The efficiency of biological motion perception

JASON M. GOLD
Indiana University, Bloomington, Indiana

DUJE TADIN
Vanderbilt University, Nashville, Tennessee

SUSAN C. COOK
Indiana University, Bloomington, Indiana

AND

RANDOLPH BLAKE
Vanderbilt University, Nashville, Tennessee

Humans can readily perceive biological motion from point-light (PL) animations, which create an image of a moving human figure by tracing the trajectories of a small number of light points affixed to a moving human body. We have applied ideal observer analysis to a standard biological motion discrimination task involving either full-figure or PL displays. Contrary to current dogma, we find that PL animations can be rich in potential stimulus information but that human observers are remarkably inefficient at exploiting this information. Although our findings do not discount the utility of PL animation, they do provide a realistic measure of the computational challenge posed by biological motion perception.

Among the moving objects we encounter in everyday life, none are more salient or behaviorally relevant than other people. Befitting the importance of perceiving human activity, our brains appear to contain specialized neural machinery for registering the kinematics of biological motion (Grossman, 2005). The ability to perceive biological motion is compellingly demonstrated by point-light (PL) animations that depict human activity using only a small number of light points strategically placed at different locations on a moving human body (Johansson, 1973; Neri, Morrone, & Burr, 1998). Viewing PL animations, human observers can make a wide variety of perceptual judgments, including immediate identification of the activity being executed as well as recognition of the identity, sex, and emotional state of the actor (Blake & Shiffrar, 2007).

Perception of human action as depicted by PL animations is often discussed in the context of two interrelated assumptions (see, e.g., Ahlström, Blake, & Ahlström, 1997; Beauchamp, Lee, Haxby, & Martin, 2003; Blake, Turner, Smoski, Pozdol, & Stone, 2003; Bülthoff, Bülthoff, & Sinha, 1998; Giese & Poggio, 2003; Grill-Spector & Malach, 2004; Heberlein, Adolphs, Tranel, & Damasio, 2004; Pavlova et al., 2005; Puce & Perrett, 2003; Thornton & Vuong, 2004). The first assumption is that PL animations represent an impoverished version of the original, whole-body animation. Although this assumption has intuitive

appeal, it can be substantiated only by carrying out careful physical measurements of the intrinsic difficulty of a given task and the associated biological motion stimulus used in that task. Such an analysis must take into account the physical dissimilarity of the figures being discriminated, the specifics of the task performed by the observers, and the statistics of stimulus noise that limits performance.

A second common assumption about biological motion perception is that human observers are remarkably efficient at using the impoverished information available in PL animations (Ahlström et al., 1997). Again, although it is intuitively appealing, this assumption rests on knowing the intrinsic difficulty of a task involving the discrimination of PL stimuli. To ascertain this difficulty, we need a method that provides a true quantitative measure of how efficiently an observer uses the information provided by the PL figures in a given task. In the present study, we tested these widely held but unsubstantiated assumptions using *ideal observer analysis* (Geisler, 2003; Green & Swets, 1966), which estimates the performance of an observer who is limited only by the physical availability of information in a given task. The decision rule used by such an observer can be derived mathematically using Bayesian inference (Geisler, 2003). Measuring the performance of the ideal observer provides a benchmark of optimal performance from which one can estimate the amount of information used by an actual (e.g., human)

J. M. Gold, jgold@indiana.edu

observer. Comparing actual human performance with that of the ideal observer (a measure known as *efficiency*) factors out the physical limits that are imposed on any human observer performing the task. *Any* departures in performance from that of the ideal must then reflect information loss taking place within the observer, independent of the intrinsic difficulty of the task (Gold, Bennett, & Sekuler, 1999; Tjan, Braje, Legge, & Kersten, 1995).

Using a nonideal model, Pollick, Lestou, Ryu, and Cho (2002) established an upper bound on efficiency for the detection of arm movements using PL stimuli. Their analyses consisted of comparing human performance with the performance of a suboptimal neural network trained to detect arm movements, rather than with the performance of an ideal observer. This approach allowed the establishment of an upper bound on efficiency. This is because an ideal observer performs better than any given suboptimal observer (including Pollick et al.'s, 2002, neural network model observer), thus producing an overestimate of efficiency when human performance is compared with that of a nonideal rather than an ideal observer. Although this approach was effective at establishing an upper bound on efficiency, it did not allow for a true measure of efficiency in the task. Efficiency, measured relative to a suboptimal model observer, is quite hard to interpret, since it reflects a combination of physical task constraints (normally factored out by measuring true ideal performance) and inefficiencies introduced by the processing of information by both the suboptimal model observer and the human observer. That is, comparing human performance with that of a subideal observer in this fashion introduces an additional source of unspecified inefficiency into the efficiency measure, rather than factoring out the physical constraints on performance.

More recently, Pollick, Kay, Heim, and Stringer (2005) made a similar attempt at measuring efficiency with PL displays, this time in the context of gender discrimination. They used an approach similar to that used by Liu, Knill, and Kersten (1995) in which a statistically optimal observer is derived with respect to only a subset of the entire range of physically available information. The utility of this approach is that makes it possible to determine whether a human observer is using more information than that in the subset given to the information-restricted optimal observer. However, this approach differs from ours in that it does not measure the performance of a truly optimal observer that has access to all of the available information in a given task.

In our experiments, we measured both ideal and human performance in a typical biological motion task, using full-figure (FF) stimuli as well as stimuli defined only by points of light. The goal of our experiments was to test the two assumptions described above, namely: (1) that PL stimuli are informationally impoverished relative to FF stimuli and (2) that human observers are highly efficient at using the information available in PL displays. Our results force reconsideration of the impoverished nature of PL animations and the efficiency with which they are perceived.

METHOD

For our experiments, we digitally recorded a human actor walking to the right on a moving treadmill. The actor wore a black spandex body suit that covered his entire body. Fluorescent spots were attached to 13 points on the surface of the actor's suit (one point on each of the actor's shoulders, elbows, hands, hips, knees and feet, and one on the actor's head). Long rods with fluorescent tips at each end were also attached, at their centers, to each point on the actor's left side, which allowed us to recover the exact location of any point that became occluded during the sequence. This was done by finding the midpoint between the fluorescent ends of the rod that corresponded to the occluded point. The background behind the walker was a uniform white field created by an illuminated projection screen.

The walker completed one complete step cycle (i.e., two steps), which we chose to film as a 20-frame movie sequence. We then transformed each frame of this sequence into either a FF silhouette or a PL version of the walker figure (Figure 1). We used Adobe Photoshop 7.0 selection tools to segregate the walker from the background and to generate a silhouette version of the movie sequence. The silhouette walker figure was assigned a uniform negative contrast value and the background was set to midgray (i.e., zero contrast: 49.64 cd/m^2). A corresponding PL version of the walker was generated by tracking the locations of the 13 fluorescent body points in each frame of the sequence. This was carried out by hand using in-house software created in MATLAB 5.2. The coordinates of the points were recorded, and a new signal was generated in which negative contrast squares (2×2 pixels) were placed at each recorded location within a background of midgray. A complementary set of leftward-walking figures was generated by flipping each frame of each rightward-walking stimulus about the vertical axis. Thus, the left and right walking stimuli in each condition were identical except for a reflection about the vertical axis. Each frame of the final signals was 67 pixels high $\times$ 39 pixels wide, subtending $1.42° \times 0.82°$ of visual angle from a viewing distance of 100 cm.
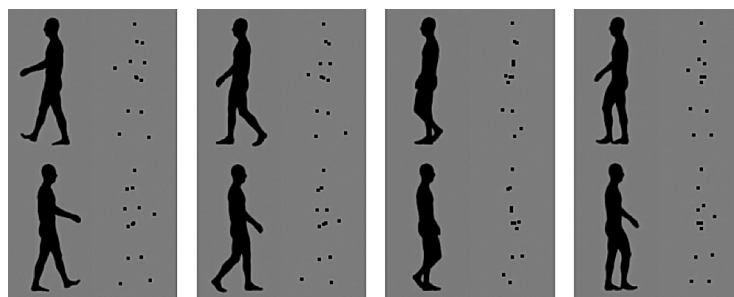

Figure 1. Four example frames from the full-figure (left side of each panel) and point-light (right side of each panel) biological motion stimuli used in the main experiment. The top and bottom halves of each panel correspond to the leftward- and rightward-walking stimuli in each condition, respectively.
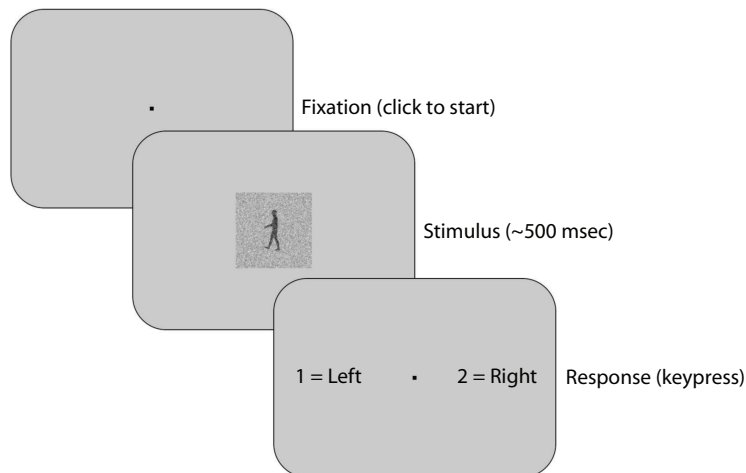
Figure 2. An illustration of the sequence of events that took place within each experimental trial (see text for details).

Observers performed a standard leftward walking versus rightward walking discrimination task commonly used in biological motion experiments (Neri et al., 1998). On any given trial, observers were shown one of two possible stimuli (either a rightward or a leftward walking figure; see Figure 2). To discourage the use of local strategies, such as using the location of a single feature (e.g., the heel of the foot or the point on a hand) to perform the task, the location of the stimulus was randomly jittered by a small amount on each trial relative to the center of the display. Specifically, the position was shifted according to two uniform probability distributions (one for vertical shift and one for horizontal shift). The uniform distributions ranged from −16 to 16 pixels. To enable the ideal observer analysis, we added to each test stimulus a unique sample of high-contrast, Gaussian, white spatiotemporal pixel noise that varied dynamically from frame to frame. This noise field covered the entire range of possible shifted locations of the stimulus as well as an additional 8×36 region of zero contrast ($107 \times 107 \times 20$ pixels, $2.27º \times 2.27º$ of visual angle). The variance of the noise distribution was 0.0625, with power spectral density (i.e., power per unit bandwidth) of 2.83×10^{-5} sec deg^2.

We measured human discrimination performance by varying the contrast energy (i.e., the integrated squared contrast measured in deg^2) of the signals (i.e., the patterns before the noise was added) across trials, using a two-down, one-up staircase procedure that converged on the 71% correct, contrast energy thresholds (300 trials per condition). On each trial, either a left- or a right-walking signal was randomly chosen with equal probability. The signal and noise combination (a series of 20 images, each repeated for two successive frames at a frame rate of 85 Hz) was shown for 470 msec, and observers were asked to make a left–right decision with a keyboard press. Observers received accuracy feedback after each trial. The conditions were blocked and randomly ordered for each observer in all of the experiments.

We were able to derive the statistically optimal decision rule for our task (see Appendix). However, the spatial uncertainty introduced by randomly jittering the location of the stimulus from trial to trial prevented an analytic measure of threshold for the ideal observer. In a typical one-of-two identification task with no externally introduced uncertainty (such as our spatial uncertainty), the ideal observer compares noise-free copies of the two possible signals (templates) with the noisy stimulus to determine which stimulus was present on a given trial. It has been shown (Murray, Bennett, & Sekuler, 2005) that the ideal observer's threshold in such a task can be derived analytically by a simple calculation that involves computing the energy of the difference between the two templates.

However, this simple analytical solution is not possible when spatial uncertainty is introduced into the task. The presence of uncertainty changes the ideal decision rule, since each possible shifted version of the two templates must be considered in order to make a decision. Thus, Monte Carlo simulations were used to measure thresholds for the ideal observer in the presence of spatial uncertainty (see the Appendix for more details about the ideal observer analysis).

RESULTS AND DISCUSSION

The two bars on the left of Figure 3A show contrast energy thresholds for the ideal observer, and the remaining bars show these thresholds for four human observers—two naive (A.S. and M.S.) and two authors (J.G. and S.C.)—in the FF (black bars) and PL (gray bars) conditions. Notice that the ideal observer's thresholds are nearly identical in the FF and the PL conditions. This counterintuitive result indicates that the two tasks are actually equivalent in terms of their intrinsic difficulty: Reducing an FF walker to a small set of points does not necessarily remove information critical to the task. Recall that observers performed a discrimination task in which the critical information was carried by those locations in space and time that differed between the two alternative classes of stimuli. Thus, what is important in this task from the perspective of the ideal observer is the amount of contrast energy that falls on nonoverlapping locations in the set of stimuli. Our ideal observer analysis has shown the contrast energy to be equivalent for both the FF and the PL discrimination tasks, at least for conditions involving spatial uncertainty about the location of the biological figure.

Considering next the task performance data, human thresholds were significantly higher in the PL than in the FF condition. Taken alone, this difference would be difficult to interpret: Human thresholds in the PL condition could be higher simply because the task was intrinsically more difficult to perform in the PL than in the FF condition. However, our ideal observer analysis reveals that the two tasks were in fact equivalent in this regard, implying that the higher human thresholds in the PL condition were
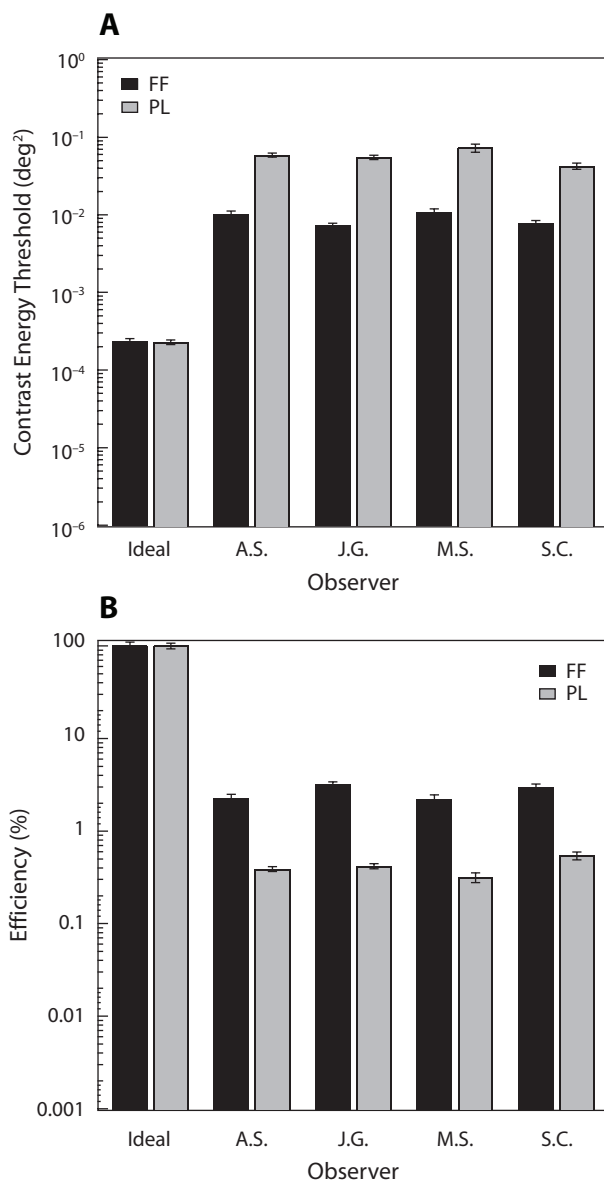
A

B

Figure 3. (A) Contrast energy thresholds and (B) efficiencies for the ideal observer and for four human observers in the main experiment. In this experiment, the spatial location of the stimulus was jittered randomly from trial to trial (see text for details). Error bars correspond to ±1 standard deviation, determined by bootstrap simulations (Efron & Tibshirani, 1993). FF, full figure; PL, point light.

due to less efficient visual processing. This difference is quantified in Figure 3B, which plots efficiency (ideal and human observer thresholds) in each condition for each observer. By definition, ideal efficiency is 100%; human efficiency is far lower, averaging approximately 2.5% in the FF condition and approximately .4% in the PL condition.

Sources of Lower PL Efficiency

Why were human observers so much less efficient at discriminating PL than FF displays? Perhaps human observers were less efficient in the PL than in the FF con-

dition because they used the same FF spatial strategy (or template) in both conditions. It is common to model human pattern discrimination as a noisy template-matching process in which a set of internally stored templates is compared with a stimulus and a decision is made on the basis of their relative similarities (Eckstein, Ahumada, & Watson, 1997; Lange, Georg, & Lappe, 2006; Murray et al., 2005). If observers were using such a strategy in our task and the templates they used in both conditions were similar to the FF stimuli, we would expect efficiency to be far lower in the PL than in the FF condition due to the greater mismatch between the stimuli and the templates. However, Lu and Liu (2006) have recently applied a reverse correlation technique to the analysis of discrimination of PL displays. Their results suggest that in PL displays, observers may make judgments on the basis of the locations where points appear rather than the locations in between points. Such results are inconsistent with the idea that observers use inefficient FF templates when viewing PL animations.

One could argue that the spatial jitter that we introduced to the stimuli had a greater negative impact on visual processing in the PL condition than in the FF condition. We tested this possibility by repeating our experiment without spatial jitter for three new human observers (all naive). All other aspects of the experiment were identical. As Figure 4 shows, removing spatial jitter had very little effect on human thresholds, suggesting that the amount of spatial jitter introduced in the previous experiment was less than or equal to the amount of intrinsic spatial uncertainty introduced by human visual processing. However, this was not the case for the ideal observer: As expected, removing spatial jitter significantly lowered the ideal observer's thresholds in both the PL and FF conditions. Interestingly, the improvement for the ideal observer was greater for the PL than for the FF condition, indicating that the PL stimuli actually carried *more* information than did the FF stimuli in the absence of spatial uncertainty. This is likely due to the facts that the effect of spatial jitter is similar to the effect of spatial blur (i.e., removal of high spatial frequencies) and that the PL stimuli have more energy at high spatial frequencies than FF stimuli do. Figure 5 plots normalized power as a function of spatial frequency for the FF (solid line) and PL (dashed line) stimuli. The power has been averaged across orientations and across frames in each stimulus sequence. This figure shows that power in the PL condition is evenly distributed across spatial frequencies, whereas power is much more concentrated at lower spatial frequencies in the FF condition. Thus, we would expect spatial jitter and/or blur to have a larger impact on performance in the PL condition than in the FF condition for an observer who is not limited by any additional intrinsic spatial uncertainty.

It is also possible that the relatively low efficiency in the PL condition was related to the difference in the distribution of power across spatial frequencies in the two types of stimuli. Efficiency tends to fall off monotonically beyond the peak of the contrast sensitivity function (~4 c/deg), in large part because of preneural factors such as filtering by human optics (Banks, Geisler, & Bennett, 1987; Campbell & Gubisch, 1966). We tested the pos-
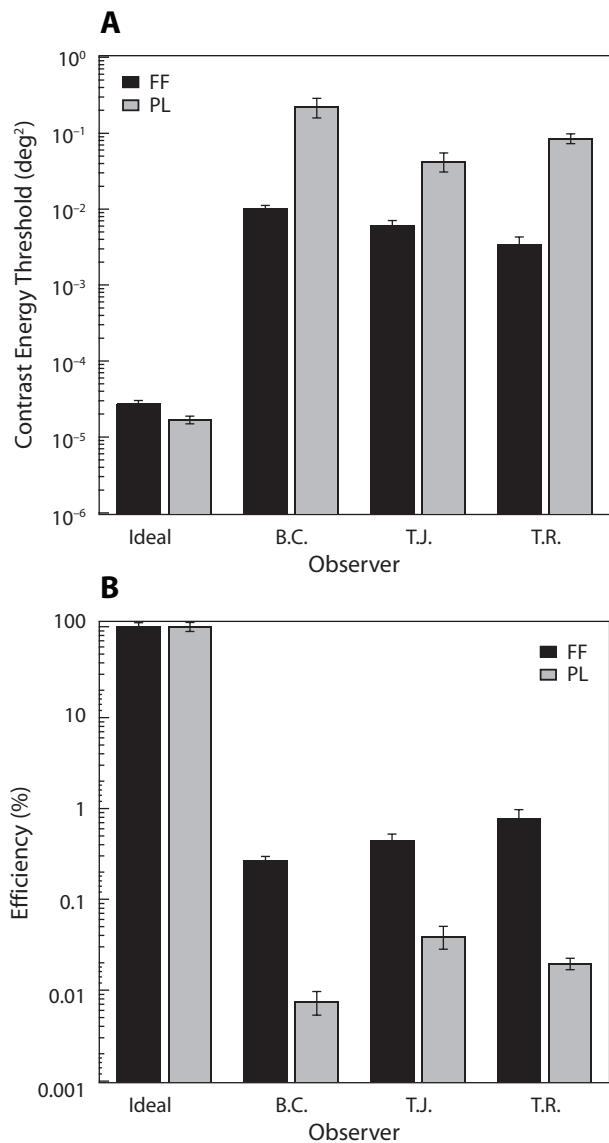
A

B

Figure 4. (A) Contrast energy thresholds and (B) efficiencies for the ideal observer and for three human observers in the no-uncertainty experiment. In this experiment, the spatial location of the stimulus did not vary from trial to trial. Instead, it always appeared at the center of the display. Error bars correspond to ±1 standard deviation. FF, full figure; PL, point light.

sibility that the differences in the distribution of power across spatial frequencies were responsible for the differences in efficiency in the PL and FF conditions by swapping the amplitude spectra of the stimuli and remeasuring thresholds (Gold et al., 1999; Kleiner & Banks, 1987). Switching the amplitude spectra (i.e., the power distribution across frequency components) between patterns while preserving their phase spectra (i.e., the relative positions of the frequency components) had little effect on the spatial structure of the patterns (Oppenheim & Lim, 1981; Piotrowski & Campbell, 1982). All other aspects of the stimuli and experiment were the same as in the initial experiment. If the differences in efficiency for the PL and

FF stimuli were due to the differences in their amplitude spectra, switching the amplitude spectra should reverse the effects found in our first two experiments.

This is precisely what we found (Figure 6). For the human observers (two authors, two naive), thresholds were predicted by the amplitude spectra, not the phase spectra. Thresholds were highest and nearly identical for the two conditions in which the PL amplitude spectrum was used, regardless of the differing phase spectra of the two classes of stimuli. Similarly, thresholds were lowest and nearly identical for the two conditions in which the FF amplitude spectrum was used, strongly suggesting that the differences in amplitude spectra were responsible for the differences in efficiency between the PL and FF conditions. This result also suggests that motion perception performance measured with traditional structure-from-motion stimuli, such as PL walker displays and kinetic surface displays, may be underestimated due to the concentration of a great deal of power at higher frequencies, which are used less efficiently by the human visual system in such tasks. Indeed, in other work, it has been found that manipulations (i.e., diffusion and blur) that selectively degrade high spatial frequency signals in PL animations have no measurable consequence on perception of these stimuli (Ahlström et al., 1997; Barclay, Cutting, & Kozlowski, 1978).

CONCLUSION

Our results lead to some counterintuitive conclusions that call into question several prevailing assumptions about PL stimuli and biological motion perception. First, we
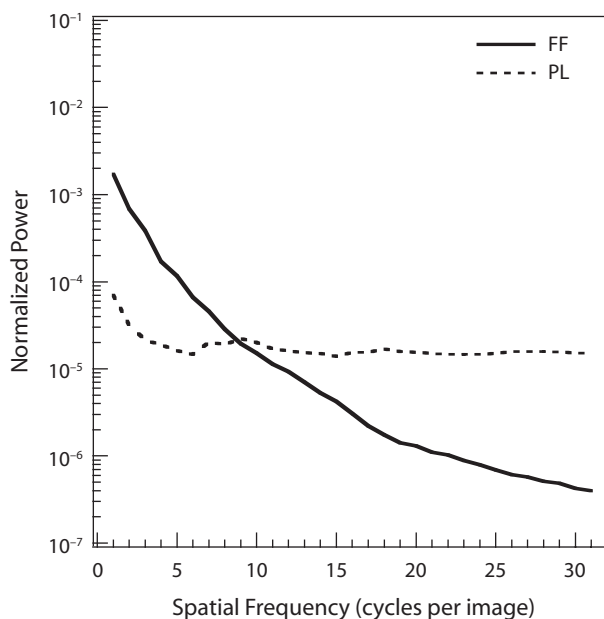

Figure 5. Spatial frequency power spectra (plotted in cycles per image) for the full-figure (FF) and point-light (PL) stimuli. The power was averaged across orientations and across frames in each stimulus movie sequence. Each spectrum has been normalized to have an integrated power of 1.
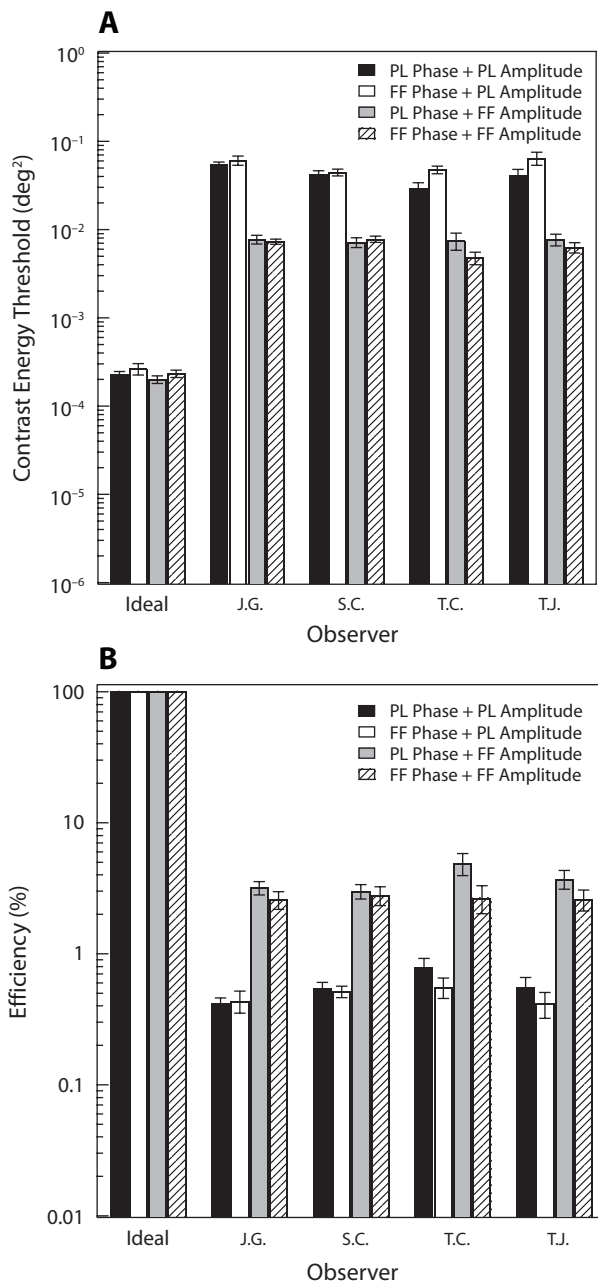
Figure 6. (A) Contrast energy thresholds and (B) efficiencies for the ideal observer and for four human observers in the amplitude-swapping experiment. In this experiment, the amplitude spectrum of each stimulus was paired with either the original phase spectrum or the phase spectrum of the corresponding stimulus in the other condition. Error bars correspond to ±1 standard deviation. PL, point light; FF, full figure.

tained in PL displays. These results offer a cautionary note about conclusions that can be drawn about performance with PL stimuli in the absence of a quantitative measure of informational content. Alternatively, it is also important to keep in mind that the results of an ideal observer analysis are always specific to a particular task, and thus the generalizability of our results to other tasks and stimuli (e.g., gender discrimination and person identification) remains an open question that we are currently exploring.

In addition to considering how our results speak to the use of PL displays, exploring why human observers are so highly inefficient at using the rich information conveyed by our biological motion displays (both PL and FF) also merits attention. The efficiencies in our tasks ranged from about 0.4% (for the PL stimuli) to about 2.5% (for the FF stimuli). Although these values might seem quite low, they do fall within the (lower) range of values obtained in other tasks involving the discrimination of complex patterns such as faces (Gold et al., 1999; Näsänen, 1999), objects (Tjan et al., 1995), and letters (Gold et al., 1999; Pelli, Farell, & Moore, 2003; Tjan, 1996). The maximal efficiency achieved by the ideal observer in these kinds of tasks is a result of the fact that the ideal decision rule has been optimally designed for the particular experimental task at hand—in our case, the discrimination of two perfectly known signals embedded in white Gaussian noise. In contrast, human observers are limited by factors such as internal noise, which can place a low ceiling on efficiency, depending on the task (Gold, Sekuler & Bennett, 2004; Pelli, 1990). Moreover, the human visual system has evolved into its current form for the much more general purpose of detecting, discriminating, and identifying a wide variety of stimuli under highly variable circumstances. For a system such as this, it makes far more sense to emphasize flexibility over specialization, even though this flexibility might come at the cost of statistical inefficiency under certain circumstances.

AUTHOR NOTE

REFERENCES

Ahlström, V., Blake, R., & Ahlström, U. (1997). Perception of biological motion. *Perception*, **26**, 1539-1548.

Banks, M. S., Geisler, W. S., & Bennett, P. J. (1987). The physical limits of grating visibility. *Vision Research*, **27**, 1915-1924.

Barclay, C. D., Cutting, J. E., & Kozlowski, L. T. (1978). Temporal and spatial factors in gait perception that influence gender recognition. *Perception & Psychophysics*, **23**, 145-152.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2003). fMRI responses to video and point-light displays of moving humans and manipulable objects. *Journal of Cognitive Neuroscience*, **15**, 991-1001.

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology*, **58**, 47-73.

Blake, R., Turner, L. M., Smoski, M. J., Pozdol, S. L., & Stone, W. L.

have demonstrated that transforming a FF stimulus into a PL stimulus does not necessarily reduce the amount of information carried by the stimulus. In fact, without added spatial jitter, our PL displays actually carried more information than did the FF stimuli that were used to generate those PL displays. Second, we found that human observers were particularly inefficient at using the information con-

(2003). Visual recognition of biological motion is impaired in children with autism. *Psychological Science*, **14**, 151-157.

BÜLTHOFF, I., BÜLTHOFF, H., & SINHA, P. (1998). Top-down influences on stereoscopic depth-perception. *Nature Neuroscience*, **1**, 254-257.

CAMPBELL, F. W., & GUBISCH, R. W. (1966). Optical quality of the human eye. *Journal of Physiology*, **186**, 558-578.

ECKSTEIN, M. P., AHUMADA, A. J., JR., & WATSON, A. B. (1997). Visual signal detection in structured backgrounds: II. Effects of contrast gain control, background variations, and white noise. *Journal of the Optical Society of America A*, **14**, 2406-2419.

ECKSTEIN, M. P., & WHITING, J. S. (1996). Visual signal detection in structured backgrounds. I. Effect of number of possible spatial locations and signal contrast. *Journal of the Optical Society of America A*, **13**, 1777-1787.

EFRON, B., & TIBSHIRANI, R. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

GEISLER, W. S. (2003). Ideal observer analysis. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (pp. 825-837). Cambridge, MA: MIT Press.

GIESE, M. A., & POGGIO, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*, **4**, 179-192.

GOLD, J. [M.], BENNETT, P. J., & SEKULER, A. B. (1999). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Research*, **39**, 3537-3560.

GOLD, J. M., SEKULER, A. B., & BENNETT, P. J. (2004). Characterizing perceptual learning with external noise. *Cognitive Science*, **28**, 167-207.

GREEN, D. M., & SWETS, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

GRILL-SPECTOR, K., & MALACH, R. (2004). The human visual cortex. *Annual Review of Neuroscience*, **27**, 649-677.

GROSSMAN, E. (2005). Evidence for a network of brain areas involved in perception of biological motion. In G. Knoblich, I. M. Thornton, M. Grosjean, & M. Shiffrar (Eds.), *Human body perception from the inside out* (pp. 361-384). New York: Oxford University Press.

HEBERLEIN, A. S., ADOLPHS, R., TRANEL, D., & DAMASIO, H. (2004). Cortical regions for judgments of emotions and personality traits from point-light walkers. *Journal of Cognitive Neuroscience*, **16**, 1143-1158.

JOHANSSON, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, **14**, 201-211.

KLEINER, K. A., & BANKS, M. S. (1987). Stimulus energy does not account for 2-month-olds' face preferences. *Journal of Experimental Psychology: Human Perception & Performance*, **13**, 594-600.

LANGE, J., GEORG, K., & LAPPE, M. (2006). Visual perception of biological motion by form: A template-matching analysis. *Journal of Vision*, **6**, 836-849.

LIU, Z., KNILL, D. C., & KERSTEN, D. (1995). Object classification for human and ideal observers. *Vision Research*, **35**, 549-568.

LU, H., & LIU, Z. (2006). Computing dynamic classification images from correlation maps. *Journal of Vision*, **6**, 475-483.

MURRAY, R. F., BENNETT, P. J., & SEKULER, A. B. (2005). Classification images predict absolute efficiency. *Journal of Vision*, **5**, 139-149.

NÄSÄNEN, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision Research*, **39**, 3824-3833.

NERI, P., MORRONE, M. C., & BURR, D. C. (1998). Seeing biological motion. *Nature*, **395**, 894-896.

OPPENHEIM, A. V., & LIM, J. S. (1981). The importance of phase in signals. *Proceedings of the IEEE*, **69**, 529-541.

PAVLOVA, M., SOKOLOV, A., STAUDT, M., MARCONATO, F., BIRBAUMER, N., & KRÄGELOH-MANN, I. (2005). Recruitment of periventricular parietal regions in processing cluttered point-light biological motion. *Cerebral Cortex*, **15**, 594-601.

PELLI, D. G. (1990). The quantum efficiency of vision. In C. Blakemore (Ed.), *Vision: Coding and efficiency* (pp. 3-24). New York: Cambridge University Press.

PELLI, D. G., FARELL, B., & MOORE, D. C. (2003). The remarkable inefficiency of word recognition. *Nature*, **423**, 752-756.

PIOTROWSKI, L. N., & CAMPBELL, F. W. (1982). A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase. *Perception*, **11**, 337-346.

POLLICK, F. E., KAY, J. W., HEIM, K., & STRINGER, R. (2005). Gender recognition from point-light walkers. *Journal of Experimental Psychology: Human Perception & Performance*, **31**, 1247-1265.

POLLICK, F. E., LESTOU, V., RYU, J., & CHO, S. B. (2002). Estimating the efficiency of recognizing gender and affect from biological motion. *Vision Research*, **42**, 2345-2355.

PUCE, A., & PERRETT, D. (2003). Electrophysiology and brain imaging of biological motion. *Philosophical Transactions of the Royal Society of London B*, **358**, 435-445.

THORNTON, I. M., & VUONG, Q. C. (2004). Incidental processing of biological motion. *Current Biology*, **14**, 1084-1089.

TJAN, B. S. (1996). *Ideal observer analysis of object recognition*. Unpublished doctoral dissertation, University of Minnesota, Minneapolis.

TJAN, B. S., BRAJE, W. L., LEGGE, G. E., & KERSTEN, D. (1995). Human efficiency for recognizing 3-D objects in luminance noise. *Vision Research*, **35**, 3053-3069.

APPENDIX

The ideal observer for our task and stimuli can be derived using Bayes's rule (Geisler, 2003; Green & Swets, 1966; Tjan et al., 1995). Observers are asked to choose the signal, S_i (where i refers to the ith of N possible signals that could be shown), that most likely appeared within the noisy stimulus data, D. According to Bayes's rule, the a posteriori probability of S_i having been presented given D can be expressed as

$$P(S_i \mid D) = P(S_i) P(D \mid S_i) / P(D). \tag{A1}$$

For our task and stimuli, $P(S_i)$ and $P(D)$ are scaling constants and can be removed. Thus, the ideal observer chooses the signal that maximizes $P(D \mid S_i)$. For the case containing random spatial jitter, the ideal observer must compute this probability for all m possible spatially shifted versions of each signal (all of which are equally probable), resulting in the following probability function:

$$P(D \mid S_i) = \sum_{k=1}^{m} \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{\sqrt{2\sigma^2}}(D_j - S_{ijk})^2}, \tag{A2}$$

where n is the total number of pixels in the stimulus and σ is the standard deviation of the Gaussian distribution from which the external noise was generated. The ideal decision rule is to choose the S_i that maximizes this function. For the case containing no spatial jitter ($m = 1$), the summation falls out, and it can be shown (Eckstein & Whiting, 1996; Tjan et al., 1995) that the ideal decision rule reduces to choosing the signal that produces the higher cross correlation with the noisy stimulus—that is, $\max(D \otimes S_i)$. Notice that, for the ideal observer, the computation that it performs for a spatiotemporal pattern is no different from what it performs for a purely spatial pattern. The inclusion of successive frames in a stimulus is formally equivalent to creating one large spatial image that comprises the concatenation of all the frames in the movie (or even a single vector containing all of the pixels in the movie). As long as the correspondence between the pixels in the ideal observer's templates and the configuration of the pixels in the stimulus remains intact, reorganizing the pixels in any fashion will have no effect on ideal performance.

We used these decision rules with Monte Carlo simulations to estimate the ideal observer's threshold in each of our experimental conditions. The same adaptive threshold-measurement procedures used with the human observers were also used with the ideal observer simulations. A minimum of 1,000 simulated trials were used to estimate thresholds in each experimental condition.