Perceptual Multistability in a Temporal Illusion

Joseph Marrama
Department of Symbolic Systems
Stanford University
jmarrama@stanford.edu

Alden Timme
Department of Math and Computational Sciences
Stanford University
aotimme@stanford.edu

Abstract

In this project, we model the cognitive conception of a visual illusion that can be perceived in four primary modes. We model perceptual multistability using a generative Bayesian network that captures the relations between high level features that are extracted from the illusion and the different overall conceptions of the illusion that the viewer can have. Due to the high complexity of the illusion itself, we simplify the visual stimulus down to a single unit that corresponds to the viewer's observation of the visual stimulus. Despite this simplifying assumption, our model is sophisticated enough to allow it to exhibit human-like perceptual multistability using Markov Chain Monte Carlo (MCMC) methods to sample from the model conditioned on the observed stimulus. (once we actually get results we should write the rest)

1 Perceptual Multistability

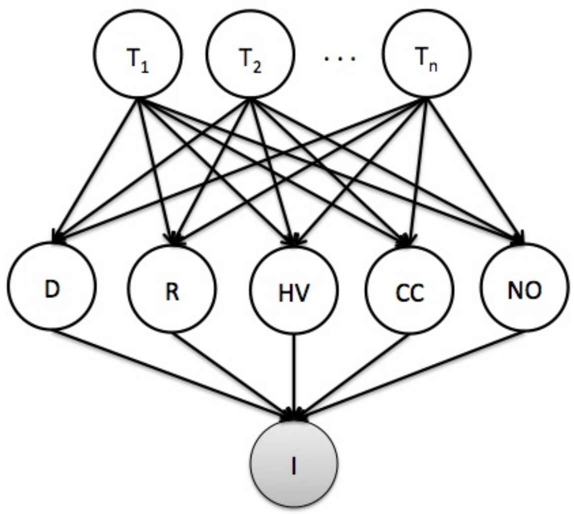
The mental phenomenon of perceptual multistability is a commonly modeled aspect of human cognition, because it is a well-defined and well-measured phenomenon, and it is fairly easy to model simple illusions that give rise to multistability. However, little research has gone into modeling perceptual multistability in complex illusions, such as those with a temporal component, multiple stable percepts, and complex structure. In our project, we aim to model a high-dimensional temporal illusion with 6 stable percepts (which can be found at bit.ly/c3uGK2) and accurately reconstruct the dynamics of perceptual multistability as it naturally arises from the illusion using MCMC sampling.

2 Our Model

Due to the high complexity of the illusion, a number of simplifying assumptions have to be made in order to construct a computationally tractable model. Before we constructed our model, we decided to reduce the number of stable percepts in the illusion to 4 instead of 6, because the mirror images of the double helix and wave forms (see the illusion) are very similar, to the point where they are practically interchangable. The difference between the helix, wave, horizontal motion, and bouncing dots percepts is much more significant than the difference between the 2 wave and helix configurations.

Using this simplification, we constructed a 3-layer generative Bayesian network for our model. Each unit in each layer (except the bottom layer) contains directed edges going from that unit to every unit in the layer immediately below it. The top layer contains 4 'percept' units, where each unit can take on a value 1 to 4 that corresponds to a different percept of the illusion. The top layer generates the second layer, which contains intermediate 'high-level feature' units that correspond to different characteristics one might observe. We have five of these units, which correspond to observed dimensionality (2d or 3d), planar rotation, horizontal velocity, correlation between dots in each column, and number of objects in the illusion. Lastly, the bottom layer captures the observed illusion. For the bottom layer, we model the input of the illusion as a single 'illusion unit', which has two values corresponding to whether the illusion is being observed or not. This is a very large but necessary simplification, because modeling configurations of the dots in the illusion would've taken hundreds of variables and a temporal model. Also, our approach is still wholly valid, because we can encode the configurations of high-level features that are likely to be observed in the illusion in the conditional probability distribution of the bottom-layer unit. For example, if a configuration of high-level features are observed that *don't* correspond to any percept of the illusion (say, rotation and 2-dimensionality were observed), the illusion unit has a near-zero probability of being active, whereas if a configuration of high-level features that correspond to one of the percepts is observed, then the illusion unit has a near-one probability of being active.

Our model can be viewed as a cognitive representation of the reception of a visual field. Staying true to the concepts taught in class, our model is a generative model, where each higher level of abstraction generates the lower. The model can be viewed as a small subset of the complete visual system, where we only use levels of abstraction and units that are relevant for modeling the illusion. To model the stimulus of the illusion, we condense all lower layers of abstraction/visual processing into a single layer of abstraction that represents all possible visual stimuli, and then only include the one unit that represents the stimuli of the illusion.



A graphical representation of our model, with a generalized amount of top-layer percept variables. Our model uses 4 top-layer variables unless stated otherwise.

The CPDs of all the variables in the model can be parameterized as follows. Let the second-layer 'high-level feature' units be denoted as $\{D, R, HV, CC, NO\}$, where they are in the same order as their enumeration above. Let the top-level percept units be denoted as T_1, T_2, T_3, T_4, the bottom-layer illusion unit as I, and finally, let the number of values a unit U can take on be denoted as $|U|$.

In the illusion, each stable percept corresponds to a unique assignment to the high-level feature units. Let c_1 denote this configuration for viewing the illusion as ovals moving horizontally, c_2 the configuration for the double helix, c_3 for the wave, and c_4 for the dots moving up and down. These are defined as:

$$
\begin{aligned}
c_1 &= \{D_1, R_1, HV_2, CC_2, NO_2\} \\
c_2 &= \{D_2, R_2, HV_1, CC_2, NO_1\} \\
c_3 &= \{D_2, R_1, HV_1, CC_2, NO_1\} \\
c_4 &= \{D_1, R_1, HV_1, CC_1, NO_3\}
\end{aligned}
$$

The probability distributions for all top level units T_i taking on different percepts is simply distributed according to a multinomial distribution whose parameters are given by a dirichlet distribution with hyper-parameters $\alpha = 200$ (in order to make it give a multinomial distribution close to the uniform distribution) and with a dimension of 4 (because there are 4 possible percepts).

In order to describe the probability distributions for all mid-layer units ML, lets define the following function:

$$\begin{aligned} \phi(T_i, ML_x) &= (1 - (|ML| - 1) * \epsilon) \text{ if } T_i = j \text{ and } ML_x \in c_j \\ &= (\epsilon) \text{ if } T_i = j \text{ and } ML_x \notin c_j \end{aligned}$$

Now, we can define:

$$P(ML = x | T_1, T_2, T_3, T_4) = \frac{\sum_{i=1}^{4} \phi(T_i, ML_x)}{4}$$

In more understandable terms, the probability $P(ML_x | T_1 ... T_4)$ is simply the average of all of the contributions of $T_1 ... T_4$, where the contribution of T_i is simply $(1 - \epsilon)$ if ML_x corresponds to the assigned percept of T_i, or ϵ if not. ϵ can be viewed as a noise term.

Lastly, the CPD for the bottom unit I is:

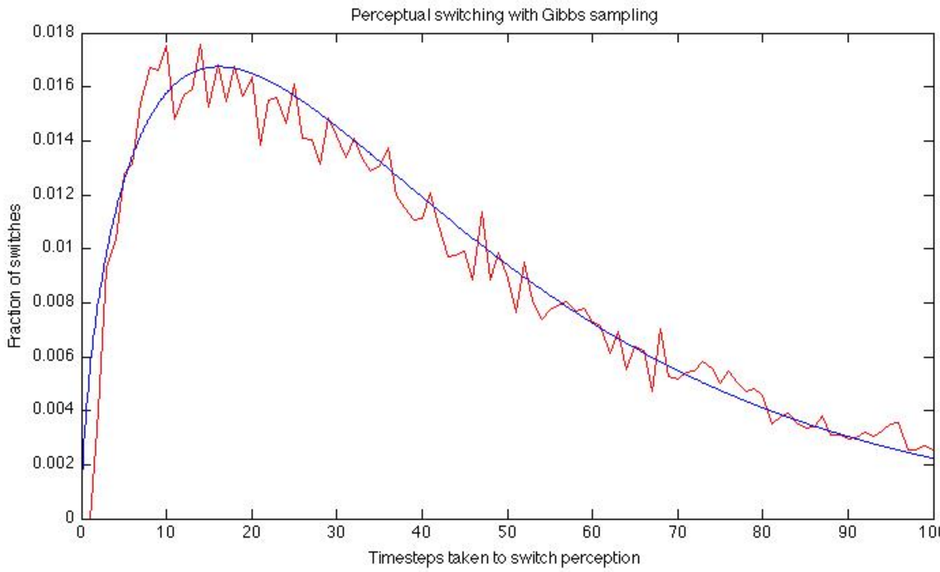
$$\begin{aligned} P(I = 1 | D, R, HV, CC, NO) &= (1 - \epsilon) \text{ if } c_i = \{D, R, HV, CC, NO\} \\ &= \epsilon \text{ if } c_i \neq \{D, R, HV, CC, NO\} \end{aligned}$$

3 Results

For all experiments, we used standard Gibbs sampling as our MCMC sampling method, and sampled in order to approximate the posterior $P(M_{-I} | I_1)$, where I_1 means the illusion is on and M_{-I} is the whole model sans I. Unless otherwise stated, we first sample all variables using forward sampling (and set $I = 1$), then use a burn-in time of 10,000 Gibbs samples, and then collect 1,000,000 samples. For the experiments discussed below, we use a noise parameter of $\epsilon = 0.01$ and 4 top-level units T_i unless otherwise stated.

3.1 Distribution of dominance durations

As you can see in the figure below, the time to switch between dominant percepts is roughly distributed according to a gamma distribution. We define a 'dominant percept' as anytime 3 or more top-layer percept units are assigned the same percept. It has been thoroughly documented that switching times in humans between different stable percepts tend to follow a Gamma distribution. Our model has similar behavior. This is likely because each top-layer percept unit has switching times roughly distributed along an exponential distribution, and the Gamma distribution is just a sum of exponential distributions. Nonetheless, our model exhibits human-like characteristics, which is a success of our model.



The percentage of perceptual switches as a function of the timesteps taken to switch percepts is shown in red, and a fitted gamma distribution with a shape parameter of 1.7 and a scale parameter $\Theta = 27$ is shown in blue.

3.2 Visual Cues

In the illusion, there are certain visual cues you can set that help you see different percepts. For example, the horizontal motion cue displays a dot moving along with each object to cue the viewer into the fact that there are multiple objects moving along at a high horizontal velocity. This provides a unique opportunity to test our model by holding these cues constant and sampling from the new posterior $P(M_{-I-cues}|I = 1, cues)$, and then using the resulting samples to approximate $P(T_i|I = 1, cues)$. Intuitively, we should expect $P(T_i = j|I = 1, cues)$ to be much higher than $P(T_i = j|I = 1)$, assuming the cues are for percept j.

Using the gathered samples from Gibbs sampling on the posterior $P(M_{-I}|I = 1)$, we can easily compute the posterior $P(T_i = j|I = 1)$ by summing over samples. Computing each $P(T_i = j|I = 1)$ and then averaging over all top layer percepts T_i, we are left with the following probabilities for $P(T = j|I = 1)$:
.247 for j = Horizontal moving shapes
.26 for j = Rotating double helix
.24 for j = Wave
.253 for j = Uncorrelated bouncing dots

For the horizontal cue, we decided that the two obvious features the cue prompted the viewer to notice are horizontal motion and multiple objects. Setting $HV = 2$ and $NO = 2$, and using 200,000 iterations of Gibbs sampling to calculate $P(T = j|I = 1, HV = 2, NO = 2)$, we get the following probabilities:
.512 for j = Horizontal moving shapes
.155 for j = Rotating double helix
.153 for j = Wave
.168 for j = Uncorrelated bouncing dots

This is exactly what we would expect to see; observing the cues for Horizontal moving shapes drastically increases the probability of being in that percept. We can do the same for the Helix cue, which tips us off to the features that the illusion is 3-dimensional and rotating. Using 200,000 iterations of Gibbs sampling, $P(T = j|I = 1, R = 2, D = 2)$ =
.13 for j = Horizontal moving shapes
.472 for j = Rotating double helix
.229 for j = Wave

.16 for j = Uncorrelated bouncing dots

Again, this is exactly what we would like to see. Repeating the same process with the other 2 cues leads to similar results; $P(T = j|I, cues)$ is generally around .5 when the cues favor percept j.

4 Discussion