Supplementary Material

Amplitude and Phase Spectra Defined Scene Statistics
There exists a long history of describing the low-level properties of complex images in the spatial domain (i.e., the domain which constitutes the image itself and is made up of pixel luminance values) with respect to the statistical relationships between the pixel luminance values that comprise any given image (Julesz, Gilbert, Shepp, & Frisch, 1973; Julesz, Gilbert, & Victor, 1978; Klein & Tyler, 1986; Thomson, 1999a; Victor, Thengone, & Conte, 2013). Specifically, the pixel luminance histogram (i.e., the probability distribution of luminance) is referred to as the first moment (or first-order statistic) of the image. The degree of correlation between pairs of pixels as a function of all possible physical distances is referred to as the 2nd-order statistic, and can be conceivied of as the variance of the pixel luminance distribution (i.e., the luminance contrast of the image). The relationship between more than two pixel luminance values (e.g., pixel triplets or quadruplets) is referred to as the higher-order image statistics of an image, and have been shown to carry information about the relative edges, lines, contours, etc., that form the structures that provide the meaningful content of scenes (Thomson, 1997; 1999a; 1999b; 2001a; 2001b; Victor & Conte, 1996; Zetsche, Krieger, & Wegmann, 1999). One historically important approach to quantifying the image statistics of scene images comes from the global 2-dimensional (2-D) Discrete Fourier Transform (DFT) and related methods developed from it. The 2-D DFT treats an image as a complex 2-D luminance waveform, which can be represented as the sum of sinusoidal waveforms of different amplitudes, frequencies, orientations, and phases (referenced in the Fourier domain, which is the linear transformation of the image in the spatial domain).

The global Fourier amplitude spectrum (or power spectrum – the amplitude spectrum squared) of a given scene image contains only relative measurements pertaining to image contrast as a function of spatial frequency and orientation in an image, without regard to their locations in the spatial domain. The most consistently observed statistical regularity in the amplitude spectra of scene imagery involves a physical property where the contrast at different spatial frequencies (f) falls with increasing spatial frequency, following a 1/fα relationship (e.g., Billock, 2000; Field, 1987; Ruderman & Bialek, 1994; Tolhurst et al., 1992; van der Schaaf & van Hateren, 1996), with α being the slope on logarithmic axes, and typically being observed to be near 1.0. The exact α measured for a given scene image is therefore typically referred to as the slope of the orientation averaged amplitude spectrum of that image. What this amounts to is that scene imagery generally contains much more contrast at lower spatial frequencies relative to higher spatial frequencies, with the slope of the amplitude spectrum reflecting the degree of luminance contrast at lower relative to higher spatial frequencies. Additionally, regarding the orientation of scene content, scenes typically have a bias for the vertical (0°) and horizontal (90°) orientations (i.e., the cardinal orientations) compared to the 45° and 135° oblique orientations (Hansen & Essock, 2004; Keil & Cristóbal, 2000; Switkes, Mayer, & Sloan, 1978; Torralba & Oliva, 2003; van der Schaaf & van Hateren, 1996). Since the distribution of amplitude as a function of spatial frequency and orientation in the Fourier domain is a direct assessment of the degree of correlation between the luminance values of all possible pixel pairs in the spatial (i.e., image) domain (e.g., Klein & Tyler, 1986), the Fourier amplitude spectrum (or power spectrum) is a direct measurement of the 2nd-order statistical relationships of the pixel luminance values of scene images.

The phase (from 0-2π) of the waveforms obtained from the DFT plotted as a function of orientation and spatial frequency is referred to as the phase spectrum (Shapley and Lennie, 1985; Smith, 2007) and has been shown to provide useful information regarding the higher-order statistical relationships of pixel luminance values in the image domain (e.g., Gerhard, Wichmann, & Bethge, 2013; Julesz, Gilbert, Shepp, & Frisch, 1973; Julesz, Gilbert, & Victor, 1978; Klein & Tyler, 1986; Oppenheim & Lim, 1981; Piotrowski & Campbell, 1982; Thomson, 1999a; 2001a; 2001b). The phase spectrum of a given scene image determines where in a given scene different image frequencies are aligned (Kovesi,
1999; Marr, 1979; Morrone & Burr, 1988; Morrone & Owens, 1987; Thomson 1999a; 1999b; Wang & Simoncelli, 2004; Wichmann, Braun, & Gegenfurtner, 2006). Therefore, the phase spectrum is considered a global representation that defines the spatial relationships between different spatial frequencies that form specific local image features within an image (though the representation itself is not local). Accordingly, any scene image that lacks phase coherence (e.g., a phase-randomized image) would have its contrast as a function of spatial frequency and orientation randomly distributed across the image, which would produce an image with non-localized structure. Such an image would still possess amplitude-defined information (provided that there are relative differences in contrast across spatial frequency and orientation), but would lack phase-defined statistical relationships, including the edges and lines that comprise recognizable structure within scene images. See the illustration below for additional detail. Lastly, it is important to note that while one can create image structure defined only by amplitude regularities (i.e., a phase randomized scene image), the same is not true for phase-defined structure. That is, some level of contrast difference is needed in order to see phase-defined structure (e.g., Westheimer, 2001). Even an image with an intact phase spectrum that has been whitened (i.e., flat amplitude spectrum, described above) will possess local contrast differences in the vicinity of the phase-defined edges.

(a) Top left, illustration of an edge, below that is an illustration of a series of sinusoidal waveforms that have “arrival phase” convergence at the central position, which, over a full series of sinusoidal waveforms of increasingly higher spatial frequencies (not shown) will sum up to the edge shown above. The next four illustrations show how the structure of that edge is corrupted as the alignment of the sinusoidal waveforms is increasingly perturbed. (b) Two-dimensional (2-D) illustration of how the structure of the scene’s edges can be disrupted as a function of increasing (left-to-right) phase randomization. Note that for both (a) and (b), the global 2nd-order image statistics are identical for all images in the row, but in the far right image (phase randomized), the higher-order statistical relationships have been destroyed.


