

The role of Fourier theory and of modulation in the prediction of visible moiré effects

Isaac Amidror* and Roger D. Hersch

Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

(Received 20 March 2009; final version received 7 April 2009)

Modern moiré theory is largely based on the Fourier approach. And yet, it seems that the Fourier theory by itself cannot answer all questions related to moiré effects. We present some visible phenomena in the superposition of periodic structures (such as line gratings) that are not captured by the Fourier approach. We discuss their significance, and provide possible explanations. In particular, we introduce the infinite module spanned by the frequencies of the original individual layers as a main tool in the moiré theory. We discuss its significance on the visible periodicities in the layer superposition – either real moiré effects, or pseudo-moiré (modulation) effects having no corresponding impulses in the Fourier spectral domain.

Keywords: moiré; Fourier theory; modulation; spectrum

1. Introduction

When using a mathematical model to explain a physical or an experimental phenomenon, it often happens that although the model being used is well adapted to the situation at hand, in some particular circumstances it gives unexpected results, or simply fails. In some cases this may indeed indicate that we have reached the limits of the present mathematical model (like the limits of classical mechanics, using a well-known example), and that a new refinement of the model or even a new theoretical breakthrough (like relativistic or quantum mechanics, in the example above) may be needed. In other cases the reason may be more prosaic, such as some misunderstood or disregarded conditions, an inadapted hypothesis, or simply an error (in the model, in the interpretation of the results, or even in a computer program).

In the following, we discuss this question in the context of the moiré theory. We describe cases in which the mathematical Fourier-based model being used seems to deviate from our real observations, and we try to provide some plausible explanations.

It should be noted that the moiré theory is, indeed, a good test case for studying the limitations of a mathematical model, since in this case no approximations (such as least square fits or statistical considerations) are explicitly being used, and in principle the model should be able to precisely describe the phenomena in question, without any approximation errors or other inherent error sources. Of course, the use of any mathematical model to describe the reality implicitly implies some assumptions, simplifications

2. Background: the Fourier basis of moiré theory

The Fourier theory is, indeed, the cornerstone of the modern moiré theory. As explained in Chapter 2 of [1], the most adapted approach for investigating phenomena in the superposition of periodic structures is the spectral approach, which takes its roots in the Fourier theory. The spectral approach is based on the duality between functions or images in the spatial image domain and their spectra in the spatial frequency domain, through the Fourier transform. A key property of the Fourier transform is its ability to allow one to examine a function or an image from the perspective of both the space and frequency domains. This approach enables us, therefore, to analyze properties not only in the original layers and in their superposition but also in their spectral representations, and thus it offers a more profound insight into the problem and provides indispensable tools for exploring it. The advantages that the spectral approach offers in

or approximations. For example, when using the Fourier theory we may simplify things a lot by assuming that our structures in the image domain (gratings, etc.) are perfectly periodic, and that they extend to infinity in all directions (in order to obtain pure impulses in the spectral domain). But this 'idealization' of the real world does not introduce into the model explicit error considerations or uncertainties as in models that are based on best fit or statistical approaches.

^{*}Corresponding author.

the study of moiré phenomena are reviewed in detail in Section 2.14 of [1].

And yet, it turns out that the Fourier theory by itself cannot answer all questions regarding the moiré effects. For example, although the Fourier theory is very well suited to deal with global or macrostructure effects in the layer superposition, it is not adapted to deal with local, microstructure effects such as rosette shapes in the periodic case (see [1], Chapter 8) or dot trajectories in the aperiodic case (see [2], Chapter 4).

But while this limitation of the Fourier approach is well understood and easily accepted,¹ it is more surprising to realize that there exist even some macrostructure phenomena related to the moiré theory that are not readily captured by the Fourier approach. In the following sections we will review such cases and try to provide some plausible explanations. But before that, let us provide here a short reminder on the Fourier basics of the moiré theory.

As explained in Section 2.2 of [1], each monochrome image can be represented in the image domain by a *reflectance* function, which assigns to any point (x, y) of the image a value between 0 and 1 representing its light reflectance: 0 for black (i.e. no reflected light), 1 for white (i.e. full light reflectance), and intermediate values for in-between shades. In the case of transparencies, the reflectance function is replaced by a trans*mittance* function which is defined in a similar way. A superposition of such images can be obtained by means of overprinting, or by laying printed transparencies on top of each other. Since the superposition of black and any other shade always gives here black, this suggests a multiplicative model for the superposition of monochrome images. Thus, when m monochrome images are superposed, the reflectance of the resulting image is given by the product of the reflectance functions of the individual images:

$$r(x, y) = r_1(x, y) \cdot \dots \cdot r_m(x, y). \tag{1}$$

The same rule applies also to the superposition of monochrome transparencies, in which case $r_i(x, y)$ and r(x, y) simply represent transmittance rather than reflectance functions. Now, according to the convolution theorem (see [3], p. 244), the Fourier transform of a function product is the convolution of the Fourier transforms of the individual functions. Therefore, if we denote the Fourier transform of each function by the respective capital letter and the 2D convolution by **, the spectrum of the superposition is given by:

$$R(u, v) = R_1(u, v) * * \cdots * * R_m(u, v).$$
(2)

However, the multiplicative model is not the only possible superposition rule, and in other situations different superposition rules can be appropriate.

For example, when images are superposed by projection onto the same frame (such as a white screen on the wall) or by making multiple exposures on a positive photographic film (assuming that we do not exceed the linear part of the film's response), intensities at each point are summed up, which implies an *additive* rule of superposition. In another example, when images are superposed by making multiple exposures on a negative photographic film (again, assuming a linear response) an *inverse additive* rule can be appropriate. More exotic superposition rules (involving, for example, various Boolean operations etc.) can be artificially generated by computer, even if they do not correspond to any physical reality. The interested reader may find examples which illustrate various superposition rules in references like [4,5] or Chapter 3 of [6]. Note that different superposition rules in the image domain will have different spectrum composition rules in the spectral domain, which are determined by properties of the Fourier transform. For example, in the case of the additive superposition rule, where Equation (1) is replaced by

$$r(x, y) = r_1(x, y) + \dots + r_m(x, y),$$
 (3)

the spectrum of the superposition is no longer the spectrum-convolution given by Equation (2), but rather the *sum* of the individual spectra:

$$R(u, v) = R_1(u, v) + \dots + R_m(u, v).$$
(4)

To illustrate the above discussion, consider the multiplicative superpositions shown in Figures 1 and 2. In the remainder of this section we explain these two cases, that will serve us later, in the following sections, as a basis for comparison.

In the simplest case, shown in Figure 1, both layers are pure cosinusoidal gratings (or rather 'raised' cosinusoidal gratings, since reflectance functions always take values between 0 and 1). The first layer is given here by the reflectance function:²

$$r_1(x, y) = \frac{1}{2}\cos(2\pi f_1 x) + \frac{1}{2}.$$
 (5)

This periodic function (Figure 1(*a*)) has a frequency of f_1 cycles per unit, i.e. its period is $T_1 = 1/f_1$ units in the *x* direction, $\theta_1 = 0$. Similarly, the reflectance function of the second layer is a cosinusoidal grating with a frequency of f_2 which is rotated by angle θ_2 (Figure 1(*b*)):

$$r_2(x, y) = \frac{1}{2}\cos(2\pi f_2[x\cos\theta_2 + y\sin\theta_2]) + \frac{1}{2}.$$
 (6)

In this simple case the 2D Fourier transform of each of the reflectance functions consists of exactly three impulses (see Figures 1(*d*) and (*e*)): two symmetric impulses of amplitude $\frac{1}{4}$ located at a distance of $f_i = 1/T_i$ to both sides of the origin along

(b)(c)()(d)(e) (f)

Figure 1. Raised cosinusoidal gratings (a) and (b) and their multiplicative superposition (c) in the image domain; their respective spectra are the impulse triplets shown in (d) and (e) and their convolution (f). These spectra were obtained by applying FFT to each of the first-row images. Note the two new impulse pairs which have appeared in the spectrum convolution (f): one new impulse pair is located inside the visibility circle, at the difference frequencies $f_1 - f_2$ and $f_2 - f_1$, and the other new impulse pair is located far beyond the visibility circle, at the sum frequencies $f_1 + f_2$ and $-f_1 - f_2$. The moiré effect that is visible in the layer superposition (c) corresponds, indeed, to the new low frequency that is represented by the new impulse pair at the difference frequencies. The circle in the center of each spectrum represents the visibility circle for a viewing distance where the original gratings are no longer visible.

the direction θ_i , representing the first (and only) harmonic component of the structure, plus the Fourier transform of the additional constant $\frac{1}{2}$, which is an impulse of amplitude $\frac{1}{2}$ at the origin (the DC impulse). Therefore, according to Equations (1) and (2) the spectrum of the superposition of $r_1(x, y)$ and $r_2(x, y)$ is the Fourier transform of their product, i.e. the convolution of their individual spectra $R_1(u, v)$ and $R_2(u, v)$. Since in our case each of these spectra consists of three impulses, their convolution consists of nine impulses (see Figure 1(f)); as explained in detail in Section 2.3 of [1], these impulses are obtained by placing a centered, parallel copy of $R_2(u, v)$ on top of each impulse of $R_1(u, v)$, or vice versa. This spectrum convolution contains all the impulses of the original spectra (only their amplitudes have been modified, but not their geometric locations). But in addition to the original impulses, two new impulse pairs which did not exist in any of the original spectra also appear in the convolution (see Figure 1(f)). The geometric locations of these new impulse pairs are the vectorial sum and the vectorial difference of the frequency vectors of the original impulses, namely: $f_1 + f_2$, $-f_1 - f_2$, and $f_1 - f_2$, $f_2 - f_1$.

Since each impulse pair in the spectrum reflects a periodic component with the corresponding frequency and angle in the image domain, these two new impulse pairs suggest that the superposition of the two original images includes two new periodic components which did not exist in either of the original images. And indeed, two new periodic components are present in the superposed image, Figure 1(c): the more obvious one has the frequency and the direction of the difference vector, $f_1 - f_2$, which is often much smaller than the original layer frequencies f_1 and f_2 , and is therefore located inside the visibility circle;³ this frequency corresponds, indeed, to the moiré effect that we see in the superposition (Figure 1(c)). The other periodic component, with the frequency and the direction of the sum vector, $f_1 + f_2$, contributes to the superposition fine, high-frequency details, but its isolated contribution is not easily discerned by the eye in the superposition due to its high frequency, which may largely exceed the frequencies of the original layers, and is therefore located outside the visibility circle.4

And indeed, if we extract from the spectrum convolution the three impulses having the orientation



Figure 2. Binary gratings (a) and (b) and their multiplicative superposition (c) in the image domain; their respective spectra are the infinite combs shown in (d) and (e) and their convolution (f). In this case the spectra (d), (e) of the original layers contain infinitely many harmonics, in addition to the first harmonic that appears in Figure 1; therefore, the spectrum convolution (f), i.e. the spectrum of the superposition, consist of infinitely many impulses. Note that like in Figure 1, the visibility circle in the center of spectrum (f) contains the impulse pair whose frequency vectors are $f_1 - f_2$ and $f_2 - f_1$; this is the fundamental impulse pair of the moiré seen in (c), and it spans the infinite impulse comb in the spectrum convolution (f) which defines the intensity profile of the moiré. Here, too, the spectra (d)–(f) were obtained by applying FFT to each of the first-row images; notice in the spectra (e) and (f) the false impulses that are due to the aliasing (folding-over) artifacts of the FFT.

of the moiré effect, namely the impulses located at $f_1 - f_2$, **0** and $f_2 - f_1$, and apply to them an inverse Fourier transform, we obtain back in the image domain a raised cosinusoidal structure that corresponds exactly to the moiré effect that we see in Figure 1(c).

The case involving pure raised cosinusoidal gratings is indeed the simplest, since each such grating contains only one harmonic component (i.e. one impulse to each side of the origin in the spectrum). In more complex cases, where both layers consist of more harmonic components, the convolution and the resulting moiré also consist of more impulses. For example, in the case of binary gratings with a square wave profile which takes only the values 0 and 1 (see Figure 2), the spectrum of each of the layers is a comb consisting of an infinite number of harmonics in addition to the first (and only) harmonic of the purely cosinusoidal case. Let us now consider this case in more detail.

Suppose that r(x, y) is a binary (0, 1 valued) periodic square wave grating, i.e. a sequence of straight lines, with period T and opening (i.e. *white* line width)

 τ ; we may assume here that the square wave grating is symmetrically centered about the origin, so that both the original image and its spectrum are real-valued and symmetric. This square wave is expressed by:

$$r(x, y) = \sum_{n = -\infty}^{\infty} \operatorname{rect}\left(\frac{x - nT}{\tau}\right),\tag{7}$$

where:

$$\operatorname{rect}(x) = \begin{cases} 1, & |x| \le \frac{1}{2}, \\ 0, & |x| > \frac{1}{2}. \end{cases}$$

According to the Fourier theory the periodic square wave grating r(x, y) can be expressed, by means of its two-sided Fourier series expansion, as an infinite series of weighted sine and cosine functions at the fundamental frequency of 1/T and all its harmonics (see, for example, [3], p. 205):

$$r(x,y) = \sum_{n=-\infty}^{\infty} a_n \cos(2\pi nx/T) + \sum_{n=-\infty}^{\infty} b_n \sin(2\pi nx/T),$$
(8)



Figure 3. A symmetric square wave with period *T* and opening τ and its Fourier transform. The dotted line indicates the envelope of the impulse comb. The opening ratio in this case is $\tau/T = \frac{1}{5}$, and therefore every fifth impulse in the spectrum has a zero amplitude.

where the Fourier series coefficients are given by:

$$a_{0} = \frac{\tau}{T},$$

$$a_{n} = \frac{1}{n\pi} \sin\left(\frac{\pi n\tau}{T}\right) = \frac{\tau}{T} \operatorname{sinc}\left(\frac{n\tau}{T}\right),$$

$$b_{n} = 0.$$
(9)

The fact that the square wave grating r(x, y) can be expressed as a constant term a_0 plus an infinite sum of cosine functions implies that the Fourier transform of the square wave contains a DC impulse whose amplitude is a_0 , plus an infinite series of impulse pairs (the *n*th harmonics) that are located to both sides of the origin at the frequencies $\pm n/T$, and whose amplitudes are given by the coefficients a_n (Equation (9)). And indeed, the spectrum of the square wave r(x, y) is given, according to Equation (A.8) in Appendix A of [1] or [7], p. 112, by:

$$R(u,v) = \sum_{n=-\infty}^{\infty} a_n \delta(u - n/T), \qquad (10)$$

where $\delta(u)$ is the impulse symbol (see [3], Chapter 5). This is an impulse train (or a *comb*) which samples the continuous function ('envelope') $g(u) = (\tau/T)\operatorname{sinc}(\tau u)$ at the frequency u = 1/T and all its harmonics, u = n/T (see Figure 3). The amplitude of the impulses oscillates and fades out symmetrically in both directions from the center.

Remark 1: It follows from the formula of the Fourier coefficients a_n in the case of a square wave (Equation (9)) that if the opening ratio τ/T of a given square wave is rational, i.e. if it can be expressed as a ratio m/k

between two integers, then for any *n* that is a multiple of *k* the impulse amplitude a_n is zero. For instance, if $\tau/T = \frac{1}{2}$ then every even impulse in the comb has a zero amplitude, and if $\tau/T = \frac{1}{4}$ or $\tau/T = \frac{3}{4}$ then every fourth impulse in the comb is zero. We will return to this fact in Section 5 below.

Let us now return to the line grating superposition of Figure 2. The reflectance function of the first linegrating, having period T_1 in the x direction (see Figure 2(*a*)), is given by the two-sided Fourier series:

$$r_1(x, y) = \sum_{n=-\infty}^{\infty} a_n^{(1)} \cos(2\pi n x/T_1),$$
 (11)

where the Fourier coefficients $a_n^{(1)}$ are given by Equation (9), with T and τ being equal to T_1 and τ_1 , respectively. Similarly, the reflectance function of the second grating, having period T_2 and angle θ_2 (Figure 2(*b*)), is given by:

$$r_2(x, y) = \sum_{n=-\infty}^{\infty} a_n^{(2)} \cos(2\pi n [x \cos \theta_2 + y \sin \theta_2]/T_2),$$
(12)

where $a_n^{(2)}$ are again the same Fourier coefficients, but with T_2 and τ_2 .

As indicated by Equation (10) above, the Fourier transform $R_1(u, v)$ of the reflectance function $r_1(x, y)$ is a symmetric 1D impulse comb on the *u* axis (Figure 2(*d*)); the intervals between the impulses are $1/T_1$ and their amplitudes are $a_n^{(1)}$. Similarly, the Fourier transform $R_2(u, v)$ of the reflectance function $r_2(x, y)$ is a symmetric 1D impulse comb lying on a straight line through the origin of the *u*, *v* plane whose orientation is given by θ_2 ; its impulse intervals equal $1/T_2$ and its impulse amplitudes are $a_n^{(2)}$ (Figure 2(*e*)).

Let us now consider the superposition of the two line-gratings, $r(x, y) = r_1(x, y)r_2(x, y)$. Its spectrum R(u, v) is, according to the convolution theorem (Equation (2)), the convolution of spectra $R_1(u, v)$ and $R_2(u, v)$. This convolution is an infinite oblique nailbed on the *u*, *v* plane, which is obtained by placing a centered, parallel copy of the comb $R_2(u, v)$ on top of each impulse of the comb $R_1(u, v)$, or vice versa (see Figure 2(f), and compare with the case of cosinusoidal gratings shown in Figure 1(f)). More precisely, the results of this convolution can be described as follows. (a) The impulse location (frequency-vector) of the (m, n)th impulse of the convolution in the u, v plane is the vectorial sum of the frequency-vector of the *m*th impulse in the first comb and the frequency-vector of the *n*th impulse in the second comb. (b) The *amplitude* of the (m, n)th impulse is the product of the amplitudes

of the *m*th impulse in the first comb and the *n*th impulse in the second comb:

$$a_{m,n} = a_m^{(1)} a_n^{(2)}.$$
 (13)

In our example of symmetric square-wave gratings this gives, according to Equation (9):

$$a_{m,n} = \frac{1}{mn\pi^2} \sin\left(\frac{\pi m\tau_1}{T_1}\right) \sin\left(\frac{\pi n\tau_2}{T_2}\right)$$

$$= \frac{\tau_1 \tau_2}{T_1 T_2} \operatorname{sinc}\left(\frac{m\tau_1}{T_1}\right) \operatorname{sinc}\left(\frac{n\tau_2}{T_2}\right).$$
(14)

As in the cosinusoidal case, we see that here, too, the superposition of gratings introduces new impulses in the spectrum. If any of these impulses fall inside the visibility circle, as the impulses located at $f_1 - f_2$, $f_2 - f_1$ in Figure 2(f), this indicates that in the superposed image there exists a visible periodic component (i.e. a moiré effect) at the corresponding direction and frequency. And indeed, if we extract from the spectrum convolution the impulse comb consisting of the new, low-frequency impulses at $f_1 - f_2$, $f_2 - f_1$ and all their harmonics and apply to this comb an inverse Fourier transform, we obtain back in the image domain a periodic structure that corresponds exactly to the moiré effect that we see in Figure 2(c).

Having reviewed the required background we can now proceed to the 'problematic' Fourier-based cases, our main subject of interest in the present study.

3. Pseudo (1, -1)-moiré bands in additive superpositions

Let us consider again the simplest case in which both layers are pure, raised cosinusoidal gratings (see Equations (5) and (6) above). But this time, unlike in Figure 1, we superpose these layers using the additive superposition rule, as explained above (note that in order not to exceed the value 1 in the resulting reflectance function, we can normalize it back into the range 0...1 by taking the *average* of the individual layers rather than their sum). This additive superposition is shown in Figure 4. As we have seen in Equation (4), in this case the spectrum of the superposition is given by the sum (or, respectively, by the average) of the spectra of the original layers, and not by their convolution. It follows therefore that unlike in the multiplicative case (Figure 1(f)), in the present case the spectrum of the superposition (Figure 4(f)) does not contain any new impulses at the sum or difference frequencies, but only the original impulses of the two original layers.



Figure 4. Same as in Figure 1, but this time using an additive superposition rule. In this case the spectrum of the superposition (f) containd only the impulses of the two original spectra (d) and (e), and no new impulse pairs are generated at the sum or difference frequencies. And yet, a new low-frequency structure is visible in the superposition (c) exactly at the same locations as in the multiplicative case shown in Figure 1(c).

Journal of Modern Optics

(c)(d)(e) (f)

Figure 5. The multiplicative superposition of two raised cosinusoidal gratings having the same orientation but slightly different frequencies, and the respective spectra.

A similar situation is also shown in Figures 5 and 6. where both of the superposed layers have the same orientation but slightly different frequencies. Here, too, the spectrum of the additive superposition (Figure 6(f)) does not contain any new impulses at the sum or difference frequencies, unlike in the multiplicative case (Figure 5(f)).

And yet, when visually comparing the multiplicative superposition and the additive superposition, we see that in the additive case, too, there still appears a new low-frequency structure, exactly at the same locations as in the multiplicative case: note the washedout gray bands which appear in Figure 4(c) (or in Figure 6(c) exactly where the dark (1, -1)-moiré bands appear in Figure 1(c) (or, respectively, Figure 5(c)). These washed-out gray bands are not a real (1, -1)-moiré, since unlike in the multiplicative case the superposition does not contain a lowfrequency component at the difference frequency (we have no 'energy', i.e. no impulses, at that frequency in the spectrum).

Theoretically, if we hang Figures 1(c) and 4(c) (or Figures 5(c) and 6(c) side by side on a wall and observe them from a sufficient distance, so that the original gratings are no longer discerned by the eye, we should see a (1, -1)-moiré in the multiplicative superposition, but not in the additive superposition. However, in practice we do see a similar structure in the additive case, too, even though it looks much

weaker than the real (1, -1)-moiré bands in the multiplicative superposition. We may call this phantom structure a *pseudo* (1, -1)-moiré, since according to the Fourier theory it should not exist. How can we explain the existence of this phantom structure, in spite of the theoretical predictions? We survey below some plausible classical answers, along with a new possible explanation (that will be further developed in Section 6). The truth is probably a combination of some or all of these points.

(1) The effect of non-linearities. Remember that printers (and many other display devices) normally have a significantly non-linear response (or reproduction curve).⁵ This may explain why weak phantom traces of a (1, -1)moiré are visible in the additive superposition, too: indeed, according to a well-known result (that we will henceforth call the non-linearity *principle*), when a non-linearity is applied to a sum of cosines with frequencies f_1 and f_2 (for instance by displaying the sum using a printer or any other device having a non-linear behaviour), new frequencies of the form $k_1f_1 + k_2f_2$ (with $k_1, k_2 \in \mathbb{Z}$) may be generated in the spectrum. These frequencies correspond to new moiré effects that may become visible in the superposition if they fall inside the visibility circle (see Remark 2.3 in [1], as well as [8],



Figure 6. The additive superposition of two raised cosinusoidal gratings having the same orientation but slightly different frequencies, and the respective spectra.

pp 91–92, [4], p. 88 and in particular [9]). In our additive case (Figure 4) the new difference frequency $f_1 - f_2$ that is generated by the nonlinearity falls, indeed, inside the visibility circle (it is 'added' to Figure 4(f) exactly in the same location as in Figure 1(f)), and hence it results in a visible (1, -1)-moiré in the additive superposition, although this moiré is much weaker than in the multiplicative case.

(2) Microstructure versus macrostructure effects. Secondly, it can be argued that the pseudomoiré bands in the additive superposition only correspond to microstructure variations in the superposition, but unlike in the multiplicative superposition, they do not indicate the presence of macrostructure variations (a real low frequency component):⁶ in spite of the repetitive variations in the microstructure, the local average value⁷ of the additive superposition remains constant throughout, while in the multiplicative superposition the local average value really varies, indicating the existence of darker and brighter moiré bands (this is easier to see in the 1D counterpart shown in Figure 7). Note in particular that in the additive case the dark bands are, in fact, gray (the microstructure is oscillating there around the value $\frac{1}{2}$, just like anywhere else in the superposition), while in the multiplicative case

the dark bands are really black (the microstructure is oscillating there around the value 0, much lower than in the bright bands). This explains why in the additive case these bands are so weak, and why, at least theoretically, they are only visible as long as the observer can still resolve the individual lines of the original gratings – while in the multiplicative case the moiré bands are clearly visible even from a longer distance (see, for example, [8], pp 89–92). The use of more localized versions of the Fourier transform such as a suitable wavelet transform may better take into account local microstructure variations than the global Fourier transform of the entire image.

(3) The human visual system. While in the multiplicative case the large bands in question are indeed represented in the spectrum by a low-frequency impulse pair (at the difference frequency), in the additive case no such impulse pair exists in the spectrum, and it is only the human eye-brain system that connects adjacent microstructure details into moiré-like bands that are 'seen' in the additive case (see, for example, [10,11]). This means that the human visual system is, indeed, more sophisticated than just a simple Fourier analyzer. Alternatively, it could be also argued that human vision has a non-linear nature, so that





Figure 7. (a) The 1D counterpart of the multiplicative superposition of two raised cosinusoidal waves with similar frequencies $f_1 \approx f_2$, namely, $[\frac{1}{2}\cos(2\pi f_1x) + \frac{1}{2}][\frac{1}{2}\cos(2\pi f_2x) + \frac{1}{2}]$. (b) The 1D counterpart of the additive superposition of two raised cosinusoidal waves with similar frequencies $f_1 \approx f_2$, namely, $(\frac{1}{2}\cos(2\pi f_1x) + \frac{1}{2}] + [\frac{1}{2}\cos(2\pi f_2x) + \frac{1}{2}])/2$. Note that in order not to exceed the value 1 in the resulting reflectance function we normalize it back into the range 0...1 by taking the average of the individual waves rather than their sum. (c) The envelope of the modulated wave (b) is given by the raised cosine $\frac{1}{2}\cos(2\pi [(f_1 - f_2)/2]x) + \frac{1}{2})$, but its period is half of the period of this cosine. The frequency of the envelope is therefore twice $(f_1 - f_2)/2$, namely $f_1 - f_2$. Note the real (1, -1)-moiré effect at the difference frequency $f_1 - f_2$ in (a), and the pseudo (1, -1)-moiré effect at the very same frequency in (b). Compare with the 2D Figures 5 and 6 (which are drawn, however, for other frequencies). For the sake of completeness, we show in (d) the two original cosines being used in the present figure: the raised cosine $\frac{1}{2}\cos(2\pi f_1x) + \frac{1}{2}$ with $f_1 = 1$ is drawn in black (left), and the raised cosine $\frac{1}{2}\cos(2\pi f_2x) + \frac{1}{2}$ with $f_2 = 1.1$ is dashed (right); both curves are overprinted in the central part of (d) to allow a better understanding of their product and of their average in (a) and (b).

≯ the missing low-frequency impulses are introduced, in the same manner as in point (1) above, by the human visual system itself.

(4) Modulation. A further, related point is that the image obtained in the additive superposition is in fact a 2D amplitude modulation of a cosinusoidal wave having the average frequency (which plays the role of the *carrier*) by a cosinusoidal wave having the difference frequency (which plays the role of the modulator or the envelope). And indeed, it turns out that the pseudo-moiré bands in the additive superposition, which are the envelope of the 2D amplitude modulation (see Figure 4(c) or Figure 6(c), have the same period and angle as the real (1, -1)-moiré bands in the multiplicative superposition (Figure 1(c) or, respectively, Figure 5(c)). Let us show now that this is always true, and not just a happy coincidence that occurs in the particular case of Figures 1 and 4 or Figures 5 and 6.

For the sake of simplicity, we show first the 1D counterpart of this claim, namely, that the pseudomoiré in the additive superposition of Figure 7(b) has the same period as the real moiré in the multiplicative superposition of Figure 7(a). Based on the well-known trigonometric identity

$$\cos \alpha + \cos \beta = 2\cos \frac{\alpha + \beta}{2}\cos \frac{\alpha - \beta}{2},$$

we obtain in the additive case:

$$\left(\left[\frac{1}{2}\cos(2\pi f_1 x) + \frac{1}{2}\right] + \left[\frac{1}{2}\cos(2\pi f_2 x) + \frac{1}{2}\right]\right)/2$$

= $\frac{1}{4}\left[\cos(2\pi f_1 x) + \cos(2\pi f_2 x)\right] + \frac{1}{2}$ (15)
= $\frac{1}{2}\cos\left(2\pi \frac{f_1 + f_2}{2}x\right)\cos\left(2\pi \frac{f_1 - f_2}{2}x\right) + \frac{1}{2}.$

Thus, the sum (or average) of the two cosinusoidal layers gives a product of cosines. Clearly, if the original layer frequencies f_1 and f_2 are similar, this cosine product corresponds to a modulation effect, where the cosine with the higher frequency $(f_1+f_2)/2$ represents the carrier and the cosine with the low frequency $(f_1-f_2)/2$ represents the envelope (see Figure 7(*b*)).⁸ However, because the envelope periodicity is in fact half of one full cosine period (see Figure 7(*c*)) it turns out that the envelope's frequency is equal to the frequency difference f_1-f_2 .

Let us now return to the multiplicative superposition of Figure 7(a). Using the well-known trigonometric identity

$$\cos\alpha\cos\beta = \frac{1}{2}\cos(\alpha + \beta) + \frac{1}{2}\cos(\alpha - \beta),$$

we obtain:

$$\begin{bmatrix} \frac{1}{2}\cos(2\pi f_1 x) + \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2}\cos(2\pi f_2 x) + \frac{1}{2} \end{bmatrix}$$

= $\frac{1}{4} + \frac{1}{4}\cos(2\pi f_1 x) + \frac{1}{4}\cos(2\pi f_2 x)$
+ $\frac{1}{8}\cos(2\pi [f_1 + f_2]x)$
+ $\frac{1}{8}\cos(2\pi [f_1 - f_2]x),$ (16)

where the last cosine corresponds to the visible moiré effect, whose frequency is, again, the frequency difference $f_1 - f_2$.

We see therefore that both in the additive and in the multiplicative cases the visible low-frequency phenomenon has the same frequency, $f_1 - f_2$.

The situation in 2D modulation is similar to the 1D case of Figure 7(*b*), but a little more complicated since the carrier and the envelope waves in the 2D case may have different orientations (as we can see, for example, in Figure 4(c)). To simplify the presentation in the 2D case, we adopt the following vector notation:

$$\cos(2\pi f_i[x\cos\theta_i + y\sin\theta_i]) = \cos(2\pi f_i \cdot x),$$

where $f_i = f_i(\cos \theta_i, \sin \theta_i)$ and $\mathbf{x} = (x, y)$. Using this vector notation the generalization of Equation (15) to the 2D additive case of Figure 4 is:

$$\left(\left[\frac{1}{2}\cos(2\pi f_{1}\cdot \mathbf{x})+\frac{1}{2}\right]+\left[\frac{1}{2}\cos(2\pi f_{2}\cdot \mathbf{x})+\frac{1}{2}\right]\right)/2$$

= $\frac{1}{4}\left[\cos(2\pi f_{1}\cdot \mathbf{x})+\cos(2\pi f_{2}\cdot \mathbf{x})\right]+\frac{1}{2}$
= $\frac{1}{2}\cos\left(2\pi \frac{f_{1}+f_{2}}{2}\cdot \mathbf{x}\right)\cos\left(2\pi \frac{f_{1}-f_{2}}{2}\cdot \mathbf{x}\right)+\frac{1}{2}.$ (17)

Thus, here too, the sum (or average) of the two cosinusoidal layers gives a product of cosines. If the original layer frequencies f_1 and f_2 are similar, this cosine product corresponds to a 2D modulation effect, where the cosine with the higher frequency $(f_1+f_2)/2$ represents the carrier and the cosine with the low frequency $(f_1 - f_2)/2$ represents the envelope (see Figures 4(c) or 6(c)). However, because the envelope periodicity is half of one full cosine period it turns out that the envelope's frequency is equal to the frequency difference $f_1 - f_2$.

Similarly, the generalization of Equation (16) to the 2D multiplicative case of Figure 1 is:

$$\begin{aligned} \left[\frac{1}{2}\cos(2\pi f_{1}\cdot \mathbf{x})+\frac{1}{2}\right]\left[\frac{1}{2}\cos(2\pi f_{2}\cdot \mathbf{x})+\frac{1}{2}\right] \\ &=\frac{1}{4}+\frac{1}{4}\cos(2\pi f_{1}\cdot \mathbf{x})+\frac{1}{4}\cos(2\pi f_{2}\cdot \mathbf{x}) \\ &+\frac{1}{8}\cos(2\pi [f_{1}+f_{2}]\cdot \mathbf{x})+\frac{1}{8}\cos(2\pi [f_{1}-f_{2}]\cdot \mathbf{x}). \end{aligned}$$
(18)

We see therefore that both in the additive and in the multiplicative 2D cases the visible low-frequency phenomenon has the same frequency vector, $f_1 - f_2$ (and hence, the same period and the same orientation), as

shown, indeed, in Figures 4(c) and 1(c) or in Figures 6(c) and 5(c). But while in the multiplicative case the low frequency $f_1 - f_2$ that we see in the superposition corresponds, indeed, to an inpulse that is located in the spectrum at the point $f_1 - f_2$, in the additive case the modulation envelope with the low frequency $f_1 - f_2$ that we see in the superposition has no corresponding impulse at the point $f_1 - f_2$ in the spectrum. (In fact, as we will see in Section 6, this impulse does exist, but its amplitude is simply zero.)

4. Higher order pseudo moiré bands in the superposition of cosinusoidal gratings

Pseudo moiré bands which 'mimic' the behavior of higher order moirés may occur in various circumstances, even in multiplicative superpositions. The examples discussed in this and the following sections are derived from the situation shown in Figure 8 (see also Figure 2.6 in [1]) – the multiplicative superposition of two binary square-wave gratings, where the low-frequency bands that are visible in the superposition belong to the second-order (1, -2)-moiré.⁹

Consider, first, the multiplicative superposition of two raised cosinusoidal gratings having the same

angles and periods as in Figure 8. This is shown in Figure 9. The diagonal bands that we see in this superposition (Figure 9(c)) correspond to the (1, -2)-moiré, i.e. the moiré that is generated by the impulse pair $f_1 - 2f_2$, $2f_2 - f_1$ in the spectrum convolution (like in Figure 8(f)). However, as shown in Figure 9(f), in a superposition of raised cosinusoidal gratings there exist no second harmonic impulses (their amplitudes are necessarily zero). How can we explain this apparent contradiction?

Here, too, it can be said that the bands we see are partially due to the non-linearity of the printing device. Secondly, a close examination of the superposition shows that the bands in question only represent a repetitive variation in the microstructure, but not a variation in the macrostructure (since the local average remains constant throughout); this shows, once again, that the human visual system performs better (is able to capture more) than just a Fourier analysis stage. Finally, it can be also argued that what we see in Figure 9 is not a second-order moiré effect, but rather a second-order modulation effect.

Note that a similar phenomenon may occur also in an additive superposition of such cosinusoidal gratings, although it is much weaker and less visible.



Figure 8. Binary gratings (a) and (b) as in Figure 2 but with (b) having half the frequency, and their multiplicative superposition (c); their respective spectra are (d), (e) and (f). The visibility circle in the center of the spectrum (f) contains the impulse pair with frequency vectors $f_1 - 2f_2$ and $2f_2 - f_1$, which originate from the second harmonic of f_2 , and represent the fundamental impulse pair of the moiré that we see in (c). Note that although this moiré is a (1, -2)-moiré, it still has the same angle and frequency as the (1, -1)-moiré of Figure 2, and only its intensity is weaker; this is easily understood by comparing the moiré impulse locations in Figures 2(f) and 8(f).



Figure 9. Raised cosinusoidal gratings (a) and (b) as in Figure 1 but with (b) having half the frequency, and their multiplicative superposition (c); their respective spectra are (d), (e) and (f). Although when looking at the superposition (c) we have the impression it contains a second-order moiré effect (having the same angle and frequency as the (1, -1)-moiré of Figure 1), the second-harmonic impulses with frequency vectors $f_1 - 2f_2$ and $2f_2 - f_1$, which would generate such a moiré effect in (c), do not exist in the spectrum convolution (f). Compare with Figure 8, where these second-harmonic impulses do exist.

5. Higher order pseudo moiré bands in the multiplicative superposition of binary gratings

Consider now the multiplicative superposition of two binary square-wave gratings having the same angles and periods as in the configuration of Figure 8, but this time both gratings having an opening ratio of $\tau/T = \frac{1}{2}$. Looking at this superposition (Figure 10(c)) we can see there diagonal bands which correspond to the (1, -2)moiré in Figure 8(c). This is shown even more clearly in Figures 11(a) and (b). As we have seen in Figure 8(f), this moiré is generated by the impulse pair $f_1 - 2f_2$, $2f_2 - f_1$ and the entire comb that it spans in the spectrum convolution. However, according to Equation (9), which gives the Fourier coefficients a_n in the case of a square wave, it follows that when the opening ratio of a square wave is $\tau/T = \frac{1}{2}$, every even impulse in the comb has a zero amplitude (see Remark 1 above). As a consequence, it follows from Equation (13) that the impulses located at $f_1 - 2f_2$, $2f_2 - f_1$ in the spectrum convolution, as well as all the other impulses on the comb of the (1, -2)-moiré, have zero amplitudes (compare Figure 10(f) with Figure 8(f)). How can we explain this apparent contradiction?

In this case, too, we may consider the same possible explanations as before – except for the non-linearity of

halftoning, since in this case all the images are purely black and white and no halftoning is used for displaying them. This does not yet completely exclude the printer's responsibility, since some of its other properties such as its dot-gain may cause the black lines of the gratings to be somewhat larger than theoretically expected (due to phenomena such as ink spreading or toner particle satellites on the paper), so that the effective (1, -2)-moiré impulses in the spectrum are not really zero.

However, in the present case, the most plausible explanation seems to be as follows: consider the dotted lines A and B in Figure 11(a). Both of these lines are drawn in parallel to the moiré bands, but while line A runs through the darkest part of a dark moiré band, line B runs through the brightest part of a bright moiré band. And indeed, we can easily see that along the line A there is more black than along the line B. This confirms the existence of the dark and bright (1, -2)-moiré bands in Figure 11(a).

Consider now Figure 11(*b*). The only difference between this figure and Figure 11(*a*) is in the line widths: in Figure 11(*b*) both gratings have an opening ratio of $\tau/T = \frac{1}{2}$, so that theoretically no (1, -2)-moiré bands should exist in their superposition. And indeed, if we concentrate on lines A and B in Figure 11(*b*), we



Figure 10. Same as in Figure 8, but this time both of the superposed binary gratings have the opening ratio $\tau/T = \frac{1}{2}$. This implies that all the even impulses in the spectra (d) and (e) have a zero amplitude (compare with the spectra (d) and (e) in Figure 8). Therefore, all the impulses contained in the second-order moiré comb that is spanned in the spectrum convolution (f) by the second-harmonic impulses with frequency vectors $f_1 - 2f_2$ and $2f_2 - f_1$ have a zero amplitude (except for the DC impulse). This means that the (1, -2)-moiré of Figure 8 does not exist in the present case; and yet, we have the impression that such moiré bands exist in (c) exactly in the same locations as in Figure 8(c).

see that unlike in Figure 11(a) the amount of black ink along line A is identical to the amount of black ink along line B; in fact, the amount of black ink along any line parallel to lines A and B in this figure is identical. This indicates, indeed, that unlike in Figure 11(a), the banding phenomenon we see in Figure 11(b) does not originate from a real physical difference in the black ink coverage rate in dark and white bands, but only from the different spatial distribution of the same quantity of black ink. In other words, these pseudomoiré bands correspond to a periodic variation in the local microstructure, but not to a real, physical periodicity in the global average gray level (the macrostructure). This is why no moiré impulses are visible in the Fourier spectrum of the superposition (Figure 10(f)), while they are visible in Figure 8(f). And indeed, if we hang Figures 11(a) and (b) on a distant wall and observe them from a sufficient distance, i.e. a distance from which the eye no longer resolves the microstructure, then (at least in principle) no moiré bands should be visible in Figure 11(b). And if they are visible, this can be attributed, as explained above, to factors extrinsic to the Fourier-based model such as non-linearities or a more sophisticated behaviour of the human visual system.

6. A new explanation based on the role of the module $Md(f_1, \dots, f_m)$

As we can see, not every periodicity in the image superposition is represented by a corresponding impulse pair (or impulse comb) in the spectrum. And indeed, such periodicities having no 'energy' in the spectrum (i.e. whose impulse amplitudes are zero) may give rise to pseudo-moiré effects, that can in some circumstances mimic the behaviour of their real moiré counterparts, as we have seen in the examples above. In the case discussed in Section 3 this pseudo-moiré effect is simply due to an amplitude modulation having the frequency $f_1 - f_2$, but in the cases discussed in Sections 4 and 5 the pseudo-moiré effect is due to a 'microstructure modulation', i.e. periodic variations in the microstructure having the frequency $f_1 - f_2$.

It should be noted that such phenomena are not limited to superpositions of two layers, and they may occur also in superpositions of three or more layers; in such cases there may even exist hybrid phenomena that are generated by a combination of real and pseudo moirés. From this point of view it can be said that not only non-zero amplitude impulses in the spectrum contribute visible low



Figure 11. (a) A multiplicative superposition of two binary gratings having periods T and 2T (i.e. frequencies f and f/2) gives a (1, -2)-moiré as explained in Figure 8. The dotted lines A and B indicate the center of the dark and bright bands of the moiré, respectively. (b) Same as in (a), but this time both of the gratings have an opening ratio of $\tau/T = \frac{1}{2}$, like in Figure 10. The same dotted lines A and B pass, again, along the center of the visible bands, but this time the ratio of black to white along the two lines is identical.

frequencies in the image domain. In fact, if f_1, \ldots, f_m are the frequency vectors of the original layers, then *all* the impulse locations in the spectrum that belong to the set $Md(f_1, \ldots, f_m) = \{k_1f_1 + \cdots + k_mf_m | k_i \in \mathbb{Z}\}$ play a role in the superposition,¹⁰ and if any of them is located within the visibility circle it may correspond to a visible low frequency in the layer superposition: if its amplitude is zero it corresponds to a pseudo-moiré, otherwise it corresponds to a real moiré effect. This is, indeed, a two-fold generalization of the well-known non-linearity principle that we mentioned in Point (1) of Section 3: on the one hand we generalize it here to m rather than 2 frequency vectors (or superposed layers), but more importantly, we also generalize it here to impulse locations having zero amplitudes (i.e. no energy). In other words, even impulses with zero amplitudes may contribute something to the visible phenomena in the superposition.

According to the non-linearity principle, it is the application of a non-linearity that makes visible the new impulses of the module $Md(f_1, f_2) = \{k_1f_1 + k_2f_2 | k_i \in \mathbb{Z}\}$ by giving them a nonzero amplitude. But our result here is even stronger, since it says that any impulse of this module (that falls inside the visibility circle) contributes to the visible result, even if its amplitude is zero. But of course, if a non-linearity is applied and hence endows such an impulse with a non-zero amplitude (by virtue of the the non-linearity principle), this will simply further strengthen the visible effect.

Note also that we must always consider the entire infinite module $Md(f_1, \ldots, f_m)$. This is true not only in the case of periodic structures whose spectra consist of an infinite comb of impulses (some of which may happen to have a zero amplitude, like in Figure 10), but also in the case of periodic structures such as pure cosines who only have a single pair of non-zero harmonic impulses (like in Figures 4 and 9). It should be understood that even in this last case all the higher harmonic impulses do exist (although their amplitude is zero), and indeed they may reappear as soon as we apply even the slightest deformation on the pure cosinusoidal profile of the structure. This is, indeed, an extension of the classical Fourier approach where considered the non-zero impulses one only that appear in the spectrum of the layer superposition. The set of non-zero impulses that is considered in the classical Fourier approach is clearly a subset of the module that is spanned by the frequencies f_1, \ldots, f_m of the original periodic layers, $Md(f_1, \ldots, f_m)$.

7. Conclusion

The Fourier theory by itself cannot answer all questions related to moiré effects. We show several cases in which periodic structures that are generated in the superposition of periodic layers do not have corresponding impulses in the Fourier spectrum. We survey several plausible classical explanations, but also provide a new possible approach for the understanding of such phenomena. We propose to extend the scope of the Fourier theory by considering in the spectrum of the superposition all the impulses of the module $Md(f_1, \ldots, f_m)$, including impulses having zero amplitudes. This approach allows us to understand limit cases where low-frequency impulses that are located inside the visibility circle happen to have zero amplitudes. Although such 'impulses' do not correspond to an 'energy' at that frequency in the spectrum, they still contribute a visible structure in the image domain that corresponds to a microstructure modulation effect (which reduces, in the simplest cases, into amplitude modulation). This provides a natural

explanation of various phenomena (as illustrated in Sections 3–5) which could not be explained so far by the classical Fourier approach alone, without having to use extrinsic arguments such as the effect of non-linearities.

Notes

- 1. Since the Fourier transform is a global operation that is applied to the entire spatial image domain, local microstructure effects are averaged together and buried in the global spectrum of the entire image. And even if we apply the Fourier transform (or a localized version thereof such as a wavelet transform) to different local areas of the entire image, it will only help us to distinguish between the different local microstructures and to identify and analyze their particular spectral properties; but this will not yet explain the various geometric shapes of the microstructure elements (rosettes, etc.).
- 2. For the sake of simplicity we prefer to use here *cosinusoidal* gratings rather than *sinusoidal* gratings, since the cosine function is symmetrically centered about the origin, so that its spectrum is real valued. If we use here sines (or cosines having any other phases) our results will remain unchanged, except that the spectrum impulses will have complex rather than real valued amplitudes.
- 3. The visibility circle is a theoretic circle around the spectrum origin which approximately represents the cut-off frequency of the human visual system (i.e. the highest frequency it can resolve) under some given viewing conditions, such as viewing distance, light conditions, etc.
- 4. Note that for the sake of clarity all the enclosed figures show largely magnified gratings; however, the visibility circle drawn in their spectra corresponds to a more realistic situation where the frequencies of the original layers are no longer visible. With the present magnified gratings this may imply a viewing distance of a few meters.
- 5. Note that gray levels are generated in many display and printing devices using a halftoning technique, which is a non-linear operation.
- 6. Note that the term microstructure refers to structure elements whose size is of about the same order as the periods of the original layers, while the term macrostructure refers to structures that are much larger than the periods of the original layers.
- 7. i.e. the average value calculated within a moving local window.
- 8. If the original layer frequencies are very different, say $f_1 \ll f_2$, the cosine sum no longer resembles such a modulation, but rather a low frequency cosine wave of frequency f_1 having a high-frequency perturbation of frequency f_2 (or in other words, a high-frequency cosine wave of frequency f_2 riding on a low-frequency cosine wave of frequency f_1).
- 9. Note that the (1, -2)-moiré bands in Figure 8 appear exactly in the same locations as the (1, -1)-moiré bands in Figure 2. The reason is that in Figure 8 the period of the second layer has been doubled with respect to Figure 2 (meaning that the impulse distance in its comb has been halved), while the angles in both cases remain

identical (compare the impulse locations in the bottom row of Figures 2 and 8).

10. Note that this infinite set of points in the spectrum of the superposition, which is spanned by the vector frequencies f_1, \ldots, f_m of all the individual layers, is not a dot lattice but rather a *module*. For a more detailed explanation see Section 5.2 in [1]. This set is the support of the spectrum convolution, and it contains the geometric locations of *all* the possible harmonic impulses, including those whose amplitude is zero.

References

- [1] Amidror, I. *The Theory of the Moiré Phenomenon*; Kluwer Academic Publishers: Dordrecht, 2000.
- [2] Amidror, I. The Theory of the Moiré Phenomenon, Volume II: Aperiodic Layers; Springer: Dordrecht, 2007.

- [3] Bracewell, R.N. *The Fourier Transform and its Applications*, 2nd ed.; McGraw-Hill Publishing Company: Reading, NY, 1986.
- [4] Bryngdahl, O. J. Opt. Soc. Am. 1976, 66, 87-94.
- [5] Asundi, A.K. Opt. Eng. 1993, 32, 107-116.
- [6] Patorski, K. Handbook of the Moiré Fringe Technique; Elsevier: Amsterdam, 1993.
- [7] Gaskill, J.D. Linear Systems, Fourier Transforms, and Optics; Wiley: New York, 1978.
- [8] Post, D.; Han, B.; Ifju, P. High Sensitivity Moiré: Experimental Analysis for Mechanics and Materials; Springer-Verlag: New York, 1994.
- [9] Eschbach, R. J. Opt. Soc. Am. A 1988, 5, 1828–1835.
- [10] Badcock, D.R.; Derrington, A.M. Vision Res. 1985, 25, 1253–1258.
- [11] Badcock, D.R.; Derrington, A.M. Vision Res. 1989, 29, 731–739.