\pi$ in each pair of conjugate variables. Such cells are the information quanta for 3D signals (Fig. 23). The elementary information cells can be indexed by sextuples of quantum numbers, three spatiotemporal and three spectral, $\mathbf{m} = (m_1, m_2, m_3)$, $\mathbf{n} = (n_1, n_2, n_3)$, so that

$$\begin{aligned}
p = m_1\Delta x, &\quad u = n_1\Delta\zeta, \\
q = m_2\Delta y, &\quad v = n_2\Delta\eta, \\
r = m_3\Delta t, &\quad w = n_3\Delta\nu.
\end{aligned}$$

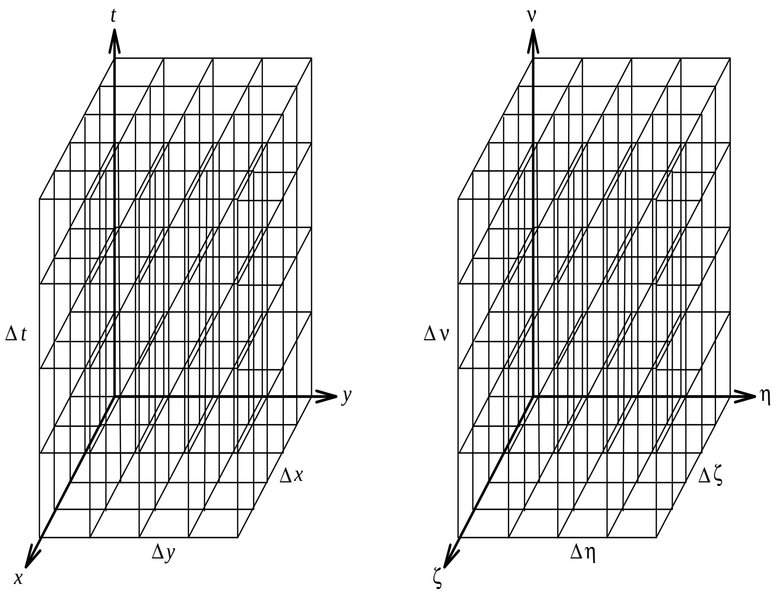

Figure 23: Information cells defining a 3D signal. For clarity the three spatiotemporal dimensions are shown separate from the three spectral dimensions, but it must be born in mind that each information cell is a six-dimensional rectangular space. Each such cell represents a quantum of information in Fourier space.

Then the complex numbers $c_{\mathbf{mn}}$ are the Gabor coefficients of ψ if and only if

$$\psi = \sum_{\mathbf{m}} \sum_{\mathbf{n}} c_{\mathbf{mn}} \phi_{pqruvw},$$

where the indices $\mathbf{m}$ and $\mathbf{n}$ have ranges appropriate to the spatiotemporal extent and bandwidth of ψ. Of course, ψ could be equally well represented by real coefficients and Gaussian sinusoids in quadrature phase.

Next we consider the significance to vision of this quantization of Fourier space. Although it might seem natural to interpret it as a mathematical fiction, Daugman's research has already given us good reason to believe that the visual cortex is organized around a spatial Gabor representation. This suggests that we take the spatiotemporal Gabor representation quite literally and interpret the visual cortex as a bank of filters tuned to spatial frequency bands of width $\Delta\zeta$ and $\Delta\eta$, a temporal frequency band of width $\Delta\nu$, and localized to a spatial region of size $\Delta x\,\Delta y$. We further hypothesize that these filters accumulate information over an interval of time that is a small multiple of Δt, and produce the Gabor coefficients at the end of this interval (perhaps by relaxation during the next interval).[25]

It is natural to identify this interval with the principal rhythm of the occipital (visual) cortex, the alpha rhythm. Slow rhythms, such as the alpha, seem to clock

[25] We may compare the inhibitory wave that seems to reset cerebellar computation every 500 msec. (Pribram 1991, p. 127).

the generation of spike trains, just as we would expect if a set of rate-encoded Gabor coefficients were computed during each interval.[26] During periods of greater activity the alpha rhythm "desynchronizes" and is replaced by a higher frequency oscillation (40–60 Hz.). The results of such a decrease in Δt include greater temporal resolution, poorer temporal-frequency resolution, and less accurate computation of the Gabor coefficients — all reasonable tradeoffs in situations demanding action.

If the hypothesized correlation of Δt with the alpha rhythm is correct, then from the resting alpha frequency, 8 to 12 Hz., we can estimate the resting interval $T_\alpha \approx 100$ msec., with a range of perhaps 80 to 125 msec. Since Δt is the standard deviation of the Gabor elementary function, we can expect that T_α must be 3 or 4 times Δt (so that T_α contains 90–95% of the wave-packet). Since in Appendix B we show $\Delta t = \gamma \,/\, 2\sqrt{\pi}$ (Eq. 30), for mathematical convenience we estimate

$$T_\alpha \approx 2\sqrt{\pi}\Delta t = \gamma.$$

This implies a resting temporal frequency resolution of

$$\Delta w_\alpha = \frac{1}{4\pi\Delta t} \approx \frac{1}{4\pi(T_\alpha \,/\, 2\sqrt{\pi})} = \frac{1}{2\sqrt{\pi}T_\alpha} \approx 2.8 \text{ Hz.}$$

The Gabor representation also sheds light on the perception of form and motion.[27] The parameters u, v and w determine the orientation of the elementary signal in space-time. For example, if $w = 0$ then the wave packet is perpendicular to the t-axis (Fig. 24) and we have the effect of a 2D (spatial) Gabor function, but localized in time. Conversely, if $w \neq 0$ then the elementary function is inclined to the time and space axes (Fig. 25). We can see that such a filter would respond to a grating patch moving at a fixed velocity perpendicular to the fringes. We hypothesize that 3D Gabor elementary functions of this kind explain the response characteristics of *complex* and *hypercomplex* cells in the visual cortex, which have been shown to respond to moving bars and gratings.

We can easily calculate the *phase velocity* at which the fringes move:

$$v_\mathrm{p} = w/f,$$

where $f = \|\mathbf{u}\|$ is the spatial frequency of the grating patch. The fringes move in a direction opposite to the (spatial) wave vector $\mathbf{u}$ (Fig. 25), so the velocity vector $\mathbf{v}$ of the fringes is $-v_\mathrm{p}$ times the *wave normal* $\mathbf{u}/\|\mathbf{u}\|$:

$$\mathbf{v} = -v_\mathrm{p}\mathbf{u}/\|\mathbf{u}\| = -w\mathbf{u}/f^2.$$

[26] We refer here to Bland's studies of the theta rhythm in the dentate gyrus of the rabbit (Bland & al. 1978), but the same principle applies to the alpha rhythm.

[27] Of course, so would other representations in terms of signals localized in both the space-time and spectral domains, such as other 3D Fourier transforms windowed in space-time, or 3D wavelets.

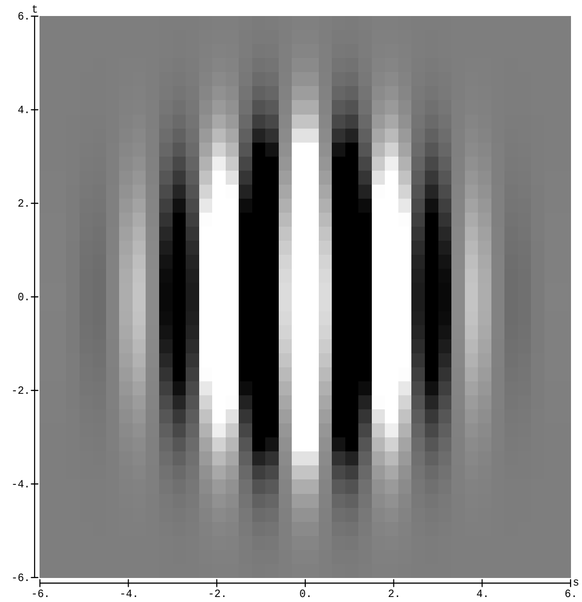

Figure 24: Slice through even (cosine) component of 3D Gabor elementary function oriented perpendicular ($w = 0$) to the time axis (ordinate). The abscissa is taken to be along the wave-vector $\mathbf{u}$, and so perpendicular to the spatial wavefronts. Such a filter is selective for stationary spatial frequency f, localized in both space and time. This function has $f = 1/2$, $\alpha^2 + \beta^2 = \gamma^2 = 20$, $u = v = 1/\sqrt{8}$, $w = 0$, $\mathbf{p} = 0$ and $r = 0$. Lighter regions are more positive, darker more negative.

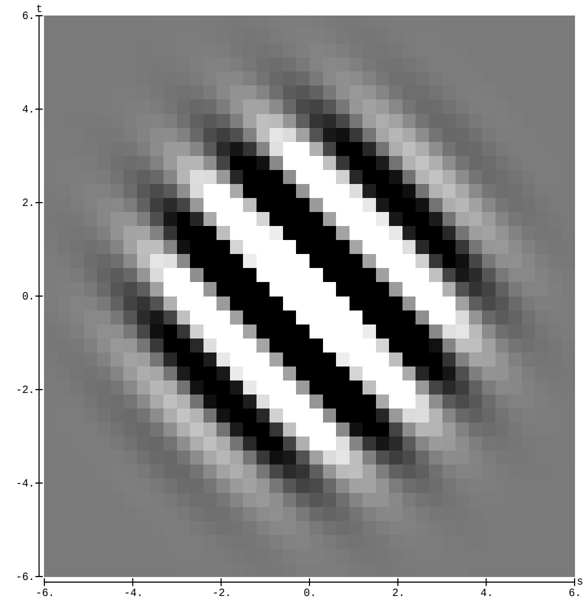

Figure 25: Slice through even (cosine) component of 3D Gabor elementary function inclined to the time axis ($w \neq 0$). Such a filter is selective for a spatial grating of frequency f, moving at velocity v_p, and localized in both space and time. This function has $\alpha^2 + \beta^2 = \gamma^2 = 20$, $u = v = 1/4$, $w = 1/\sqrt{8}$, $\mathbf{p} = 0$ and $r = 0$. It is selective for fringes of frequency $f = 1/\sqrt{8}$ moving at a phase velocity $v_\mathrm{p} = w/f = 1$ to the left.

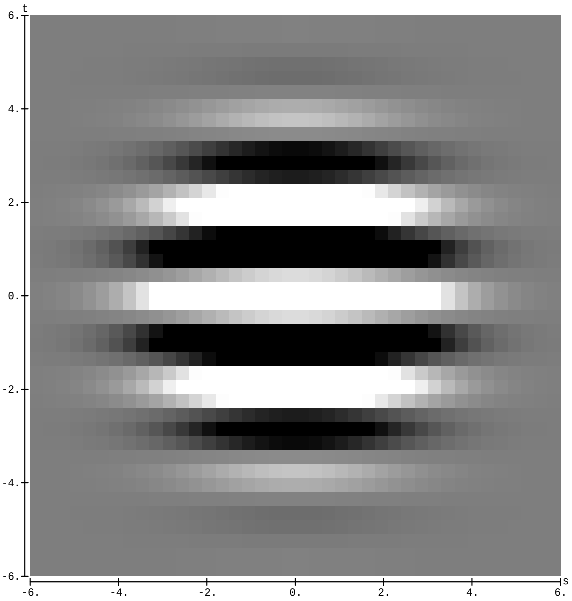

Figure 26: Slice through even (cosine) component of 3D Gabor elementary function oriented parallel to the time axis ($\mathbf{u} = 0$). Such a filter is selective for a spatially uniform intensity oscillating at frequency w, localized in both space and time. This function has $\alpha^2 + \beta^2 = \gamma^2 = 20$, $u = v = 0$, $w = 1/2$, $\mathbf{p} = 0$ and $r = 0$.

We consider briefly the case in which the Gabor function is parallel to the time axis (Fig. 26), that is, $\mathbf{u} = 0$. Such a filter would respond to a uniform intensity (within its spatial receptive field) oscillating at a frequency w. We are unaware of research looking for cells with this kind of response, but it is interesting that much of the work on receptive fields has made use of *flashing* spots, and so might be consistent with the existence of such cells.

11 Conclusions

We have reviewed Gabor's Uncertainty Principle and Daugman's evidence for 2D Gabor filters in the visual cortex. We compared the Gabor representation with wavelet-based representations, and concluded that the Gabor representation is preferable. This is in spite of the Gabor functions not being orthogonal. Indeed we argued that a nonorthogonal set of elementary functions might be preferable in a biological context. Finally we argued that since vision must be understood in terms of images evolving in time, the appropriate representational primitives are 3D Gabor functions. We suggested that these functions could explain the selectivity for moving edges exhibited by complex and hypercomplex cells in visual cortex, and we suggested that the alpha rhythm may correspond to the interval at which the Gabor coefficients are computed. More concrete predictions will depend on finding empirical data to constrain the parameters of the Gabor elementary functions.

A Proof of General Gabor Uncertainty Principle

We prove a general Gabor Uncertainty Principle for n-dimensional functions.[28] Let ϕ be a function and Φ its Fourier transform; for convenience we assume $\|\phi\| = \|\Phi\| = 1$ (this is just a matter of units). We will also assume that ϕ decays to 0 at infinity; specifically we assume $s_k\phi^2(s_1, \ldots, s_n) \to 0$ for all k. For this proof we will not be able to use the simple definition of nominal spread that we used for 1D signals; instead we define the nominal spread of a signal along the kth axis to be its standard deviation along that axis. Thus the spread along the kth axis is given by

$$\Delta s_k = \sqrt{\mathrm{Var}_k\{\phi\}},$$

where Var_k is the variance along the kth axis. Similarly, in the spectral domain we define,

$$\Delta\sigma_k = \sqrt{\mathrm{Var}_k\{\Phi\}}.$$

The variances are given by:

$$\begin{array}{rclclcl} \mathrm{Var}_k\{\phi\} & = & \|s_k\phi\|^2 & = & \int_{\mathcal{S}} |s_k\phi|^2 \mathrm{d}\mathbf{s} & = & \int_{\mathcal{S}} \phi s_k^2 \phi^* \mathrm{d}\mathbf{s}, \\ \mathrm{Var}_k\{\Phi\} & = & \|\sigma_k\Phi\|^2 & = & \int_{\Sigma} |\sigma_k\Phi|^2 \mathrm{d}\boldsymbol{\sigma} & = & \int_{\Sigma} \Phi\sigma_k^2\Phi^* \mathrm{d}\boldsymbol{\sigma}, \end{array}$$

where $\mathcal{S} = \Sigma = \mathbf{R}^n$, $\mathbf{s} = (s_1, \ldots, s_n) \in \mathcal{S}$ and $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n) \in \Sigma$. Our goal will be to show

$$\mathrm{Var}_k\{\phi\}\,\mathrm{Var}_k\{\Phi\} \geq \frac{1}{16\pi^2}.$$

By the Schwartz inequality we know

$$\|s_k\phi\|^2 \left\| \frac{\partial\phi}{\partial s_k} \right\|^2 \geq \left\langle s_k\phi, \frac{\partial\phi}{\partial s_k} \right\rangle^2, \tag{22}$$

where the bracketed expression on the right is an inner product. Since the Fourier transform is an isometry, it preserves the norm, so the norm of $\partial\phi\,/\,\partial s_k$ is the same as the norm of its Fourier transform, which is $2\pi i\sigma_k\Phi$. Therefore we can rewrite the left-hand side of Eq. 22 as follows:

$$\begin{aligned} \|s_k\phi\|^2 \|\partial\phi\,/\,\partial s_k\|^2 &= \|s_k\phi\|^2 \|2\pi i\sigma_k\Phi\|^2 \\ &= 4\pi^2 \|s_k\phi\|^2 \|\sigma_k\Phi\|^2 \\ &= 4\pi^2 \mathrm{Var}_k\{\phi\}\mathrm{Var}_k\{\Phi\}. \end{aligned} \tag{23}$$

[28] The proof is a generalization of that in Hamming (1989, pp. 181–184), which is based on Gabor (1946), which is in turn based on the Heisenberg-Weyl derivation of the uncertainty principle in physics.

Now we work on the right-hand side of Eq. 22:

$$\begin{aligned}\left\langle s_k\phi, \frac{\partial \phi}{\partial s_k}\right\rangle &= \int_{\mathcal{S}} s_k\phi\frac{\partial \phi^*}{\partial s_k}\mathrm{d}\mathbf{s} \\ &= \int_{\mathcal{S}-\mathbf{R}}\int_{\mathbf{R}} s_k\phi\frac{\partial \phi^*}{\partial s_k}\mathrm{d}s_k\mathrm{d}\mathbf{s}' \end{aligned} \tag{24}$$

where $\mathbf{s}' = (s_1, \ldots, s_{k-1}, s_{k+1}, \ldots, s_n) \in \mathcal{S} - \mathbf{R}$. We apply integration by parts to the innermost integral ($U = s_k\phi$, $V = \phi^*$):

$$\begin{aligned}\int s_k\phi\frac{\partial \phi^*}{\partial s_k}\mathrm{d}s_k &= \int s_k\phi\mathrm{d}\phi^* \\ &= s_k\phi\phi^*|_{-\infty}^{\infty} - \int \phi^*\mathrm{d}(s_k\phi) \\ &= s_k|\phi|^2\Big|_{-\infty}^{\infty} - \int \phi^* s_k\mathrm{d}\phi - \int \phi^*\phi\mathrm{d}s_k.\end{aligned}$$

By our assumption that $s_k\phi^2 \to 0$ we know $s_k|\phi|^2\,|_{-\infty}^{\infty} = 0$, so

$$\int s_k\phi\mathrm{d}\phi^* = -\int s_k\phi^*\mathrm{d}\phi - \int \phi^*\phi\mathrm{d}s_k$$

and so

$$\int s_k\phi\frac{\partial \phi^*}{\partial s_k}\mathrm{d}s_k = -\frac{1}{2}\int \phi^*\phi\mathrm{d}s_k.$$

Substituting this into Eq. 24 yields

$$\begin{aligned}\left\langle s_k\phi, \frac{\partial \phi}{\partial s_k}\right\rangle &= \int_{\mathcal{S}-\mathbf{R}}\left(-\frac{1}{2}\int \phi^*\phi\mathrm{d}s_k\right)\mathrm{d}\mathbf{s}' \\ &= -\frac{1}{2}\int_{\mathcal{S}}\phi^*\phi\,\mathrm{d}\mathbf{s} \\ &= -1/2, \end{aligned} \tag{25}$$

since ϕ is normalized (by assumption). Therefore, combining Eqs. 22, 23 and 25,

$$4\pi^2\mathrm{Var}_k\{\phi\}\mathrm{Var}_k\{\Phi\} \geq 1/4,$$

and so,

$$\mathrm{Var}_k\{\phi\}\mathrm{Var}_k\{\Phi\} \geq 1\,/\,16\pi^2.$$

We have proved the general Uncertainty Principle,

$$\Delta s_k\,\Delta\sigma_k \geq 1\,/\,4\pi.$$

B Proof of Optimality of General Gabor Elementary Functions

Our task is to show that the n-dimensional Gabor elementary functions achieve the minimum area in $2n$-dimensional Fourier space. Thus we must show

$$\frac{\|s_k\phi\|^2}{\|\phi\|^2}\frac{\|\sigma_k\Phi\|^2}{\|\Phi\|^2} = \frac{1}{16\pi^2}. \tag{26}$$

Without loss of generality we assume that ϕ and Φ are centered at the origin (since this won't alter their variance). Notice that both functions can be written as a *cisoid* (complex sinusoid) times a product of Gaussians:

$$\phi(\mathbf{s}) = \exp(2\pi i\mathbf{s}\cdot\mathbf{u})\prod_{j=1}^{n}\exp(-\pi s_j^2/\alpha_j^2), \tag{27}$$

$$\Phi(\boldsymbol{\sigma}) = \exp(2\pi i\mathbf{p}\cdot\boldsymbol{\sigma})\prod_{j=1}^{n}\exp(-\pi\sigma_j^2\alpha_j^2) \tag{28}$$

When we compute the norms and variances of these functions the periodic parts can be ignored, since they have a constant modulus 1. Therefore,

$$\begin{aligned}\|\phi\|^2 &= \int\cdots\int\left\{\prod_{j=1}^{n}\exp(-2\pi s_j^2/\alpha_j^2)\right\}\mathrm{d}s_1\cdots\mathrm{d}s_n\\ &= \prod_{j=1}^{n}\int\exp(-2\pi s_j^2/\alpha_j^2)\mathrm{d}s_j.\end{aligned} \tag{29}$$

The integrands are Gaussians, so to understand their structure better, rewrite them as normal distributions:

$$\exp(-2\pi s_j^2/\alpha_j^2) = \frac{\alpha_j}{\sqrt{2}}\left\{\frac{1}{\sqrt{2\pi}(\alpha_j\,/\,2\sqrt{\pi})}\exp\left[-\frac{s_j^2}{2(\alpha_j^2\,/\,4\pi)}\right]\right\}.$$

The expression in curly braces is a normal distribution with mean $= 0$ and variance $\sigma^2 = \alpha_j^2\,/\,4\pi$. Therefore rewrite it $N_\sigma(s_j)$:

$$\exp(-2\pi s_j^2/\alpha_j^2) = \frac{\alpha_j}{\sqrt{2}}N_\sigma(s_k).$$

Since N_σ is a probability distribution, $\int N_\sigma(s_j)\mathrm{d}s_j = 1$, so from Eq. 29, $\|\phi\|^2$ is the product of the normalization factors $\alpha_j/\sqrt{2}$:

$$\|\phi\|^2 = \frac{\prod_j\alpha_j}{2^{n/2}},$$

which is the first formula we need.

Next consider the variance of ϕ. It too can be rewritten, as a product of Gaussians and a quadratic factor:

$$\begin{aligned}\|s_k\phi\|^2 &= \int\cdots\int s_k^2 \exp[-2\pi(s_1^2/\alpha_1^2+\cdots+s_n^2/\alpha_n^2)]\mathrm{d}s_1\cdots\mathrm{d}s_n \\ &= \int s_k^2\exp(-2\pi s_k^2/\alpha_k^2)\mathrm{d}s_k \times \prod_{j\neq k}\int \exp(-2\pi s_j^2/\alpha_j^2)\mathrm{d}s_j.\end{aligned}$$

These can be rewritten in terms of normal distributions:

$$\|s_k\phi\|^2 = \frac{\alpha_k}{\sqrt{2}}\int s_k^2 N_\sigma(s_k)\mathrm{d}s_k \times \prod_{j\neq k}\frac{\alpha_j}{\sqrt{2}}\int N_\sigma(s_j)\mathrm{d}s_j.$$

The first integral is the variance of the normal distribution, which we saw to be $\sigma^2 = \alpha_k^2\,/\,4\pi$; the remaining integrals are again 1, so

$$\begin{aligned}\|s_k\phi\|^2 &= \frac{\alpha_k}{\sqrt{2}}\frac{\alpha_k^2}{4\pi}\prod_{j\neq k}\frac{\alpha_j}{\sqrt{2}} \\ &= \frac{\alpha_k^2}{4\pi}\frac{\prod\alpha_j}{2^{n/2}}.\end{aligned}$$

Hence we see that the normalized variance of ϕ around the kth axis is

$$\frac{\|s_k\phi\|^2}{\|\phi\|^2} = \frac{\alpha_k^2}{4\pi}\frac{2^{-n/2}\prod\alpha_j}{2^{-n/2}\prod\alpha_j} = \frac{\alpha_k^2}{4\pi}.$$

Thus α_k is the standard deviation along the kth axis, scaled by $2\sqrt{\pi}$:

$$\Delta s_k = \frac{\alpha_k}{2\sqrt{\pi}}. \tag{30}$$

Exactly the same analysis can be applied to the Fourier transform Φ, except that the variance of the normal distribution N_σ is $\sigma^2 = 1\,/\,4\pi\alpha_j^2$ and the normalization factor is $1\,/\,\sqrt{2}\alpha_j$. Hence,

$$\begin{aligned}\|\Phi\|^2 &= \frac{1}{2^{n/2}\prod_j\alpha_j}, \\ \|\sigma_k\Phi\|^2 &= \frac{1}{4\pi\alpha_k^2}\frac{1}{2^{n/2}\prod_j\alpha_j}.\end{aligned}$$

Hence,

$$\frac{\|\sigma_k\Phi\|^2}{\|\Phi\|^2} = \frac{1}{4\pi\alpha_k^2},$$

and we see that the standard deviation along the kth spectral axis is α_k^{-1}, scaled by $2\sqrt{\pi}$.

Now the product of the variances in the space-time and spectral domains is easy to compute:

$$\frac{||s_k\phi||^2}{||\phi||^2}\frac{||\sigma_k\Phi||^2}{||\Phi||^2} = \frac{\alpha_k^2}{4\pi}\frac{1}{4\pi\alpha_k^2} = \frac{1}{16\pi^2}$$

and we see that the n-dimensional Gabor elementary functions achieve the minimum area given by the uncertainty principle.

References

[Bland & al., 1978] Bland, B. H., Anderson, P., Ganes, T., and Sveen, O. "Automated Analysis of Rhythmicity of Physiologically Identified Hippocampal Formation Neurons." *Experimental Brain Research.* Vol. 38 (1978), 205–219.

[Brillouin, 1956] Brillouin, L. *Science and Information Theory.* New York, NY: Academic Press, 1956.

[Cherry, 1978] Cherry, C. *On Human Communication.* Cambridge, MA: MIT Press, 1978.

[Daubechies, 1988] Daubechies, I. "Orthonormal Bases of Compactly Supported Wavelets." *Comm. Pure Applied Math.* Vol. 41 (1988), 909–996.

[Daubechies & al., 1986] Daubechies, I., Grossman, A., and Meyer, Y. "Painless Nonorthogonal Expansions." *J. Math. Phys.* Vol. 27 (1986), 1271–1283.

[Daugman, 1980] Daugman, J. G. "Two-Dimensional Spectral Analysis of Cortical Receptive Field Profiles." *Vision Research.* Vol. 20 (1980), 847–856.

[Daugman, 1984] Daugman, J. G. "Spatial Visual Channels in the Fourier Plane." *Vision Research.* Vol. 24 (1984), 891–910.

[Daugman, 1985a] Daugman, J. G. "Representational Issues and Local Filter Models of Two-dimensional Spatial Visual Encoding." In: Rose, D., & Dobson, V. G. (eds.). *Models of the Visual Cortex.* New York: John Wiley & Sons, 1985, 96–107.

[Daugman, 1985b] Daugman, J. G. "Uncertainty Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized by Two-dimensional Visual Cortical Filters." *Journal of the Optical Society of America A.* Vol. 2 No. 7 (1985), 1160–1169.

[Daugman, 1988] Daugman, J. G. "An Information-theoretic View of Analog Representation in Striate Cortex." In: Schwartz, E. L. (ed.). *Computational Neuroscience.* Cambridge, MA: MIT Press, 1988, 403–423.

[Gabor, 1946] Gabor, D. "Theory of Communication." *Journal of the Institution of Electrical Engineers*. Vol 93 Part III (1946), 429–457.

[Hamming, 1989] Hamming, R. W. *Digital Filters*, third edition. Englewood-Cliffs, NJ: Prentice-Hall, 1989.

[Hartley, 1928] Hartley, R. V. L. "Transmission of Information." *Bell System Technical Journal*. Vol. 7 (1928), 535–563.

[Heil & Walnut, 1989] Heil, C. E., & Walnut, D. F. "Continuous and Discrete Wavelet Transforms." *SIAM Review*. Vol. 31 No. 4 (December 1989), 628–666.

[Jones & Palmer, 1987] Jones, J., & Palmer, L. "An Evaluation of the Two-dimensional Gabor Filter Model of Simple Receptive Fields in Cat Striate Cortex." *Journal of Neurophysiology*. Vol. 58, No. 6 (1987), 1233–1258.

[MacKay, 1969] MacKay, D. M. *Information, Mechanism and Meaning*. Cambridge, MA: MIT Press, 1969.

[MacLennan & Pribram, 1991] MacLennan, B. J., & Pribram, K. H. "Neural Computation Without Sigmoids." In preparation.

[Mallat, 1989a] Mallat, S. "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11 (1989), 674–693.

[Mallat, 1989b] Mallat, S. "Multiresolution Approximations and Wavelet Orthonormal Bases of $L^2(R)$." *Transactions American Mathematical Society*. Vol. 315 No. 1 (September 1989), 69–87.

[Marčelja, 1980] Marčelja, S. "Mathematical Description of the Responses of Simple Cortical Cells." *Journal of the Optical Society of America*. Vol. 70 (1980), 1297–1300.

[Movshon, 1979] Movshon, J. "Two-dimensional Spatial Frequency Tuning of Cat Striate Cortical Neurons." *Society for Neuroscience Abstracts*. Vol. 9 (1979), 799.

[Pollen & Ronner, 1981] Pollen, D., & Ronner, S. "Phase Relationships Between Adjacent Simple Cells in the Visual Cortex." *Science*. Vol. 212 (1981), 1409–1411.

[Pribram, 1991] Pribram, Karl H. *Brain and Perception: Holonomy and Structure in Figural Processing*. Hillsdale, NJ: Lawrence Erlbaum, 1991.

[Pribram & Carlton, 1986] Pribram, K. H., & Carlton, E. H. "Holonomic Brain Theory in Imaging and Object Perception." *Acta Psychologica.* Vol. 63 (1986), 175–210.

[Shannon, 1948] Shannon, C. E. "A Mathematical Theory of Communication." *Bell System Technical Journal*, Vol. 27 (1928), 379–423, 623–656. Reprinted in: Claude E. Shannon & Warren Weaver, *The Mathematical Theory of Communication*, Urbana: University of Illinois Press, 1963.

[Strang, 1989] Strang, G. "Wavelets and Dilation Equations: A Brief Introduction." *SIAM Review.* Vol. 31 No. 4 (December 1989), 614–627.

[Watson, 1982] Watson, A. B. "Summation of Grating Patches Indicate Many Types of Detector at One Retinal Location." *Vision Research.* Vol. 22 (1982), 17–25.

[Yu, 1976] Yu, F. T. S. *Optics and Information Theory.* New York: John Wiley & Sons, 1976.