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. (figure would be nice) The top layer contains (4? a variable number?) '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.

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:

$$
\begin{aligned}
c_1 &= \{D_0, R_0, HV_1, CC_1, NO_1\} \\
c_2 &= \{D_1, R_1, HV_0, CC_1, NO_0\} \\
c_3 &= \{D_1, R_0, HV_0, CC_1, NO_0\} \\
c_4 &= \{D_0, R_0, HV_0, CC_0, NO_2\}
\end{aligned}
$$

The probability distributions for all top level units T_i are simply:

$$
\begin{aligned}
P(T_i = 1) &= 0.35 \\
P(T_i = 2) &= 0.35 \\
P(T_i = 3) &= 0.2 \\
P(T_i = 4) &= 0.1
\end{aligned}
$$

The probability distributions for all mid-layer units ML is:

$$
\begin{aligned}
P(ML = x | T_1, T_2, T_3, T_4) &= (1 - (|ML| - 1) * \epsilon) \text{ if } T_1 = T_2 = T_3 = T_4 = i \text{ and } ML_x \in c_i \\
&= (\epsilon) \text{ if } T_1 = T_2 = T_3 = T_4 = i \text{ and } ML_x \notin c_i \\
&= \frac{1}{|ML|} \text{ if } T_i \neq T_j
\end{aligned}
$$

In more understandable terms, $P(ML|T_1...T_4)$ is $1 - (|ML| - 1)\epsilon$ whenever all top-layer units have the same percept i and the assignment to ML is in the corresponding c_i, and if the assignment isn't in c_i, then the probability of getting it is ϵ. ϵ can be viewed as a noise term. If the top-layer units aren't the same value, then the value of ML is chosen from the uniform distribution. Lastly, the CPD for the bottom unit I is:

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

2

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 $\epsilon = 0.05$ 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.

(insert image and small text beneath it...)

3.2 Visual Cues

4 Discussion

Table 1: Sample table title

PART	DESCRIPTION
Dendrite	Input terminal
Axon	Output terminal
Soma	Cell body (contains cell nucleus)

COOL LaTeX 2_ε STUFF!!! alright! LOL?

4.1 Footnotes

Indicate footnotes with a number[1] in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a horizontal rule of 2 inches (12 picas).[2]

4.2 Figures

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction; art work should not be hand-drawn. The figure number and caption always appear after the figure. Place one line space before the figure caption, and one line space after the figure. The figure caption is lower case (except for first word and proper nouns); figures are numbered consecutively.

Make sure the figure caption does not get separated from the figure. Leave sufficient space to avoid splitting the figure and figure caption.

You may use color figures. However, it is best for the figure captions and the paper body to make sense if the paper is printed either in black/white or in color.
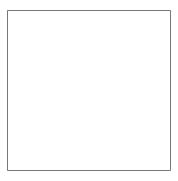
Figure 1: Sample figure caption.

4.3 Tables

All tables must be centered, neat, clean and legible. Do not use hand-drawn tables. The table number and title always appear before the table. See Table 1.

Place one line space before the table title, one line space after the table title, and one line space after the table. The table title must be lower case (except for first word and proper nouns); tables are numbered consecutively.

LaTeX users:

- Consider directly generating PDF files using `pdflatex` (especially if you are a MiKTeX user). PDF figures must be substituted for EPS figures, however.

---

[1] Sample of the first footnote
[2] Sample of the second footnote

- Otherwise, please generate your PostScript and PDF files with the following commands:

```
dvips mypaper.dvi -t letter -Ppdf -G0 -o mypaper.ps
ps2pdf mypaper.ps mypaper.pdf
```

  Check that the PDF files only contains Type 1 fonts.
- xfig "patterned" shapes are implemented with bitmap fonts. Use "solid" shapes instead.
- The `\bbold` package almost always uses bitmap fonts. You can try the equivalent AMS Fonts with command

```
\usepackage[psamsfonts]{amssymb}
```

  or use the following workaround for reals, natural and complex:

```
\newcommand{\RR}{I\!\!R} %real numbers
\newcommand{\Nat}{I\!\!N} %natural numbers
\newcommand{\CC}{I\!\!\!\!C} %complex numbers
```

- Sometimes the problematic fonts are used in figures included in LaTeX files. The ghostscript program `eps2eps` is the simplest way to clean such figures. For black and white figures, slightly better results can be achieved with program `potrace`.

Acknowledgments

References

References follow the acknowledgments. Use unnumbered third level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to 'small' (9-point) when listing the references. **Remember that this year you can use a ninth page as long as it contains *only* cited references.**

[1] Alexander, J.A. & Mozer, M.C. (1995) Template-based algorithms for connectionist rule extraction. In G. Tesauro, D. S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609-616. Cambridge, MA: MIT Press.

[2] Bower, J.M. & Beeman, D. (1995) *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System.* New York: TELOS/Springer-Verlag.

[3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.