# An Oscillatory Correlation Model of Human Motion Perception

**Erdoğan Çeşmeli**
General Electric
Corp. Res. and Dev.
Niskayuna, NY 12309
*cesmeli@cis.ohio-state.edu*

**Delwin T. Lindsey**
Dept. Psychology
The Ohio State University
Columbus, OH 43210
*lindsey.43@osu.edu*

**DeLiang L. Wang**
Dept. Comp. Info. Sci.
The Ohio State University
Columbus, OH 43210
*dwang@cis.ohio-state.edu*

## Abstract

An oscillatory correlation model of human motion perception is proposed based on the integration of motion and luminance information. The model is composed of two parallel pathways that segment the input scene based on motion and luminance, respectively. Combining these segmentations, the model refines the motion estimates in the integration stage to obtain the final segmentation in the motion pathway. For segmentation, LEGION (Locally Excitatory Globally Inhibitory Oscillator Networks) is employed whereby the phases of oscillators are used for region labeling. The model performance is demonstrated using a set of psychophysical data.

## 1 Introduction

A central problem to computational investigation of motion perception is the selective integration of local motion estimates. Most approaches assume that motion integration can be addressed using only motion information. However, this assumption has been questioned by recent studies, e.g. [4, 1], using *plaids*, which are constructed by the superposition of two moving gratings at different orientations. When presented, observers report either a coherent motion corresponding to that of the plaid or a pair of component (non-coherent) motions belonging to the gratings. Without changing the underlying motion, when the luminance at the intersection of the gratings is varied so that gratings appear to be on two different surfaces, a non-coherent motion is more frequently perceived. Consistent with this observation, other studies, (see [4]), also indicate that non-motion cues, e.g. stereo, play a role in motion integration. Similarly, computational studies demonstrate that the inclusion of luminance information improves motion-based segmentation.

Motivated by these studies, we propose a motion perception model based on oscillatory correlation. Our model consists of two parallel pathways for motion and luminance, respectively. The motion pathway has two stages. First, local motions are estimated by an adaptive temporal block matcher, a variation of the Reichardt detector [3]. In the second stage, locations are grouped based on motion similarity in a multilayer LEGION network [6]. In order to complement the initial motion segmentation, a stationary scene analysis is performed in the parallel luminance pathway also using a LEGION network. Next, the integration stage refines motion estimates based on which the final segmentation is obtained in the motion network.

The following two sections describe our model. Next is the demonstration of its performance using synthetic and psychophysical data. Finally, conclusions are drawn.

## 2 Background

### 2.1 Temporal Block Matcher (TBM)

TBM compares luminance within a block, $N_B$, instead of a single location [3]. Selectivity to different velocities is achieved by sliding correlation blocks in different quantities and different directions on the previous snapshot (frame) of the scene. The correlation corresponding to displacement $\mathbf{r} = (r_x, r_y)$, at location $(x, y)$ and time $t$ can be expressed as:

$$\tilde{v}_{\mathbf{r}}(x, y, t) = \sum_{(j,k) \in N_B(x,y)} \frac{I(j, k, t)}{|I(j, k, t) - I(j - r_x, k - r_y, t - 1)|} \tag{1}$$

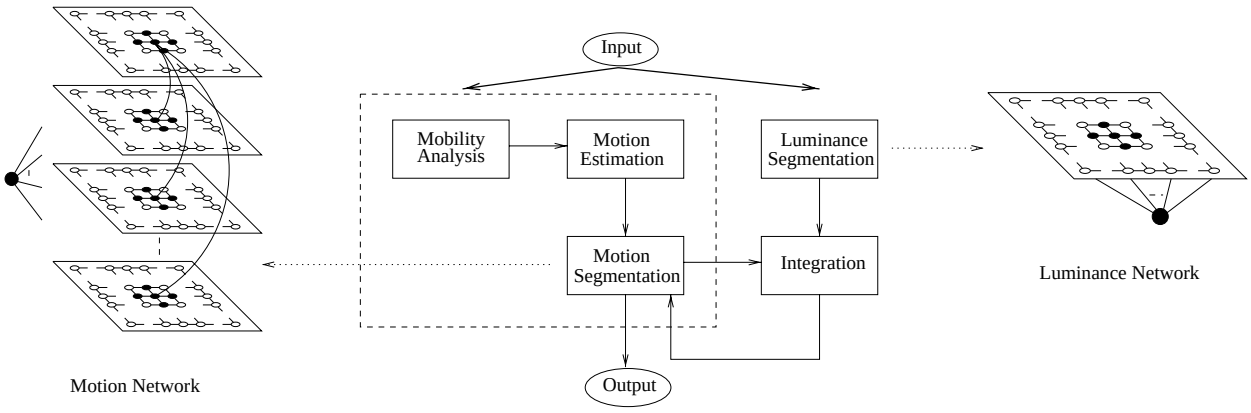


Figure 1: The flow diagram of the proposed model. Processing starts from the top and proceeds downward. Also shown are the multilayer motion and the luminance networks. Ellipses and circles represent oscillators and inhibitors, respectively.

where $I(x, y, t)$ is the luminance at location $(x, y)$ in the image frame $t$ and $N_B$ is centered at location $(x, y)$. Note that a large $\tilde{v}_{\mathbf{r}}$ implies a high probability of $\mathbf{r}$.

### 2.2 LEGION

As an alternative to the grandmother cell representation, regions can be represented efficiently by dynamic linking of corresponding features. Oscillatory correlation, the idea behind LEGION, is one such representation where feature binding is achieved based on collective activities of multiple oscillators [6, 5]. LEGION is composed of interacting relaxation oscillators, each corresponding to one location in the scene. Their interaction includes a variable called lateral potential and coupling from neighboring oscillators, e.g. four nearest neighbors, and a global inhibitor. The potential is introduced for each oscillator to distinguish a homogeneous region from a noisy one. Initially high potentials continuously decay. Unlike noisy areas, homogeneous regions support *leaders*, oscillators that can keep their potentials high. Leaders inititate segments but other oscillators, *followers*, cannot. The global inhibitor is on only when an oscillator is active. Its inhibition ensures that only a single segment becomes active at a time. A critical design issue in LEGION is the definition of the local coupling weight, $W_{ik}$, between two oscillators, $i$ and $k$. When $i$ and $k$ correspond to similar features, e.g. luminance, $W_{ik}$ is assigned a larger value when they do not.

## 3 Adaptive TBM and LEGION

Our model is shown in Figure 1. The first pathway (dashed-box) performs motion analysis. The second pathway segments the scene based on luminance. The two segmentation results are combined in the integration stage to refine the motion estimates. The final segmentation is performed in the motion network based on the refined estimates.

### 3.1 Motion Pathway

**Motion Estimation:** Inspired by the visual system (see [2]), the motion pathway has two stages for motion estimation and segmentation, respectively. For estimation, first, a mobility value is calculated at each location by quantifying temporal luminance changes at each location. A large mobility value at a location indicates a high probability of motion at that location. Subsequently, at each location, a correlation block, $N_B$, is selected only when it contains a large sum of mobility values and an average mobility without having its size reaching an upper limit.

Having selected $N_B$ at reliable locations, we apply TBM to three consecutive frames. The correlation corresponding to displacement $\mathbf{r}$ is given by $v_{\mathbf{r}}(x, y, t) = \tilde{v}_{\mathbf{r}}(x, y, t) + \tilde{v}_{\mathbf{r}}(x, y, t + 1)$ where $\tilde{v}_{\mathbf{r}}(x, y, t)$ and $\tilde{v}_{\mathbf{r}}(x, y, t + 1)$ are the correlations at $(x, y)$ between the image frames at $t$ and $t - 1$ and those at $t + 1$ and $t$, as given in (1). We also match local spatial correlation surfaces (SCSs). A SCS is obtained by correlating

$N_B$ at locations within a neighborhood formed in the same frame. Two SCSs are matched where blocks of luminance are substituted by SCSs, resulting in $c_{\mathbf{r}}(x, y, t)$. Hence, the temporal correlation for displacement $\mathbf{r}$ is $\hat{V}_{\mathbf{r}}(x, y, t) = v_{\mathbf{r}}(x, y, t) + c_{\mathbf{r}}(x, y, t)$.

In the aperture problem [2], among multiple consistent local estimates, the smallest velocity is the perceptually preferred one. In order to capture this preference, we multiply the cross-correlation surface with a 2D Gaussian surface, $G$, centered at zero velocity (origin) [7]: $V_{\mathbf{r}} = G\hat{V}_{\mathbf{r}}$. Due to the shape of $G$, correlations for large displacements are suppressed more and thus, the smallest one is favored.

We determine a certainty measure for each estimate, $\mathbf{e}$, based on the shape of its correlation surface. We define a 2D coordinate system centered at $\mathbf{e}$ which corresponds to the maximum correlation, $V_{\mathbf{e}}$. The *speed axis* passes through the origin and $\mathbf{e}$, and is perpendicular to the *direction axis* at the latter. Dropping $(x, y, t)$ from the expressions for convenience, the certainty, $\omega$, of the estimate $\mathbf{e}$ is:

$$\omega_{\mathbf{e}} = \frac{(V_{\mathbf{e}} - V_{\mathbf{e}-\mathbf{k}})(V_{\mathbf{e}} - V_{\mathbf{e}+\mathbf{k}})}{2V_{\mathbf{e}} - V_{\mathbf{e}-\mathbf{k}} - V_{\mathbf{e}+\mathbf{k}}} \tag{2}$$

Here, $\mathbf{e} - \mathbf{k}$ and $\mathbf{e} + \mathbf{k}$ are the nearest neighboring displacements to $\mathbf{e}$ along the direction axis, with correlations $V_{\mathbf{e}-\mathbf{k}}$ and $V_{\mathbf{e}+\mathbf{k}}$, respectively. According to (2), the larger $\omega_{\mathbf{e}}$ is, the more certain $\mathbf{e}$ is assumed to be.

**Network:** To represent motion transparency, our model adopts a multilayer LEGION network where each layer corresponds to one velocity as depicted in Figure 1. The coupling weight between oscillators $i$ and $k$ on a velocity layer $\mathbf{r}$ is determined based on their temporal correlation: $W_{\mathbf{r},ik} = (V_{\mathbf{r}}(i) + V_{\mathbf{r}}(k))/|V_{\mathbf{r}}(i) - V_{\mathbf{r}}(k)|$. When oscillators have similar correlations for $\mathbf{r}$, they are strongly coupled in the velocity layer of $\mathbf{r}$. At locations without estimates, couplings are set to zero. The interaction is determined to be strong when it is larger than the threshold $\theta_{M,1}$. For motion transparency, the definition for potential is constrained to only allow oscillators with large mobility values to become leaders and a velocity column to have at most one leader. A leader in a column recruits oscillators on its layer through local couplings. Other oscillators in the same column become active only when they are recruited by leaders on their layers. Since multiple oscillators can become active within a column, multiple motions can be represented at a single location.

### 3.2 Luminance Pathway

In the parallel luminance pathway, the middle frame of the sequence is analyzed using a LEGION network depicted in Figure 1. The coupling weight between oscillators $i$ and $k$ is defined as $W_{ik} = (I(i) + I(k))/|I(i) - I(k)|$. When locations have relatively similar luminance, $I$, a strong coupling weight results. Oscillators corresponding to textured regions tend not to have strong couplings and thus, do not form segments. The interaction is assumed to be strong when it is larger than the threshold $\theta_B$.

### 3.3 Integration Stage

The first step of the integration stage initially considers motion segments. When all locations in a motion segment belong to a textured region, estimates in this segment are not changed. However, when all locations are from multiple untextured regions, an occlusion relationship among these regions is obtained by detecting T- and X-junctions. When a segment includes both textured and untextured regions, their motion distributions are compared to determine whether they move together.

In the second step of the integration stage, estimates along an occluding boundary are eliminated in occluded regions. The remaining ones within each luminance segment, $B$, interact iteratively and result in a segment velocity, $\mathbf{r}_B$, given by:

$$\mathbf{r}_B^{\tau} = \frac{1}{\Omega_B^{\tau}} \sum_{(x,y)\in B} \omega^{\tau}(x, y)\mathbf{r}(x, y) \tag{3a}$$

$$\omega^{\tau+1}(x, y) = \frac{\omega^{\tau}(x, y)}{\Omega_B^{\tau}} \Big(1 + \frac{\mathbf{r}(x, y) \cdot \mathbf{r}_B^{\tau}}{\sqrt{||\mathbf{r}(x, y)||\,||\mathbf{r}_B^{\tau}||}}\Big) \tag{3b}$$

Here $\mathbf{r}_B^{\tau}$ and $\Omega_B^{\tau}$ are the segment velocity and the sum of certainties, $\omega^{\tau}(x, y)$, in $B$ at iteration step $\tau$, respectively. $\mathbf{a} \cdot \mathbf{b}$ is the dot product between vectors $\mathbf{a}$ and $\mathbf{b}$, and $||\mathbf{a}||$ is the magnitude of $\mathbf{a}$. In (3a), $\mathbf{r}_B^{\tau}$ is determined by weighing the estimates in $B$ by their certainties. In (3b), $\omega^{\tau+1}(x, y)$ increases when $\mathbf{r}(x, y)$ and $\mathbf{r}_B^{\tau}$ have a similar direction. Finally, when $\omega^{\tau+1}$'s in $B$ do not change, $\mathbf{r}_B^{\tau}$ is assumed to have converged to its final value, $\mathbf{r}_B$, and is filled in at all locations in $B$.

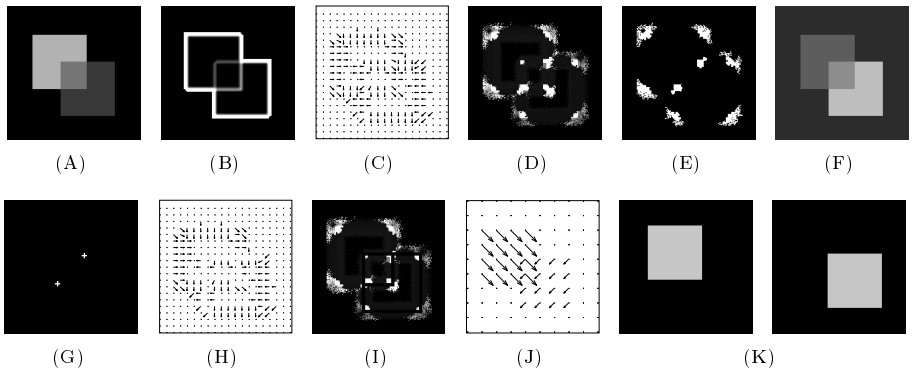


Figure 2: A) An input scene composed of transparent homogeneous regions. B) Mobility image. C-D) Estimates and certainties. E) Initial motion segmentation. F) Luminance segmentation. G) Detected X-junctions. H-I) The remaining estimates and their certainties. J) Refined motion estimates. K) Final motion segmentation.

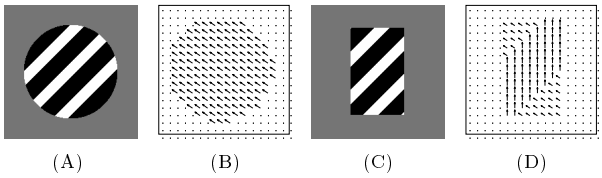


Figure 3: The barber pole illusion. Gratings are moving vertically upward, when the aperture is A-B) circular and C-D) rectangular.

Following the integration stage, couplings in the motion network are updated based on the refined estimates and the final segmentation result is obtained using $\theta_{M,2}$.

## 4 Results

Image sequences have three frames having the size of $256 \times 256$. The threshold set, $(\theta_{M,1}, \theta_{M,2}, \theta_B)$, and the maximum displacement considered in (1) are $(20, 10, 10)$ and 10, respectively, for all scenes. Our model is not sensitive to the selection of other parameters. Although model outputs have the size of $230 \times 230$, we show their spatially subsampled versions for better visualization.

**Motion transparency:** The input scene in Figure 2A is composed of two transparently overlapping square regions. The upper region moves in the rightward and downward direction while the lower one has a leftward and downward motion. Both regions have the speed of two pixels/frame. Figures 2B-I show the progress at intermediate processing steps. Due to the detected transparency, the result in Figure 2J is obtained by considering the overlapping area as a part of both regions. Thus, locations in this area are assigned both velocities. The final result includes two overlapping square segments as depicted in Figure 2K.

**The barber pole illusion:** An intriguing visual illusion occurs when gratings move behind an aperture. Gratings in Figures 3A and C move vertically upward behind a circular and a rectangular aperture, respectively, at a speed of five pixels/frame. When the aperture is circular, the motion is perceived to be in the perpendicular direction to the grating orientation, consistent with our result shown in Figure 3B. When the aperture is rectangular, our model results in a distribution of velocities parallel to the longer axis of the aperture, mimicking the barber pole illusion as shown in Figure 3D. Diagonal velocities in the upper-left and the lower-right corners are also consistent with human perception.

**Square plaids:** A square plaid can be created by the spatial repetition of a tile pattern shown in Figure 4A. *Duty cycle* of a plaid is defined as the ratio of $l_1/(l_1 + l_2)$. When regions $b$ and $d$ have the same luminance, a symmetric plaid is obtained (Fig. 4B). By assigning different luminances to regions $a$, $b$, $c$, and $d$, an asymmetric plaid can be formed (Fig. 4C).

In our results, plaids move five pixels/frame in the upward direction. Because of the spatial regularity in the luminance segments and the proximity among regions having the same luminance, (3) takes place in segments having the same luminance simultaneously. To compare the model output with the psychophysical data, an area-based probability measure is defined. For this probability, the total area of the regions having their refined motion estimates in non-coherent motion directions is calculated and normalized by the area of the aperture. Thus, when the majority of the segments have non-coherent motion, their total area increases and so does the probability of non-coherent motion perception for the input. Our results on plaids are summarized in Figure 5 where the vertical axis shows the probability of non-coherent motion perception, the solid line corresponds to the model output, and the dotted line is the psychophysical data.

We first analyze symmetric plaids where $I_a = 255$, $I_b = I_d = 100$. In Figure 5A, $I_c$ (horizontal axis) varies between 10 and 140. Examining the plot, we observe that a non-coherent motion is more frequently perceived when gratings transparently overlap, i.e., $39 \leq I_c < 100$, quantitatively consistent with the data from [4].

In Figure 5B, we consider the perceptual dependence on the duty cycle (horizontal axis). Our result is consistent with the qualitative observation that an increase in the duty cycle decreases the probability of non-coherent motion perception [4].

For an in depth study of the role of luminance in motion perception, we consider three categories of asymmetric plaids from [1]. Configurations in Category I, II, and III allow for two, one, and no transparent interpretations, respectively. Luminances of 0, 64, 128, and 255 are assigned to the corresponding regions in each configuration. For instance, $I_c = 0$, $I_d = 64$, $I_b = 255$, and $I_a = 255$ in *cdba*. Examining the plots in Figure 5C-E, we observe that the model result is in good agreement with the data reported in [1].

## 5 Discussion and Conclusions

We proposed an oscillatory model of human motion perception based on the integration of motion and luminance information. Our model, which can be expressed in a Bayesian framework, is similar to that suggested by Weiss [7], which reports a single final motion direction for an input scene. Unlike his, our model provides a spatial motion distribution, making it more useful and applicable for real scenes. For instance, diagonal estimates in Figure 3 cannot be captured by his model. Unlike other approaches, (see [7]), our model does not require specific motion mechanisms to explain the perception of plaids. Our model is also similar to the one proposed by Nowlan and Sejnowski [2], where a network is trained for estimate certainty assessment. Our model includes an analytical expression for the certainty, avoiding expensive training and the design of an appropriate network. Furthermore, our model output has a higher spatial resolution than their model does.

In summary, our model is able to represent motion transparency and to explain an intriguing perceptual phenomenon in plaids by integrating the results from two simple segmentation networks. It also mimics the barber pole illusion. Having a biologically plausible computational paradigm, our model suggests a new approach to study human visual system.

### Acknowledgments

We thank J. T. Todd for many helpful discussions. This work was supported in part by an NSF grant (IRI-9423312) and an ONR Young Investigator Award (N00014-96-1-0676) to DLW and an NSF grant (SRB-9514522) to DTL.

### References

[1] D. T. Lindsey and J. T. Todd, "On the relative contributions of motion energy and transparency to the perception of moving plaids," *Vision Res.*, **36**(2):207-222, 1996.

[2] S. J. Nowlan and T. Sejnowski, "Filter selection model for motion segmentation and velocity integration," *J. Opt. Soc. Am. A*, **11**(12):3177-3200, 1994.

[3] W. Reichardt, "Autokorrelationsauswertung als Funktionsprinzip des Zentralnervensystems" *Z. Naturforsch.*, **12b**:447-457, 1957.

[4] G. R. Stoner, T. D. Albright, and V. S. Ramachandran, "Transparency and coherence in human motion perception," *Nature*, **344**:153-155, 1990.

[5] D. Terman and D. L. Wang, "Global competition and local cooperation in a network of neural oscillators," *Physica*

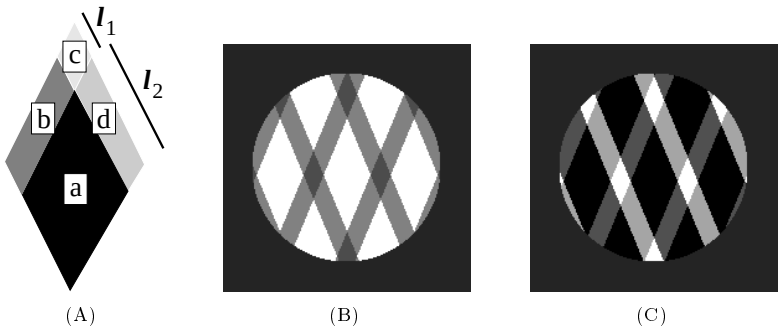


Figure 4: A) A generic tile to form B) a symmetric and C) an asymmetric plaid.

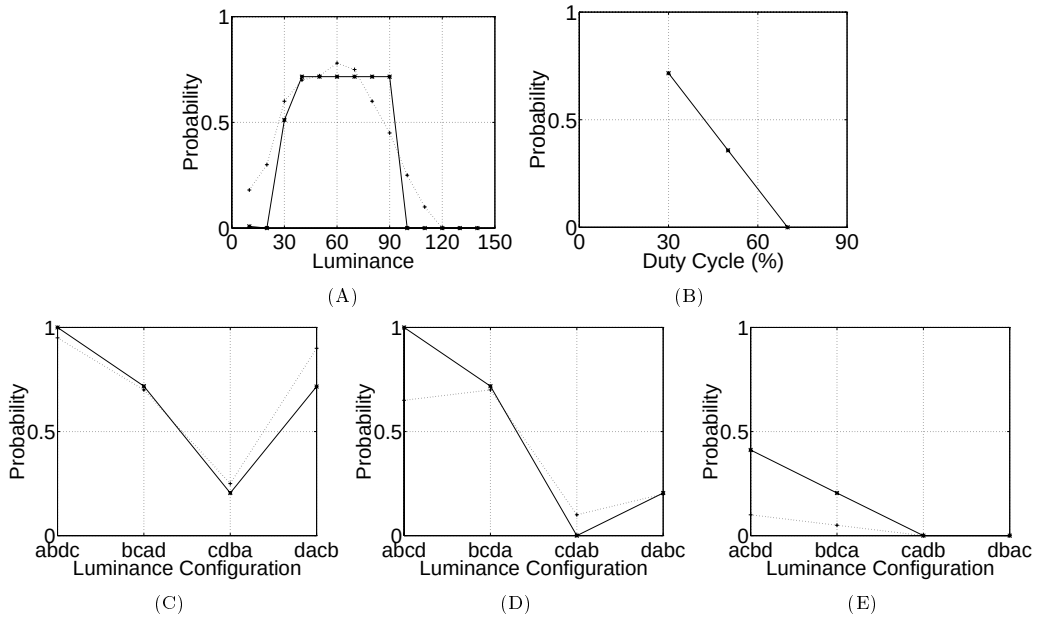


Figure 5: Model (solid line) is compared against the psychophysical data (dotted line) using the probability of non-coherent motion perception. A-B) Symmetric plaids where (A) $I_c$ and B) duty cycle are varied. C-E) Asymmetric plaids with configurations from categories C) I, D) II, and E) III.

*D*, **81**:148-176, 1995.

[6] D. L. Wang and D. Terman, "Locally excitatory globally inhibitory oscillator networks," *IEEE Trans. Neural Networks*, **6**:283-286, 1995.

[7] Y. Weiss, *Bayesian motion estimation and segmentation.* PhD thesis, Dept. Brain and Cognitive Sci., MIT, Cambridge MA, 1998.