Image Statistics of American Sign Language:  
Comparison to Faces and Natural Scenes

Rain G. Bosworth, Marian S. Bartlett, Karen R. Dobkins

1 Dept. of Psychology, University of California, San Diego  
2 Institute for Neural Computation, University of California, San Diego

*Corresponding Author:  
Karen R. Dobkins, Ph.D.  
Department of Psychology, 0109  
University of California, San Diego  
La Jolla, CA 92093  
USA  
Phone: (858) 534-5434  
Fax: (858) 534-7190  
Email: kdobkins@ucsd.edu

Keywords: Sign Language, ASL, Fourier analysis, Power Spectra, Spatial Frequency,  
Orientation, Entropy
Abstract

Several lines of evidence suggest that the image statistics of the environment shape visual abilities. To date, the image statistics of natural scenes and faces have been well characterized using Fourier analysis. Here, we employed Fourier analysis to characterize images of signs in American Sign Language (ASL). These images are highly relevant to signers who rely on ASL for communication, and thus the image statistics of ASL might influence signers’ visual abilities. Fourier analysis was conducted on 105 static images of signs, and these images were compared to analysis of 100 natural scene images and 100 face images. We obtained two metrics from our Fourier analysis: mean amplitude and entropy of the amplitude across the image set (which is a measure from information theory), as a function of spatial frequency and orientation. The results of our analyses revealed interesting differences in image statistics across the three different image sets, setting up the possibility that ASL experience may alter visual perception in predictable ways. In addition, for all image sets, the mean amplitude results were markedly different from the entropy results, which raises the interesting question of which aspect of an image set (mean amplitude or entropy of the amplitude) is better able to account for known visual abilities.
Several lines of evidence suggest that visual experience with the environment plays a role in shaping visual abilities. One of the best known examples is in the domain of orientation processing. Early studies have shown that raising animals in environments with only horizontal or vertical contours results in enhanced sensitivities to those orientations.\textsuperscript{1-3} (Similar results have been reported in children who grew up with uncorrected astigmatism\textsuperscript{4, 5}). Humans raised under normal (and unrestricted) conditions also show anisotropies for orientations, which are thought to arise from a natural orientation bias in the environment. Specifically, human subjects exhibit an “oblique effect” in which acuity and contrast sensitivity are better for cardinal (vertical and horizontal), than for oblique, orientations\textsuperscript{6-8} (but see\textsuperscript{9, 10}, and\textsuperscript{11-13} for evidence of an oblique effect has been reported in infants). It had long been speculated that the oblique effect is driven by a preponderance of cardinal orientations in the environment. In support of this notion, the results from Fourier analyses of natural scenes have shown that cardinal orientations contain more contrast (Fourier amplitude) than oblique orientations.\textsuperscript{14-18} As might be expected, the cardinal bias in the environment measured with Fourier analysis is stronger for scenes that contain man-made structures, referred to as “carpentered” environments, versus those that do not.\textsuperscript{10, 16, 19} This difference between carpentered vs. non-carpentered environments can be used to explain why people who live in less carpentered environments, such as the Cree Indians, exhibit a smaller oblique effect than people who live in highly carpentered environments.\textsuperscript{20} In sum, the results from a variety of studies investigating orientation processing suggest that visual sensitivity to different orientations is influenced by the prevalence of different orientations in the environment.
In addition to orientation, spatial frequency content of natural scenes has been statistically described using Fourier analysis (i.e., amplitude of contrast across spatial frequencies). A robust regularity that has been consistently observed across many studies is that amplitude declines with spatial frequency, where the amplitude at a given spatial frequency $f$ is approximately $1/f$.\(^{21-25}\) (Note that many studies report power which is the square of the amplitude, and the fall-off is $1/f^2$.) Recent studies have shown that visual psychophysical performance in humans is well matched to the $1/f$ function typically encountered in natural scenes. For example, object discrimination sensitivity is best for stimuli that have natural power spectra slopes compared to those that have slopes that deviate from $1/f$.\(^{26-29}\) Also, response properties of neurons in the retina, LGN, and visual cortex have been suggested to be optimally organized to process stimuli with natural image statistics,\(^{14, 18, 22, 23, 30-34}\) (and see Simoncelli and Olshausen\(^{35}\) for a review).

Statistics of other image properties such as contours and color content in natural scenes have also been shown to have a close link to perception. For example, Geisler et al\(^{36}\) showed that human performance for contour detection matched very closely to predicted ideal observer performance based upon the co-occurrence of contours in natural images. Regan\(^{37}\) found that the optimal placement of M and L cones in the retina corresponds well to the observed wavelength distribution in the environment of New World monkeys (see also \(^{38-40}\) for similar findings). Similar links between the environmental statistics of color and the cone spectral sensitivities in humans have also been reported, as well as relationships between environmental statistics and measures of color appearance, identification, and labeling.\(^{41-45}\)

Another way researchers have investigated the role of the environment in shaping visual perception has been to ask which spatial frequencies or orientations in a set of images are most
critical for *discriminating* one image in that set from another. Typically, such studies focus on a
collaborated set of images for which people have ample experience. A good example of such an
image set is printed letters. Several psychophysical studies have measured discrimination of
letters filtered or masked along a band of spatial frequencies. The general consensus from these
studies is that the most important band for letter identification peaks at an object spatial
frequency of 3 cycles per letter.\(^{46-49}\) In recent Fourier analyses of letters, it has been shown that
the highest amplitudes are also found near 3 cycles/letter.\(^{50}\) Thus, these findings indicate that the
task of letter discrimination relies most heavily on spatial frequencies that are most prevalent in
letters, and it is reasonable to assume that this link arises as a result of experience with reading.
Presumably, this “tuning” would be specific for the task of letter discrimination; for other image
sets (such as faces), the spatial frequencies that contain the highest amplitudes are likely to differ
from that of letters, and accordingly, these spatial frequencies might be most important for
discrimination of that set.

In the current study, we focused on images of signers producing American Sign
Language (ASL), and we compared these images to those from established databases of natural
scenes and faces. Our interest in ASL is based on the fact that many deaf people have lifelong
experience with ASL, relying on it for communication (and there also exist many hearing people
who converse fluently or non-fluently in ASL). Like the case for images of written letters
(described *above*), it is possible that, for ASL images, the spatial frequencies (or orientations)
that contain the highest amplitudes are also those most important in discriminating one sign from
another. In order to address this possibility, the image statistics of ASL must first be described.
To this end, we conducted Fourier analyses on a set of images containing a representative sample
of the different hand/arm positions and spatial configurations seen in ASL. The Fourier analyses
of static sign images in the current study focuses only on the position and spatial configuration of the hands/arms. Motion information is the other integral piece of information, and we address motion statistics of ASL in a separate forthcoming paper.

From our Fourier analysis of static images (signs, natural scenes and faces), we obtained two metrics. One, we quantified the mean amplitude of the Fourier component as a function of spatial frequency and orientation, which is the typically reported output from Fourier analysis. Secondly, we quantified the entropy of the amplitude as a function of spatial frequency and orientation, a metric that has yet to be described for natural image sets. Entropy describes the spread of a distribution, and is related to, although not identical to, the variance of this distribution (see Cover and Thomas\textsuperscript{51} or Atick\textsuperscript{52} for a review). Here we looked at the distribution of amplitudes across the image set for a given spatial frequency or orientation. The entropy metric comes from information theory, with the notion that a spatial frequency (or orientation) whose amplitude distribution is broadly spread across the image set contains more “information” for discriminating one image from another than a spatial frequency (or orientation) that almost always has the same amplitude across images. It is important to point out that entropy refers to information in the image set and is not to be confused with psychophysical measurements of information that test (through filtering and masking) which spatial frequencies (or orientations) are most important to the visual system for discriminating one image in a set from another. Whether the two are related (information in the image set and critical information used by the visual system) is one of the questions of the current study.

The results of our Fourier analyses revealed two main findings. First, substantial differences were observed between signs and the two other image sets, faces and natural scenes. In particular, sign images contained more amplitude for vertical, than for horizontal, contours,
while faces and natural scenes showed an opposite pattern. Given the above-mentioned relationship between the prevalence of different orientations in the environment and visual sensitivity, these results suggest that different anisotropies in orientation may exist for signers vs. non-signers. In addition, for sign images only, entropy vs. spatial frequency curves were clearly bandpass, with a peak around \(0.75 \text{ cyc/cm}\) (which is roughly \(0.19 \text{ cyc/degree}\) at a viewing distance of 5 feet). Interestingly, this peak in entropy maps on roughly to the spatial frequency that has been shown to be important for identifying signs in ASL (based on psychophysical studies by Riedl and Sperling\(^{53}\), discussed further, below), which suggests a potential link between entropy and visual discrimination.

Second, for all image sets (signs, natural scenes and faces), the shape of the curve relating mean amplitude vs. spatial frequency (and orientation) was markedly different from that relating entropy vs. spatial frequency (and orientation). These differences raise an interesting question; which aspect of an image set (mean amplitude or entropy of the amplitude) is better able to account for known perceptual abilities? In particular, in our orientation analyses of natural scenes, mean amplitudes were largest for cardinal orientations, in line with results from previous Fourier analyses, and in line with the psychophysical “oblique effect” (see above). By contrast, entropy was found to be lowest for cardinal orientations, particularly for horizontal contours. Interestingly, these results are in line with recent psychophysical data showing that the “oblique effect” (i.e., poorest contrast sensitivity for oblique contours) that is known to exist for stimuli of a single spatial frequency turns into a “horizontal effect” (i.e., poorest contrast sensitivity for horizontal contours) for stimuli of broadband spatial frequencies that are more naturalistic.\(^9,^{10}\)

Together, these current Fourier analyses and previous psychophysical results suggest a potential link between entropy and visual sensitivity.
Methods

**Photographs of American Sign Language (ASL).** To obtain a database of ASL images, we took photographs of two female signers (RB and DH) producing signs from American Sign Language. The photographs were taken with a Canon Rebel, 35 mm Single Lens Reflex camera, with a 200 speed setting and with a flash, at a distance of 63.5 inches. The signers were asked to produce 40 different signs, which were selected with the goal of photographing a range of hand shapes and arm positions commonly seen in ASL (~ 15 different hand/arm positions). Note that it was not our goal to have each image associable with a word, but rather, to capture positional information that make up the structure of signs. These sign images are segments of whole “words” and can be thought of as making up constituents within ASL, much the way spoken language has a finite set of vowels and consonants that are combined sequentially to create spoken words. For this reason, we refer to our ASL database as “sign images”. In addition to providing a representative sample of the different hand/arm positions in ASL, the 40 signs were also selected such that the proportion of signs that were one-handed vs. two handed (30% vs. 70%, respectively) and the proportion of two-handed signs that were symmetric vs. asymmetric (50%) roughly matched that which has been observed in ASL.54

To make our sign images as naturalistic as possible, signers were asked to sign a given word, but to momentarily freeze at a certain point so that the photograph of the prototypical arm/hand position for that sign could be taken. For example, in the sign CANADA (see Figure 1A), a photograph was taken with the signer’s fist on her chest. For signs that involved continuous circular motion (e.g., BICYCLE, ENJOY, GESTURE, WASH), we asked signers to freeze halfway through the motion cycle. In total, our ASL database consisted of 59 and 46 photographs of sign images.
taken of signer DH and RB, respectively. In addition to obtaining photographs of sign images, for comparison purposes, a photograph was taken of each signer with her arms resting, which we refer to as a “neutral pose” (see Figure 1B).

**Sign Image Preparation.** Each photograph was scanned using an Epson Scanner set at 600 dpi with a window size, 1800 x 2400 pixels. Mean interocular distance across all images was 138 pixels. For each image, we removed the face, the torso and the background, leaving only the arms, from shoulders to fingers, as shown in Figure 1C (sign image from the word CANADA) and Figure 1D (neutral pose). The purpose in doing so was to restrict our analyses to the signers’ arms and hands, because the arms and hands provide the lexical and grammatical information in ASL. Having the same face in each ASL image would add redundancy to the analysis. This is not to say that the face (specifically, facial expressions) does not contribute to ASL comprehension, however, the lexical identity of any sign can be comprehended without the face. (Note that we conducted a separate analysis on a set of faces obtained from a published database, where each face in the database is a different identity, see below). Finally, eliminating all but the arms in these photographs also served to eliminate noise from irrelevant portions of the image, for example, wrinkles in the clothing. Examples of several sign images used in our analyses are provided in Figure 2A.

**Images of Faces and Natural Scenes.** For comparison to sign images, we performed the same analyses on images of faces and natural scenes. Face images consisted of neutral expression images of 100 individuals of various races and genders from the AR face database (see example in Figure 2B). These images were 768 x 576 pixels with a mean interocular distance of 106 pixels. Natural scenes were taken from the Groningen natural image database, and were 1536 x 1024 pixels (see example in Figure 2C). How the different image sets (signs,
faces and natural scenes) were aligned in scale is discussed further, below. This database contains 4000 images of outdoor scenes, including buildings as well as natural landscapes and close-up foliage texture. We selected 100 images from the first 400 images of the database that contained no visible man-made structures. For all image sets (signs, faces and natural scenes), images were normalized in luminance by setting the brightest and darkest gray values of each image to 255 and 0, respectively, and linearly rescaling the gray values in between.

Scale alignment. The Fourier analysis (described below) outputs frequency in terms of cycles per image. However, the three image databases contained images of different size. The images were not rescaled to contain the same number of pixels. Instead, cycles per degree (cyc/deg) was estimated for each image database, as well as cycles per centimeter (cyc/cm) in the world where applicable. (We chose this because downscaling causes high frequency information to be lost while upscaling does not add high frequency information, and because cycles per image is not as relevant as cyc/deg or cyc/cm in the world.) For the ASL and face images, cycles per image was converted first to cycles per centimeter in the real world using the ratio of face-width in pixels to actual face-width in centimeters. Cycles per degree was then estimated by assuming a viewing distance of five feet (i.e., 2.54 cm = 1 degree). We chose this distance because it is typical for conversing signers. We also assumed that this distance would be typical for viewing faces during conversation between non-signers. For the natural scenes database, cyc/cm in the world could not be calculated since the images contain a range of scales in the foreground and background. Instead, cycles per image was converted directly to cycles per degree using the published angular resolution of the pictures, which is approximately 1 minute of arc per pixel, making a viewing angle of 17 by 25.5 degrees.\textsuperscript{56}
Fourier Analysis. Fourier analysis was conducted on each image using a discrete fast Fourier transform function in MATLAB (by MathWorks), using wrap-around with no padding. Analyses were performed on the amplitude output (defined as the square root of the power, or energy). For each image, the Fourier transform can be displayed as a 3D polar plot \((r, \theta, a)\), where the radial distance from the origin \((r)\) is spatial frequency; the angle \((\theta)\) is orientation; and the brightness at each point in the plot is the amplitude \((a)\) at a given spatial frequency and orientation. Examples of these Fourier transform plots are provided in Figure 3, for two images; a sign image from the word HEART (Figure 3A) and an image of a natural scene (Figure 3B). For all analyses, the corners of the Fourier transform were excluded since they contain high frequencies at oblique orientations that are beyond the resolution limit at the cardinal orientations. Corners were defined as radial distances that exceed the maximum radial distance for any one direction \((r > \min_\theta (\max(r_\theta)))\) where \(r_\theta\) is the radial distance in direction \(\theta\), and \(\min_\theta\) denotes the minimum over all orientations \(\theta\). In other words, the analysis was performed only for those radial distances \(r\) for which data was available at all orientations. Also, note that horizontal contours in an image have vertical Fourier energy since orientation refers to the orientation of individual sine waves. In the results presented from here forward, orientation refers to the orientation in the image, which is 90 degrees shifted from the Fourier orientation.

For all image sets, Fourier data were analyzed at spatial frequencies ranging from 0.06 to 23.0 cyc/deg (0.02 to 9.0 cyc/cm). Fourier data for orientation were analyzed from 0 to 180 degrees (with 0 and 90 degrees denoting horizontal and vertical contours, respectively). From the Fourier analyses, we computed two metrics as a function of spatial frequency and orientation, mean amplitude and entropy of the amplitude, described in detail, below.
Mean Amplitude Analyses. For each image set, mean amplitudes were calculated across spatial frequency and across orientation. Two dimensional (2D) plots of “amplitude vs. spatial frequency” were obtained by averaging across orientation within each image, and then averaging across images in a given image set. The data were then plotted in log-log coordinates, which is the conventional method of analyzing amplitude by spatial frequency data, thus allowing us to compare our data to those of previous studies. For the 2D plots of “amplitude vs. orientation”, we first logged (base 10) the values before collapsing across spatial frequency because spatial frequency data are known to be highly non-linear, whereas they are linear on a log_{10} scale. The 2D data were then averaged across the set of images.

Entropy of the Amplitude. A spatial frequency (or orientation) with high mean amplitude may nevertheless provide little information for differentiating images if it tends to frequently take the same high value, compared to a spatial frequency (or orientation) with high variability in amplitude across images. Therefore, in addition to measuring mean amplitude, we also measured the entropy of the amplitude distribution. In brief, entropy measures the spread of the probability distribution, and is highest for broad distributions and lowest for narrow distributions. For Gaussian distributions, entropy is equivalent to variance. However, it is well known that many natural signals have distributions that are highly non-Gaussian, including measures of contrast in natural scenes.\textsuperscript{57-61}

Entropy is a metric from information theory, which describes the amount of information in a signal for differentiating exemplars. (See Cover and Thomas\textsuperscript{51} for an introductory text on information theory.) Information theory defines the amount of information (I) in an outcome \(x\) as follows:
\[ I(x) = -\log_2(p(x)) \]  

where \( p(x) \) is the probability that the measure takes the value \( x \) across the set of images. (Note that the log in Equation 1 is base 2 since the definition of information comes from bits.\(^{51}\)) In the current study, \( x \) is an amplitude value (at a given spatial frequency or orientation). Entropy (\( H \)) is the expected value of the information across the full distribution of values \( x \) can assume, as follows:

\[ H = E(-\log_2(p(x))) \]  

where \( E \) is the expected value. For each image database, entropy was calculated \( across images \) for each spatial frequency and orientation. Entropy was calculated using equation 2, where \( x = \log(a_{f\theta}) \), which is the set of log amplitudes at frequency \( f \) and orientation \( \theta \) across all images in the database. Probability distributions were estimated by histogramming \( x \) into 20 equally spaced bins. Identical bins were employed at all spatial frequencies and orientations. The bin range for each image database was defined by \( \min(x) \) and \( \max(x) \) over all spatial frequencies, orientations, and images. The entropy analysis produced a single 3D entropy plot (frequency X orientation X entropy) for each image database. Like the amplitude data (\( described \ above \)), the 3D surface was then reduced to 2D by collapsing across spatial frequency and orientation separately, producing plots of “entropy vs. spatial frequency” and “entropy vs. orientation”. For our analysis, we compare the entropy peaks and shape of entropy functions for each image database, rather than compare the absolute entropy values across databases.

**Results**
Spatial Frequency: Mean Amplitude Analysis. Mean amplitudes are plotted as a function of spatial frequency in Figure 4 on a log-log scale. Data obtained for sign images (black lines) are presented separately for the two signers, RB (Figure 4A) and DH (Figure 4B). For comparison, data are also presented for a neutral pose (with arms at rest) for each signer (grey lines). Also, for comparison, data are shown for the set of faces taken from the AR database (red lines) and natural scenes taken from the Groningen database (green lines). Note that the faces and natural scenes data are plotted redundantly in Figure 4A and 4B, to allow comparison to each signer’s data.

For all image sets, amplitudes were found to be largest at low spatial frequencies, declining approximately according to 1/f where f is spatial frequency, and were approximately linear on a log-log scale. This 1/f pattern has been widely reported for natural scenes. More precisely, the amplitude spectrum of natural scenes has been found to take the form 1/f^exponent, with the exponent being equivalent to the slope of log amplitude against log frequency. In log-log coordinates, the slopes are found to be near -1, ranging from -0.7 to -1.6 across studies. (Note that some studies report results for the power spectrum, which is the square of the amplitude spectrum, and hence the exponents and slopes from these studies will be double those for the amplitude spectrum). In the current study, where we measure the amplitude spectrum in log-log coordinates, we find a slope value for natural scenes of -1.17, which is within the previously reported range. For the face database, the slope was -2.06. For sign images, the slope value was approximately -1.66, which was nearly identical for the two signers (RB = -1.68 DH = -1.64). The slope value for the neutral pose was similar to that of the sign images (RB = -1.64 DH = -1.61), with a trend to be slightly shallower. In sum, all image sets conformed to a 1/f^exponent pattern, with the steepest slopes found for faces, followed by...
sign images and then natural scenes. However, the fact that signing arms and the neutral pose yielded quite similar curves suggests that signing arms are unlikely to contain spatial frequency information much different from that of non-signing arms, an issue we return to in the Discussion.

**Spatial Frequency: Entropy Analysis.** Entropy of the amplitude distribution is plotted as a function of spatial frequency in Figure 5, separately for sign images of signer RB (*thick black line*), signer DH (*thin black line*), faces (*red line*) and natural scenes (*green line*). Note that entropy cannot be calculated for the neutral pose condition, since entropy was calculated across images and the neutral pose was a single image. For sign images from both signers, the entropy curves appeared bandpass, with a peak in entropy at 2.0 and 1.8 cyc/deg for RB and DH, respectively (0.8 and 0.7 cyc/cm). These peaks in entropy occur at roughly the same spatial frequencies that have been shown to be important for identifying signs in ASL (based on psychophysical studies⁵³), an issue we return to in the discussion. There was also a trend for a bandpass shape in the face images, with a peak at 7.0 cyc/deg (2.8 cyc/cm). For natural scenes, entropy was roughly constant up to 1 cyc/deg, increased markedly between 1 cyc/deg and 3 cyc/deg, and then remained roughly constant as spatial frequency was increased further. Note that the shapes of the entropy vs. spatial frequency curves for all image sets were very different from the shapes of the mean amplitude vs. spatial frequency curves of Figure 4. That is, for all image sets, mean amplitude decreased with increasing spatial frequency in a 1/ƒ fashion, whereas entropy curves for spatial frequency were largely bandpass in nature.

**Orientation: Mean Amplitude Analysis.** Mean amplitudes are plotted as a function of orientation in Figure 6 (0 degrees denotes horizontal, 90 degrees denotes vertical, contours in the image). Data obtained for sign images (*black lines*) and the neutral pose (*grey lines*) are
presented separately for the two signers, RB (Figure 6A) and DH (Figure 6B). For both sign images and the neutral pose, there was a clear peak in amplitude at 90 degrees. This vertical bias (which was stronger for the neutral pose), is due, of course, to the fact that arms (specifically arms at rest, as in the neutral pose) are oriented vertically. The fact that signing arms and the neutral pose yielded quite similar curves suggests that signing arms are unlikely to contain orientation information much different from that of non-signing arms, an issue we return to in the Discussion.

For comparison, data are shown for the set of faces (red lines) and natural scenes (green lines). Note that the faces and natural scenes data are plotted redundantly in Figure 6A and 6B, to allow comparison to each signer’s data. For natural scenes, peaks in amplitude were observed at the cardinal orientations (0 degrees and 90 degrees), which is in agreement with previously reported results for amplitude analyses of natural scenes.14, 17-19, 21 Similarly, for faces, peaks in amplitude were observed at the cardinal orientations, however, the peak was much stronger at 0 degrees (horizontal).

**Orientation: Entropy Analysis.** Entropy of the amplitude distribution is plotted as a function of orientation in Figure 7, separately for sign images of signer RB (thick black line), signer DH (thin black line), the set of faces (red line) and natural scenes (green line). As was the case for the spatial frequency analyses, for the orientation analyses, the entropy curves and mean amplitude curves were quite different in shape. This is revealed in several interesting ways. First, for natural scenes, the clear cardinal bias seen in the mean amplitude data (i.e., peaks in mean amplitude near 0 and 90 degrees, see Figure 6) actually reversed in terms of entropy. That is, for entropy there was a clear nadir at 0 degrees (and a smaller nadir at 90 degrees). Second, for faces, the clear peak at 0 degrees in the mean amplitude data was absent in the entropy data.
Third, for sign images, the clear peak at 90 degrees in the amplitude data (seen for both signers) was largely attenuated in the entropy data. In addition, the entropy data for signer RB showed a peak at 0 degrees, which was not observed in the amplitude data. We return to the relevance of these differences in the Discussion.

Discussion

The current study employed Fourier analysis to measure mean amplitude and entropy of the amplitude as a function of spatial frequency and orientation in three different sets of images: faces and natural scenes (from published databases) and ASL signs (from our own database). To our knowledge, this is the first study to quantify the image statistics of ASL. This research is novel in that it characterizes the low level visual content of a visual language in the same way classic phonetics research characterizes the low level auditory content of spoken languages. In addition, the current study introduces the concept of entropy across a set of images with the notion that, in addition to mean amplitude, this type of information in an image set may play a role in shaping visual perception.

In this section, we discuss the results of the current (and previous) Fourier analyses conducted on different databases, with the goal of relating the findings to known (or predicted) visual psychophysical results. Before proceeding with the discussion, it is worthwhile pointing out that there are two main types of psychophysical tests that have been used to establish links between image statistics and visual perception. One, simple sensitivities (e.g., contrast sensitivity or acuity) are determined along a given stimulus dimension (like spatial frequency or orientation). Typically, the goal has been to link these psychophysical sensitivity data to biases in the environment as measured with the mean amplitude metric of the Fourier analysis (for
example, as in the “oblique effect”). Two, using a set of real world images (within a confined set, such as printed letters or faces), different spatial frequency or orientation bands are filtered or masked to determine which bands are most critical to the visual system for discriminating one image in that set from another. Here, the goal has been to ask whether the most critical band for visual discrimination is tied to that which contains the greatest mean amplitude (as has been suggested for printed letters, see Introduction). The novel contribution of the current paper is the suggestion that contrast sensitivity and/or visual discrimination may be tied to entropy of the amplitude in an image set, rather than (or in addition to) the mean amplitude. Below, we begin with a discussion of natural scenes, followed by faces and then ASL signs.

**Natural Scenes.** With regard to spatial frequency, the results of our amplitude analyses revealed a $1/f$ function, which is in line with previous findings for natural scenes.$^{21-25}$ The slope of -1.17 for amplitude vs. spatial frequency on a log scale is similar to slopes reported previously in the literature (e.g. Tolhurst et al$^{25}$ reported a mean slope of -1.2). See Ruderman$^{62}$ and Balboa, Tyler, and Grzywacz$^{63}$ for discussions of the origins of $1/f$ scaling in natural scenes, and see Atick and Redlich$^{30}$ for a theory of how the $1/f$ scaling relates to contrast sensitivity in the retina.

The results of our entropy analyses of natural scenes looked quite different from the mean amplitude analyses. Spatial frequencies over ~2 cyc/degree had more entropy compared to spatial frequencies under ~2 cyc/degree, however, above and below this spatial frequency, entropy was relatively constant. The difference between the amplitude and entropy results raises the interesting question of which measure (if either) is more influential in shaping perception. If entropy plays a role, psychophysical studies that employ bandpass filtering or masking of natural scenes might show that middle and high spatial frequencies are more critical than low spatial frequencies for discriminating one natural scene from another.
In the orientation domain, the lowest amplitudes in natural scenes were observed for oblique orientations, which is in agreement with findings from previous studies of natural scenes.\textsuperscript{14-18} For many years it has been believed that this orientation bias in the environment is what drives the perceptual “oblique effect”; acuity and contrast sensitivity are worse at oblique orientations as compared to cardinal orientations\textsuperscript{6-8} (see Introduction). However, the results of our entropy analysis were essentially opposite to the results of our amplitude analysis (compare green lines in Figure 6 vs. Figure 7). More specifically, we observed the lowest amount of entropy for horizontal contours (i.e., troughs in entropy at 0 degrees).

At first glance, the fact that our mean amplitude results, but not our entropy results, mirror the perceptual oblique effect would seem to suggest that mean amplitude across an image set is more influential in shaping perception than is the entropy in that image set. However, recent psychophysical present a challenging point of view.\textsuperscript{9, 10} In these studies it was shown that the “oblique effect” (i.e., poorest contrast sensitivity for oblique contours) that is known to exist for stimuli of a single spatial frequency turns into a “horizontal effect” (i.e., poorest contrast sensitivity for horizontal contours) for stimuli of broadband spatial frequencies that are more naturalistic. Also, the horizontal effect was found to be strongest when the slope of the $1/f$ spatial frequency distribution in the broadband stimulus was between 0.75 and 1 (in log-log coordinates), which matches the statistics of natural scenes. The similarity between these psychophysical findings of a “horizontal effect” and the sharp decrease in entropy seen at horizontal contours in the current study suggests that entropy might, in fact, play a role in shaping perceptual sensitivity. One way in which this could occur is if the visual system allocates less dynamic range to orientations with lesser entropy in the environment.
Another explanation for the “horizontal effect”, suggested by Hansen and Essock, is that the effect is due to contrast normalization in visual cortex. Interestingly, these two hypotheses, contrast normalization (Hansen and Essock’s study) and entropy in the environment (current study), are not incompatible. Both accounts involve relying on orientations with the least redundancy, either within images as in the contrast normalization model, or across images as in the current entropy analysis. Brady and Field\textsuperscript{57} showed that divisive contrast normalization increases the entropy of neural responses to natural scenes. Brady and Field’s study was primarily a within-image analysis, whereas the present study looks at entropy across images, but it shows that the computational process involved in divisive normalization are related to information carrying capacity, a finding that is supported by Simoncelli and Schwartz.\textsuperscript{64} In any event, one way to test whether entropy plays a role in shaping visual perception would be to use the filtering or masking technique to determine psychophysically whether horizontal contours are least critical for discriminating one natural scene from another.

**Faces.** With regard to spatial frequency, the results of our amplitude analyses revealed a $1/f$ function, which is in line with previous findings for faces (e.g., Torralba and Oliva\textsuperscript{19}). Interestingly, however, the slope in log-log coordinates was steeper than for natural scenes (see Figure 4). In our entropy analysis, we observed a small peak at 2.8 cyc/cm (which converts to 7 cyc/deg, assuming a viewing distance of 5 feet). These entropy data for faces can be compared to the results of psychophysical studies using filtering or masking to determine which spatial frequencies human observers rely on most for discriminating faces. Although there is some differences across studies, the general consensus is that 10 to 15 cycles/face contains the most information used by human observers;\textsuperscript{65-68} with one study\textsuperscript{69} being slightly higher at 20 cycles/face width. The cycles per cm of the current study can be converted to cycles/face by
multiplying by the mean face width of our two signers (15.7 cm). This shows the peak in the entropy data to be 43.2 cycles/face, which is much higher than the peak cycles/face determined psychophysically. However, note that although entropy peaks at about 43.2 cycles/face in our analyses, there is a broad range in this peak, which includes the range of important frequencies reported in the psychophysical literature (10 – 15 cycles/face). Still, our entropy data seem to overestimate the contribution of high spatial frequencies.

The orientation results for faces revealed a cardinal bias in mean amplitude of faces (as was observed for natural scenes) although the effect was stronger for horizontal than vertical contours. This cardinal bias seen in the amplitude for faces, which was also reported by Torralba and Oliva\textsuperscript{19}, was largely attenuated in the entropy data of faces (compare red lines in Figure 6 vs. Figure 7).

Signs in ASL. The statistics of sign images were found to differ from those observed for natural scenes and faces in the following ways. First, although mean amplitude analyses of all image sets revealed a $1/f$ function, the slope for sign images was steeper than for natural scenes but shallower than for faces. Because the slope of the sign images falls midway between that of two other image sets for which signers should also have ample experience (faces and natural scenes), it is not entirely obvious whether spatial frequency sensitivity would be expected to differ between signers and non-signers. To date, a single study has compared contrast sensitivity to different spatial frequencies in signers vs. non-signers.\textsuperscript{70} For the three spatial frequencies tested in that study (0.5, 2 and 9 cyc/degree), no significant differences were found between groups, nor were any interactions between subject group and spatial frequency found. Such findings suggest that exposure to the spatial frequency distribution of sign images has no effect
on contrast sensitivity, although the relatively few number of spatial frequencies tested in this study may not have been sufficient for revealing subject group differences.

Second, although all image sets revealed higher mean amplitudes for cardinal, than for oblique, orientations, sign images contained more amplitude for vertical, than for horizontal, contours, while faces and natural scenes showed an opposite pattern. This difference may not be surprising since signing arms are often oriented vertically. (This is especially true for one-handed signs, where the non-dominant arm is usually in a vertical position.) If there is a relationship between the prevalence of different orientations in the environment and visual sensitivity (see Introduction), these mean amplitude analyses for orientation suggest that anisotropies in orientation perception may differ between individuals with ASL experience (signers) and those without (non-signers). Future psychophysical studies will be required to test this hypothesis.

A third way in which the statistics of ASL images were found to differ from those observed for natural scenes and faces was in the entropy analyses for spatial frequency (see Figure 5). Here, sign images revealed a bandpass pattern with a peak in entropy at approximately 0.75 cyc/cm (which converts to 1.9 cyc/deg, assuming a typical viewing distance of 5 feet). This result suggests that spatial frequencies near this peak contain the greatest amount of information for distinguishing one sign image from another. To investigate whether this entropy signal in the image set is actually used by human observers, one can ask whether the spatial frequencies that contain the highest entropy in an image set are those most critical for psychophysical discrimination of sign images. Psychophysical data from Riedl and Sperling allow us to address this possibility. In their study, they presented spatial frequency filtered signs to determine which spatial frequencies are most critical for signers to be able to discriminate one
sign from another. They used four spatial frequency filters with peaks of 0.05, 0.20, 0.35 and 0.75 cyc/cm, and found that 0.75 cyc/cm was most important. Note that the investigators were actually attempting to make it so that the four different spatial frequency bands yielded equal performance (which they attempted to do by keeping the peak frequency constant but varying the bandwidth of the filters), yet they still found that the subjects’ most accurate performance was seen at approximately 0.75 cyc/cm.

Although Riedl and Sperling\textsuperscript{53} did not test high enough spatial frequency bands to determine whether the 0.75 cyc/cm band was the peak (since it was the highest band tested), their results are, at the very least, in line with those of the current study, which showed that the image statistics of signs contain the highest entropy (i.e., the greatest amount of information) at around 0.75 cyc/cm. This consistency between the entropy statistics of signs here and psychophysical discrimination of signs reported by Riedl and Sperling suggests that the mechanisms involved in comprehension of sign language might become tuned to rely on spatial frequencies in signs that contain the greatest entropy.

Are Signing Arms/Hands Unique? On a final note, it is important to point out that if perceptual differences are found to exist between signers and non-signers, and if these differences appear to relate to the image statistics of the arms/hands in ASL, we suggest that it is more likely due to signers’ reliance on ASL, rather than to their mere exposure. This is because the case could easily be made that both signers and non-signers alike receive ample (and roughly equal) experience viewing arms/hands, whether those arms are signing (experienced by signers only), gesturing or at rest (the latter two being experienced by signers and non-signers). And in fact, the results of our mean amplitude analyses revealed only minor differences between sign images and the neutral pose (in terms of the shapes of the spatial frequency and orientation
curves). Thus, differences between signers and non-signers would most likely be attributable to that fact that signers pay much more attention to the hands/arms because this part of the body is crucial for ASL communication.

Finally, the current Fourier analysis focuses on the static information carried in ASL, i.e., in the arm/hand position that carry lexical information. However, the motion of the hands/arms also carry critical information in ASL, and we address the statistics of the motion in ASL in a forthcoming paper on the motion statistics of ASL. In any event, it is reasonable to suggest that visual perception may differ between signers and non-signers in predictable ways based on the (static or moving) image statistics of signs. Many studies have indeed reported differential spatial or motion processing between signers and non-signers, which are thought to be due to ASL experience. However, because the spatial frequency/orientation makeup of the stimuli in these previous studies were not controlled, we cannot yet determine whether differences between signers and non-signers in these studies might be linked to exposure to the visual statistics of signing arms/hands. Future experiments in our laboratory are currently underway to investigate these possibilities.

Acknowledgements

This work was supported by an NSF grant BCS-0241557 to K. R. Dobkins and an NSF IIS-0220141 to M. S. Bartlett
Figure Legends

**Figure 1.** Photography of ASL signs and Image Preparation. (A) An example of an original photograph image of DH signing CANADA and (B) Original photograph image of DH in a neutral pose (arms at side). In (C) and (D), the face, torso and background in these two images have been removed, leaving only the arms, from shoulders to fingers. Fourier and entropy analyses were conducted on the arms-only images (C and D).

**Figure 2.** Example Sign Images. (A) Several examples are shown of sign images included in our ASL sample. The *top row* has images for the sign DOCTOR and ENJOY for each RB and DH. For some signs that involve critical changes in hand shape (e.g., ARREST) or position (e.g., LIFE), two images were obtained, and these are shown in the *middle and last rows* (only for DH). (B) Example face image from the AR face database.\(^{55}\) (C) Example of a natural scene image from the van Hateren & van der Schaaf (1998) database.

**Figure 3:** Example Fourier Plots. Examples are shown for two prepared images: (A) the sign image obtained from the sign HEART shown in Figure 2A and (B) the natural scene shown in Figure 2C. For each image, the Fourier transform is displayed as a 3D polar plot, where log amplitude \(a\) is plotted as a function of both spatial frequency (distance from the origin, \(r\)) and orientation (angle, \(\theta\)). Note that horizontal objects in an image have vertical Fourier energy since orientation refers to the orientation of individual sine waves. (In the results and figures we refer to orientation as the *orientation in the image*, which is 90 degrees shifted from the Fourier
orientation.) The natural scenes plot shows high energy in the vertical orientation, which has been related to horizontal layering in distant natural scenes.\textsuperscript{19}

**Figure 4.** Amplitude vs. Spatial Frequency. Mean log amplitude is plotted as a function of log spatial frequency (data collapsed across orientation) for each signer (A) RB (n = 46 images) and (B) DH (n = 59 images). Data are shown for sign images (*black lines*) and the neutral pose (*grey lines*). For comparison, data are shown for a set of 100 faces taken from the AR database (*red lines*) and a set of 100 natural scene images taken from the van Hateren & van der Schaaf (1998) database (*green lines*). Note that the faces and natural scenes data are plotted redundantly in A and B, to allow comparison to each signer’s data.

**Figure 5.** Entropy vs. Spatial Frequency. Entropy is plotted as a function of log spatial frequency (data collapsed across orientation) for signer RB (*thick black line*) and DH (*thin black line*), faces (*red line*), and natural scenes (*green line*). Entropy could not be calculated for the neutral pose, which is a single image, as entropy is calculated from a distribution across many images.

**Figure 6.** Amplitude vs. Orientation. Mean log amplitude is plotted as a function of orientation (data collapsed across spatial frequency) for each signer (A) RB (n = 46 images) and (B) DH (n = 59 images). Data are shown for sign images (*black lines*) and the neutral pose (*grey lines*). For comparison, data are shown for the set of 100 faces taken from the AR database (*red lines*) and the set of 100 natural scene images taken from the van Hateren & van der Schaaf (1998) database (*green lines*). Note that the faces and natural scenes data are plotted redundantly in A
and B, to allow comparison to each signer’s data. Also note that the X axis is centered at 0 degrees (horizontal) to emphasize the sharpest peak in the data.

**Figure 7.** Entropy vs. Orientation. Entropy is plotted as a function of orientation (data collapsed across spatial frequency) for signer RB (*thick black line*) and DH (*thin black line*), faces (*red line*) and natural scenes (*green line*). Entropy could not be calculated for the neutral pose, which is a single image, as entropy is calculated from a distribution across many images. Note that the X axis is centered at 0 degrees (horizontal) to emphasize the sharpest trough in the data.
References


Figure 1. Photography of ASL signs and Image Preparation. (A) An example of an original photograph image of DH signing CANADA and (B) Original photograph image of DH in a neutral pose (arms at side). In (C) and (D), the face, torso and background in these two images have been removed, leaving only the arms, from shoulders to fingers. Fourier and entropy analyses were conducted on the arms-only images (C and D).
Figure 2A

Figure 2. Example Sign Images. (A) Several examples are shown of sign images included in our ASL sample. The top row has images for the sign DOCTOR and ENJOY for each RB and DH. For some signs that involve critical changes in hand shape (e.g., ARREST) or position (e.g., LIFE), two images were obtained, and these are shown in the middle and last rows (only DH). (B) Example face image from the AR face database. (C) Example of a natural scene image from the van Hateren & van der Schaaf (1998) database.
(continued)

**Figure 2.** Example Sign Images. (A) Several examples are shown of sign images included in our ASL sample. The top row has images for the sign DOCTOR and ENJOY for each RB and DH. For some signs that involve critical changes in hand shape (e.g., ARREST) or position (e.g., LIFE), two images were obtained, and these are shown in the middle and last rows (only DH). (B) Example face image from the AR face database.55 (C) Example of a natural scene image from the van Hateren & van der Schaaf (1998) database.
Figure 3: Example Fourier Plots. Examples are shown for two prepared images: (A) the sign image obtained from the sign HEART shown in Figure 2A and (B) the natural scene shown in Figure 2C. For each image, the Fourier transform is displayed as a 3D polar plot, where log amplitude ($a$) is plotted as a function of both spatial frequency (distance from the origin, $r$) and orientation (angle, $\theta$). Note that horizontal objects in an image have vertical Fourier energy since orientation refers to the orientation of individual sine waves. (In the results and figures we refer to orientation as the orientation in the image, which is 90 degrees shifted from the Fourier orientation.) The natural scenes plot shows high energy in the vertical orientation, which has been related to horizontal layering in distant natural scenes.\(^\text{19}\)
Figure 4. Amplitude vs. Spatial Frequency. Mean log amplitude is plotted as a function of log spatial frequency (data collapsed across orientation) for each signer (A) RB (n = 46 images) and (B) DH (n = 59 images). Data are shown for sign images (black lines) and the neutral pose (grey lines). For comparison, data are shown for a set of 100 faces taken from the AR database (red lines) and a set of 100 natural scene images taken from the van Hateren & van der Schaaf (1998) database (green lines). Note that the faces and natural scenes data are plotted redundantly in A and B, to allow comparison to each signer’s data.
**Figure 5.** Entropy vs. Spatial Frequency. Entropy is plotted as a function of log spatial frequency (data collapsed across orientation) for signer RB (*thick black line*) and DH (*thin black line*), faces (*red line*), and natural scenes (*green line*). Entropy could not be calculated for the neutral pose, which is a single image, as entropy is calculated from a distribution across many images.
Figure 6. Amplitude vs. Orientation. Mean log amplitude is plotted as a function of orientation (data collapsed across spatial frequency) for each signer (A) RB (n = 46 images) and (B) DH (n = 59 images). Data are shown for sign images (black lines) and the neutral pose (grey lines). For comparison, data are shown for the set of 100 faces taken from the AR database (red lines) and the set of 100 natural scene images taken from the van Hateren & van der Schaaf (1998) database (green lines). Note that the faces and natural scenes data are plotted redundantly in A and B, to allow comparison to each signer’s data. Also note that the X axis is centered at 0 degrees (horizontal) to emphasize the sharpest peak in the data.
Figure 7. Entropy vs. Orientation. Entropy is plotted as a function of orientation (data collapsed across spatial frequency) for signer RB (thick black line) and DH (thin black line), faces (red line) and natural scenes (green line). Entropy could not be calculated for the neutral pose, which is a single image, as entropy is calculated from a distribution across many images. Note that the X axis is centered at 0 degrees (horizontal) to emphasize the sharpest trough in the data.