Contrast gain control determines global form percept in tripole Glass patterns

Yih-Shiuan Lin
Department of Psychology, National Taiwan University, Taipei, Taiwan

Pao-Chou Cho
Department of Psychology, National Taiwan University, Taipei, Taiwan

Chien-Chung Chen
Department of Psychology, National Taiwan University, Taipei, Taiwan
Neurobiology and Cognitive Science Center, National Taiwan University, Taipei, Taiwan

A Glass pattern consists of randomly distributed dot pairs, or dipoles, whose orientation is determined by a geometric transform that defines the global percept for this pattern. The perception of Glass patterns involves a local process to associate paired dots into dipoles and a global process to group the dipoles into a global structure. We used a variant of Glass patterns consisting of tripoles instead of dipoles to estimate the effect of luminance contrast on the global form percept. In each tripole, an anchor dot and two context dots formed the vertices of an equilateral triangle with the anchor dot pointing toward the center of the display. Grouping the anchor dot with one context dot would result in a global percept of a clockwise (CW) spiral and grouping with the other dot a counterclockwise (CCW) spiral. We manipulated the contrast of the context dots and measured the probability of a participant judging the patterns as a CW spiral. The CW spiral judging probability first increased then decreased with the contrast of the CW context dots, resulting in an inverted U shape. The peak also shifted to the right as the contrast of the competing CCW context dots increased. Our result cannot be explained by the existing models for Glass pattern perception. Instead, the data was well fit by a divisive inhibition model, in which the response of a global pattern is the excitation raised by a power and divided by the inhibition from all global patterns plus an additive constant.

Introduction

The primary function of the biological visual system is to identify objects in an environment. To achieve this, given that in the earlier stages of visual processing an input image would be decomposed into fragmented components by neural mechanisms with localized receptive fields and specific tuning characteristics (Hubel & Wiesel, 1962, 1968), the visual system must be able to integrate image components into the percept of a coherent object. The purpose of this study is to provide a quantitative understanding of this perceptual grouping process by investigating how an observer perceives the global form in a Glass pattern. A Glass pattern (Glass, 1969; Glass & Perez, 1973) consists of randomly distributed dot pairs, or dipoles. The orientation of dipoles conforms to a predesignated geometric transform that allows an observer to perceive a global form. The merit of Glass patterns is that, to perceive the global form, an observer needs to employ at least two stages of grouping. The first is the local grouping, in which two dots are grouped to form a dipole (Burr & Ross, 2006; Chen, 2006; Dakin, 1997; Dakin & Bex, 2001; Earle, 1999; Mandelli & Kiper, 2005; Smith, Bair, & Movshon, 2002; Smith, Kohn, & Movshon, 2007; Stevens, 1978; J. A. Wilson & Switkes, 2005; J. A. Wilson, Switkes, & De Valois, 2004). The second is the global grouping, in which the dipoles are integrated into a global structure (Cardinal & Kiper, 2003; H. R. Wilson & Wilkinson, 1998; H. R. Wilson, Wilkinson, 1998).
& Asaad, 1997). Kurki, Laurinen, Peromaa, and Saarinen (2003) suggested that different neural mechanisms are behind these two grouping processes because the detection threshold has been found to be higher for local features than for global structure.

Two theories have been proposed to explain how the visual system groups local elements into a dipole in Glass patterns. The similarity theory, called the token-matching theory by some (Marr, 2010; Stevens, 1978), suggests that the grouping strength of a dipole is determined by the similarity between the dipole elements. Stevens (1978) showed that a low-intensity dot was more likely to be paired with another low-intensity dot than with a high-intensity dot to construct a global pattern. Earle (1999) reported similar results in dot contrast. J. A. Wilson et al. (2004) also found that the probability of grouping two dots in a dipole decreased with the contrast difference between them.

Another theory suggested that grouping strength depends on the energy of the local elements in a dipole. Prazdny (1984) created Glass patterns by superimposing three sets of dot patterns with different dot features and discovered that the final percept of orientation was determined by pairings between dots with higher feature energy, which was defined as the product of dot size and contrast. That is, dissimilar dots can be grouped as long as the grouped pair contains higher energy. Prazdny then suggested that it was higher energetic features rather than similarity that decided the perceptual grouping of Glass patterns. This theory may have a physiological base. It is reported that the V1 and V2 neurons of macaques respond to the dipoles of Glass patterns (Smith et al., 2002; Smith et al., 2007). Because the response of early visual cortical neurons increases with luminance contrast (Albrecht & Hamilton, 1982; Boynton, Engel, Glover, & Heeger, 1996; Goodyear & Menon, 1998; Sclar, Maunsell, & Lennie, 1990), it is reasonable to infer that feature energy plays a role in determining the final percept of Glass patterns.

These two conflicting results may be due to the limitation of the experiments used in the studies. Stevens (1978) always kept two of the three superimposed dot sets at low intensity and the third at high intensity. Earle (1999) also assigned only two contrast levels (high or low) to each dot in a Glass pattern. Thus, they could not capture subtle quantitative change in grouping performance. Similarly, Prazdny (1984) also suffered from limited contrast level conditions. In J. A. Wilson et al. (2004), the task of the participants was to detect a Glass pattern embedded in noise. Although they applied a wide range of contrast levels, there was no meaningful competition for the target. Thus, they failed to reveal an interaction between similarity and energy cues. This, as shown below, turns out to be essential for determining the grouping mechanisms underlying the perception of Glass patterns. It is possible that each theory only captures part of the truth of these grouping mechanisms.

We thus used a variant Glass pattern in the current study. Instead of using dipoles, our stimuli contained three-dot sets as local units, which we called a tripole Glass pattern (tGP). In this configuration, there are multiple ways for an observer to group dots together, so there could be more than one way to perceive the global form. By systematically changing the contrasts among the three dots, we can control the strength of either similarity or energy cues and even pitch them against each other. We can then observe how grouping performance changes with the strength of either cue. By this manipulation, we are able to provide a quantitative estimation of either effect and thus a more comprehensive picture of the grouping mechanisms.

### Method

#### Observers

Three observers (one female and two males, between 19 and 30 years old) participated in this experiment, including one of the authors (PCC) and two other observers naïve to the purpose of this study. All the observers had normal (20/20) or corrected-to-normal vision. The study was approved by the Research Ethics Committee of National Taiwan University. Written consent was obtained from each observer before the experiment.

#### Apparatus

The stimuli were presented at the center of a 19-in. ViewSonic G90fB CRT monitor with a width of 36.5 cm, height of 27.5 cm, and resolution of 800 (H) × 600 (V). The refresh rate was 75 Hz. The gamma function of the CRT monitor was measured and calibrated with a Photo Research PR-655 radiometer. The mean luminance was 48 cd/m². The viewing distance between the eye position of the observers and the monitor was set at 75 cm so that each pixel occupied 2 min of one visual angle.

The experiment was conducted in a dark room. A chin rest was used to reduce head movement. The experimental program and the generation of stimuli were written in MATLAB with the Psychtoolbox (Brainard, 1997).
Stimuli

Differing from the traditional Glass pattern, which consists of dot pairs, or dipoles, a variant Glass pattern composed of three-dot sets (tripoles), as shown in Figure 1, was used in this study. To generate a tGP, we first drew randomly distributed square dots (10° × 10° visual angle) in the display as anchor dots (illustrated as the square denoted “A”). Context dots (clockwise [CW] and counterclockwise [CCW] dots depicted by squares with “CW” and “CCW” in Figure 1), which were the same size as the anchor dot, were placed on either side of the radial line passing through the anchor dot. The line linking the anchor and the context dot intercepted the radial line at an angle of $\pi/6$, or 30°. Thus, three dots in a tripole were at the vertices of an equilateral triangle with the anchor vertex pointing toward the center of the display. The distance between two dot centers in one tripole was 28°. Linking anchor dots with the corresponding CW dots (green oval in Figure 1) produced a percept of a CW Archimedean spiral, and linking anchor dots with corresponding CCW dots (the red oval in Figure 1) produced a percept of a CCW spiral.

The dot density was set in such a way that all dots occupied 4% of the surface area in one tGP. At this relatively low density, it is unlikely that the tripoles would overlap each other. The overall size of the global Glass pattern was about 15°.

It is known that the visual system cannot group two dots of opposite luminance polarities into a dipole (Glass & Switkes, 1976; J. A. Wilson et al., 2004). Therefore, in our study, dots in one tGP were all of either positive polarity (luminance increment) or negative polarity (luminance decrement) relative to the background, which was assigned mean luminance. The luminance Weber contrast ($c$) was defined as the ratio between dot luminance increment (or decrement, depending on polarity), $|\Delta L|$, and background luminance level, $L_0$, i.e., $c = |\Delta L|/L_0$. The Weber contrast was then converted to dB units by the operation $20 \times \log_{10}(c)$. For each polarity, there were seven contrast levels for each context dot (CW and CCW dots) in the range between $-30$ dB (3%) and $-1$ dB (90%), and the anchor dot contrast remained $-20$ dB (10%).

Figure 2 shows examples of tGPs with negative polarity. The five tGPs illustrated here have the CW dot contrast increasing from the left panel to the right while the CCW dot contrast increases from the right to the left. For a more intuitive illustration of the local contrast effect on the global percept, the contrast of the anchor and context dots are selected in a way in which it is easier to perceive a CW spiral on the right side and a CCW spiral on the left.

Procedure

In each trial, a beep signaled the start of a new trial, and then a fixation cross flashed for 400 ms. After a 200-ms blank, the tGP was presented for 167 ms. The task of the observers was to press a key to indicate whether the global form was a CW or a CCW spiral. The next trial started 400 ms after response.

For each anchor dot (positive or negative polarity), there were seven levels of contrast with each context dot. For each run, two context dot contrasts were used. In each trial, one of the two contrasts was randomly assigned to the CW dots and the other was assigned to
the CCW dots. There were 60 trials in each run. Each context dot contrast pair run was repeated four times, resulting in 240 trials for each pair. The order of runs was randomized.

Results

The results of the three observers are shown in Figure 3. In each panel, the x-axis indicates the contrast level of the CW dots; the y-axis indicates the probability of the observer judging the Glass pattern as a CW spiral. Different icons denote the data collected at different CCW contrasts. The smooth curves are the fits of our model discussed below in the Model section. Each data point is the average of four measurements. The error bar represents the standard error of measurements in each condition.

At all CCW dot contrasts, the probability of CW judgment first increased with CW dot contrast until a critical point; then it decreased as CW dot contrast further increased, resulting in an inverted U shape. This trend was most obvious when the CCW dots had the lowest contrast (−30 dB), shown as the brightest curve at the top of each data panel in Figure 3. This inverted U-shape function was significant: The probability of CW judgment at −30 dB and −1 dB CW dot contrasts was significantly lower than −15 dB across all subjects and anchor dot polarities, \( t(3) \) ranged from 3.59 to 30.09, \( p < 0.05 \), except the difference between −15 dB and −1 dB of the negative anchor polarity condition of observer YCJ, \( t(3) = 2.34, p = 0.051 \). This effect was less apparent as the CCW dot contrast increased.

The CCW dot contrast had two effects. As CCW dot contrast increased, it decreased the peak probability and shifted the peak position rightward. These two effects are indicated by arrows in Figure 3 on the middle upper panel. As a reference, the dashed line in each panel indicates the contrast of the anchor dot and thus the conditions in which the anchor dot and CW dot were at the same contrast level. It is obvious that the peak probability of a CW judgment deviated more and more from this dashed line as the CCW dot contrast increased.

Our observers showed no bias in response toward either CW or CCW spirals. When the contrasts of both context dots were identical, the probability of reporting CW was 0.5106 to 0.5463 across observers, which were not statistically significant, \( t(13) \) ranged from 0.0803 to 0.8671, \( p = 0.46 \) to 0.20.

Discussion

In our experiment, we measured the probability of perceiving a global form produced by grouping the anchor dot with one of the context dots in the presence of another context dot in a tGP. We found that the probability of grouping the anchor dot and a context dot first increased and then decreased with the contrast of that context dot, resulting in an inverted U function. Furthermore, as the contrast of the second context dot increased, it not only decreased the peak of the U-shape function but also shifted the peak rightward. Our results cannot be explained by current theories of Glass pattern perception.

Prazdny (1984) argued that, in a Glass pattern, an observer tended to pair dots with greater energy, which
was defined as the product of dot size and dot contrast, regardless of contrast polarity. In our experiment, dot size was a constant. Hence, according to the energy model, our observers should group the anchor dots with the context dot with greater contrast. Thus, when the contrast of a context dot increases, the observer is more likely to group this dot with the anchor dot and, in turn, more likely to perceive the global form signaled by this context dot. For instance, as shown in Figure 4a, the probability of seeing CW patterns should increase monotonically with CW dot contrast. However, in our data, such monotonic increase only occurred when the CW dot contrast was low (less than \(-15\) dB). When the CW dot contrast reached a certain level, the probability of perceiving a CW pattern actually decreased with further increase in CW dot contrast. This inverted U-shape function cannot be explained by the energy model.

Figure 3. The probability of judging a tGP as a CW spiral across different CCW and CW dot contrasts in three participants. The symbols and curves are color-coded to represent the data and fits of the model (see Discussion) at various CCW contrasts. Dashed lines indicate the contrast level of the anchor dot (\(-20\) dB). Upper panels contain data from the positive polarity (bright) condition, lower panels the negative polarity (dark) condition. Black arrows on the upper middle panel (positive polarity data of observer PCC) point out the peak data points for each CCW dot contrast.

Figure 4. Predicted CW judging probability across different CCW dot contrasts based on previous theories. (a) The prediction of the energy model. The dashed line indicates the anchor dot contrast. (b) That of the similarity theory.
The similarity theory suggests that the likelihood of grouping two dots into a dipole increases when the similarity between the dots increases (Earle, 1999; Stevens, 1978; J. A. Wilson et al., 2004). In the context of our experiment, this theory would predict that the anchor dot is more likely to be grouped with the context dot with the smallest contrast difference. That is, as shown in Figure 4b, the similarity theory would predict that regardless of the contrast level of the competing dot (e.g., CCW dot), the probability of seeing a CW pattern should always peak when the anchor dot and CW dot have the same contrast (marked by the dashed line) and decrease when the CW dot contrast deviates from that of the anchor dot. Thus, the probability of seeing CW patterns should be an inverted U function of CW dot contrast peaking at the anchor contrast regardless of the CCW dot contrast. Our data did show an inverted U shape. However, the peak position shifted to the right as CCW dot contrast increased. That is, in most conditions, the best grouping between the anchor dot and the CW dot occurred when their contrasts were quite different. This is inconsistent with the prediction of the similarity theory.

**Divisive inhibition model**

Our result cannot be explained by either the energy or similarity theory. Here, we modified a divisive inhibition model, proposed by Chen (2009), to explain the masking effect between Glass patterns to account for our data.

This model (shown in Figure 5) has several stages. The first stage contains local linear processors that extract local dipole information (H. R. Wilson & Wilkinson, 1998; H. R. Wilson et al., 1997). Second, a global linear template integrates information from dipoles conforming to the global shape of that template and contributes the outcome as the excitation component of the global form. The response of a global form detector is the excitation of its corresponding linear template raised by a power and divided by the inhibition from all templates plus an additive constant (Chen & Foley, 2004; Chen, Foley, & Brainard, 2000; Foley, 1994). Finally, the likelihood of an observer perceiving one pattern over the other is determined by the difference in response between the two global pattern detectors.

As shown in previous studies (Smith et al., 2002; Smith et al., 2007), a dipole in a Glass pattern can be picked up by an orientation-selective V1/V2 neuron. Thus, we assume that the first stage of Glass pattern processing contains a band of orientation-selective filters whose excitation is determined by the product of their receptive fields and the spatial profile of the dipole image. That is,

\[ Ed_j = \int C_1 I_1(x, y) \times f_j(x, y) \, dx \, dy + \int C_2 I_2(x, y) \times f_j(x, y) \, dx \, dy, \]

where function \( f_j(x, y) \) is the receptive field of the \( j \)th filter, which orientation conforms to the \( j \)th global form in one Glass pattern; \( I_1(x, y) \) and \( I_2(x, y) \) are the spatial profiles; and \( C_1 \) and \( C_2 \) are the contrast of the two dots in the dipole, respectively. For simplicity, we assume that the receptive field centers are halfway between the two dots, and thus, the effects from both dots are symmetrical. Also, because we used identical square dots, the spatial profiles \( I_1 \) and \( I_2 \) are identical. Hence, Equation 1 can be simplified as

\[ Ed_j = (C_1 + C_2) \int I(x, y) \times f_j(x, y) \, dx \, dy = A_j \times (C_1 + C_2) = A_j \times C_d_j, \]

where \( C_d_j \) is the pooled contrast for the \( j \)th dipole. Here we use constant \( A_j \) to replace all factors not affected by contrast manipulation. In our tGP, there are three dots and thus three possible dipoles—the CW, CCW, and irrelevant (for decision making) dipoles (shown in Figure 5, left panel)—in one tripole, which globally produce three global forms: a CW spiral, a CCW spiral, and an irrelevant pattern. (The third pattern is deemed irrelevant because observers were not asked to respond...
to this global form.) Therefore, j can refer to any one of these three forms. All other arrangements would either produce a nonoptimal response or unsppecific orientation information.

A linear global template sums the excitation of relevant local filters together. Suppose that there are n dipoles contributing to a global template: The excitation of jth form, which is the sum of the excitation of all relevant dipoles, is

\[ E_j = n \times Ed_j = nA_j \times Cd_j = Se_j \times Cd_j. \]  

We define nAj as the excitation sensitivity to the jth global form and replace it with parameter Sej.

The inhibition toward the jth global form, Ij, is defined in a similar fashion as

\[ I_j = \sum_{k=1}^{n} \left( Si_k \times Cdk \right)^q, \]  

where q is the power parameter and Si k the inhibition sensitivity of the kth template among all three aforementioned global forms. The response of the jth global form is excitation, E, raised by a power, p, divided by the sum of the divisive inhibition, Ij, and an additive constant, z, as

\[ R_j = \frac{E_j^p}{I_j + z}. \]

The final percept of one tGP is determined by the response difference between multiple global forms. Given our two-alternative, forced-choice paradigm, we only need to consider the responses of two global form operators, the CW and CCW spirals. The response difference between them is

\[ D = R_{CW} - R_{CCW}. \]

Assume that the noise in each channel of each trial is independently and identically distributed in Gaussian; the probability of classifying an image as CW is a Gaussian cumulative distribution function with mean zero and variance one. Note that any nonstandard Gaussian parameters can be absorbed by a change in the value of other model parameters and thus are not needed here.

To implement the model, in our experiment, the observers were to judge whether the tGP was a CW or a CCW spiral. Thus Equation 5 becomes

\[ R_{cw} = \frac{E_{cw}^p}{I_{cw} + z} \quad \text{and} \quad R_{ccw} = \frac{E_{ccw}^p}{I_{ccw} + z} \]

in which the excitation (Equation 3) becomes \( E_{cw} = Se \times Cd_{cw} \) and \( E_{ccw} = Se \times Cd_{ccw} \), and inhibition (Equation 4) becomes \( I_{cw} = (Si_1 \times Cd_{cw})^q + (Si_2 \times Cd_{ccw})^q \), and \( I_{ccw} = (Si_1 \times Cd_{cw})^q + (Si_2 \times Cd_{ccw})^q \). Si1 and Si2 are the inhibition sensitivities from the other two competing global form templates. Because there was no CW judging bias in the data, we used the same set of sensitivity parameters, Se, Si1, and Si2, in our model for CW and CCW responses. Cd_{cw}, Cd_{ccw}, and Cd_{ire} correspond to the pooled contrast of the CW dipole, CCW dipole, and irrelevant dipole, respectively. We empirically found that removing the irrelevant global form template did not affect the goodness of fit; thus, the parameter Si2 was set as zero in the final fitting results.

Table 1 shows the fitting parameters and goodness of fit (R^2) for all observers. Except for Si1, Si2, q, and z, all other parameters were fixed for R^2 and did not differ empirically whether we set them free or not. The value of parameter p was set to be 1 because we empirically found that fixing it at 1.00 did not affect the goodness of fit. Overall, our model explained 87% to 97% of the variance in the data with rooted mean square errors ranging from 0.0458 to 0.1034 and mean standard error from 0.0346 to 0.0583 across all observers.

Notice that, without the inhibition terms, the response of a global form channel (Equation 5), say, the CW channel, reduces to \( (E_{cw}^p)/z \), which is a monotonic increasing function of CW dipole contrast. Thus, it becomes a version of the energy model. We actually implemented the energy model, as shown in Figure 4a, by fixing parameter q as 0 and p as 2 in our model.

Divisive inhibition is needed to explain the inverted U shape. Let us, again, use the CW channel as an example and leave out the CCW channel for now.
When CW dot contrast is low, the inhibition from the CW channel itself is small compared with the additive constant $z$. Thus, the response can be approximated by $(E_{CW}/z)$, which is a monotonic increasing function of CW contrast. This explains the rising part of the inverted U function. As CW dot contrast increases, the inhibition term becomes larger. Because the exponent parameter for inhibition, $q$, is larger than that for excitation (Table 1), the inhibition quickly catches up with the excitation and drives the response down. J. A. Wilson et al. (2004) proposed a similar theory to explain why it is easier for an observer to integrate dots of the same contrast into a dipole. Such self-inhibition, with inverted U prediction, provides a way to implement the similarity model.

However, without a divisive inhibition from other global form channels, one cannot explain the shift of peak positions in the inverted U functions. Suppose that the CCW dots have no influence on the response of the CW channel, $R_{cw} = E_{cw} / (I_{cw} + z)$, $I_{cw} = (S_{cw} \times C_{cw})^q$, but just affect the decision stage (Equation 6), in which the CCW dots only change the subtractive constant and thus have the same effect for all CW dot contrasts. That is, different CCW dot contrasts would only shift the inverted U function up and down rather than causing a lateral shift. This effect can be implemented by fixing parameter $S_{cw}$ as 0 and is visualized in Figure 4b.

Continuing with our example, with the inhibition term from the other channels at the same CW dot contrast, the relative contribution from the CW template in the denominator of the response function becomes smaller. That is, it requires a greater CW dot contrast to dominate the denominator and drive the response down. The CW dot contrast needed increases with the CCW contrast. As a result, the peak position of the inverted U function shifts rightward toward greater CW dot contrast as CCW dot contrast increases.

In this study, we applied the nonlinear response function after the summation of local dipoles by the global template, that is, an early summation model. We acknowledge that this nonlinear response may well occur between local dipole processors and the global template. This would result in a late summation model. One example of this kind of model is the linear-nonlinear (LN) model proposed by H. R. Wilson et al. (1997). (Notice that the nonlinearity in their model is provided by rectification, which is a monotonic function when all elements have the same luminance polarity. Thus the prediction of their model will be similar to that of the energy model.) In our case, the response of the late summation can be implemented by removing the constant $n$ from Equation 4 and then multiplying it by the response function in Equation 7. Either operation can be readily absorbed by a change in the values of other parameters. In the context of our current study, the two models are mathematically equivalent given the assumptions we made for the current experiment. We chose the model in the current study simply to be consistent with Chen (2009). A future experiment may be able to distinguish these two models. For instance, suppose one repeated our experiment while manipulating the number of tripoles in the patterns. For the late summation model, the effect of the number of tripoles on the decision variable (Equation 6) would be linear, and for the early summation model, its effect would be exaggerated by the nonlinearity.

### Conclusion

In the current study, we manipulated the dot contrasts of tGPs and recorded the probability of an observer judging a tGP as a CW or a CCW spiral. We found that, as the contrast of the CW dots increased, the probability of a CW spiral judgment first increased and then decreased as the contrast passed a certain level, resulting in an inverted U shape, suggesting self-inhibition during the process. The peak CW judging probability decreased and shifted to higher CW dot contrast when the CCW dot contrast increased, suggesting inhibition from other global forms presented simultaneously in one tGP and showing a competition mechanism between different forms.

Neither the energy model, which emphasizes the impact of feature energy on the grouping percept, nor the similarity theory, which stresses that similarity determines the final grouping, can fully explain all aspects of our results. We thus applied a divisive normalization model to fit the data. In our model, the response to one global form is determined by excitation gain control provided by a power and divided by the sum of divisive inhibition and an additive constant. The self-inhibition component in our model can explain the inverted U shape; the inhibition from other global form templates can explain the shifting of the peak CW judging probability with the increase in CCW dot contrast.

**Keywords:** divisive inhibition, luminance contrast, contrast gain control, global form perception

### Acknowledgments

This study is supported by Ministry of Science and Technology (MOST 103-2410-H-002-076-MY3) to CCC.

Commercial relationships: none.

Corresponding author: Chien-Chung Chen.
References


