A Cyclic Sequential Sampling Model of Bistable Auditory Perception

Michael D. Lee (mdlee@uci.edu)
Assal Habibi (habibia@uci.edu)
Department of Cognitive Sciences, 3151 Social Sciences Plaza A
University of California, Irvine, CA, 92697-5100

Abstract

We develop a cyclic sequential sampling model of bistable perception, based on the pioneering work of Vickers (1972). The model has two key parameters: a drift rate that measures the information in favor of one percept over the other; and a boundary separation that measures the evidence required by an observer to establish a percept. We implement the model within a graphical Bayesian framework, and apply it to data from several participants measuring their bistable perception for ambiguous auditory stimuli. We show that the model fits the distribution of latencies between perceptual reversals well, that the inferred drift rate parameter changes systematically as the auditory stimulus is manipulated to favor one percept over the other, and that the boundary separation parameter changes over participants to measure individual differences in their bistable perception.

Keywords: Bistable perception, bistable audition, sequential sampling models, response time modeling, diffusion model

Introduction

Bistable perception is an intriguing and important psychological phenomenon, in which a single stimulus supports two different interpretations. The key characteristic of perceptual bistability is stimulus ambiguity. In order for bistability to occur there must be more than one plausible alternative inherent in the stimulus presented to the perceptual system.

In vision, bistable perception can be achieved through ambiguous depth cues, as in the Necker cube (Necker, 1832) which is a two-dimensional representation of a three-dimensional cube, or through binocular rivalry (Helmholtz, 1925), where each eye is presented with a different and incompatible image. In both cases observers experience clear switches in perception, in the absence of any change in stimuli.

Traditional accounts of bistable phenomena propose that the basis of alternation in perception is a peripheral or sensory process, where the perceptual dominance of one stimulus is the result of activation of subset of neurons encoding that stimulus while simultaneously inhibiting those that perceive the alternative stimulus. Over time, fatigue or satiation in the system pushes the subjective state to reverse (Koehler & Wallach, 1944). More recent models suggest an alternative view of bistable phenomena under which alternation is a sign of responses to active, programmed events initiated by brain areas that integrate sensory and non-sensory information to guide behavior (Leopold & Logothetis, 1999).

Under this more recent view, bistable perception can be considered a result of the exploration of the sensory environments, and so reflects a fundamental aspect of cognition supporting flexible decision making (Kim, Grabowecky, & Suzuki, 2006). In addition, there is considerable research interest in bistable perception from the perspective of investigating correlates of conscious perception, since changes in perceptual awareness can be experienced in the absence of changes in stimulus.

According to Leopold and Logothetis (1999), the key characteristics of bistability are exclusivity, randomness and inevitability. Exclusivity refers to the existence of two possible yet mutually exclusive alternative interpretations of the sensory input. Randomness characterizes the statistical distribution of the time spent in each percept. Inevitability refers to the finite duration of perceptual dominance. That is, even when the intention is to hold onto one interpretation, observers only have limited volitional control on perceptual alternation.

As a concrete example of these properties, the alternation of bistable perception for the Necker cube is typically estimated to be equally distributed between the two percepts, and the rate of reversals is estimated to level off to an average of 16–20 times per minute after a period of initial learning.

Auditory Bistability

In this paper, our focus in on modeling data from an auditory perception task that induces a bistability. Although audition is less studied than vision, it is a basic and important question for cognitive science to understanding how people perceive ambiguous auditory stimuli. In language, for example, listeners must segment words and phrases from the ongoing speech stream in order to make sense of the incoming signal. In music, comprehending melodic structure involves segmenting tone sequences into smaller coherent chunks in order to discern larger patterns.

The bistable phenomenon we use involves a series of low tones (L), high tones (H) and silences (-) being presented one after the other with a fixed interstimulus interval (i.e., L-H-L-H-L-H). When the frequency difference

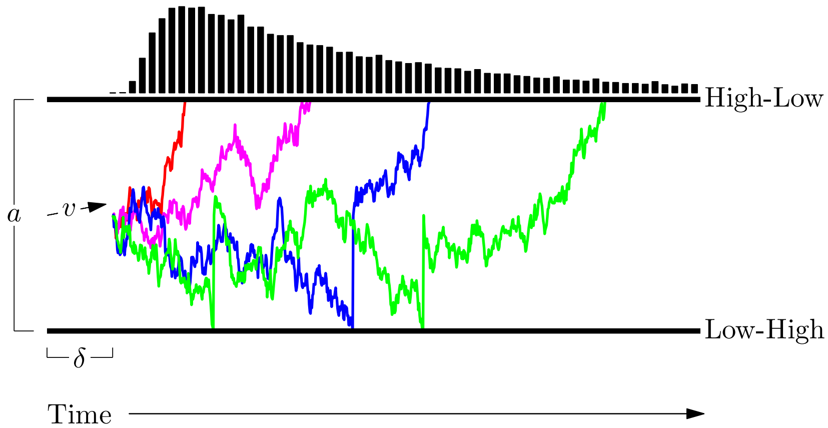

Figure 1: Schematic representation of the cyclic sequential sampling model.

between L and H, denoted Δf, is small and repetition rate is slow, listeners typically report a single LHL "galloping" pattern. When Δf is large enough and the repetition rate is sufficiently fast, listeners report hearing two separate streams of tones, each in a metronome-like rhythm (i.e., L-L-L-L- and H-H-H-H-). At intermediate Δf, the stimulus is ambiguous and perception can alternate between interpretations of one and two streams following the initial buildup (Pressnitzer & Hupé, 2006). The perception can alternate between interpretations LH and HL (Bregman & Campbell, 1971). A number of recent studies (e.g., Cusack, 2005; Gutschalk et al., 2005; Snyder, Alain, & Picton, 2006; Winkler, Takegata, & Sussman, 2005) have similarly exploited the bistability of auditory perception in investigating the neural correlates of auditory perceptual organization.

Overview

Our model of perceptual bistability uses the sequential sampling framework developed in mathematical psychology. The basic idea is to assume people accrue information from a stimulus by ongoing observation, and, even when a percept is established, continually re-evaluate the incoming information in terms of competing possible perceptual interpretations. For ambiguous stimuli, this process of re-evaluation will eventually favor the rival interpretation, at which point a perceptual reversal occurs. We develop a formal quantitative model implementing these ideas, and evaluate it against data from a number of auditory L-H-L-H- sequences, in which the lengths of the inter-stimulus silence intervals are systematically manipulated to bias in favor of one percept over the other.

The structure of the paper is as follows: In the next section we describe cyclic sequential sampling models, as they can be applied to modeling bistable perceptual decision-making, and provide the formal details of our model. We then describe the experimental procedures used to gather test data, and apply our the model to the data. We discuss the ability of the model to fit the data, and infer meaningful parameter values. We conclude with a discussion of the usefulness of the modeling approach in measuring and understanding bistable perception.

A Cyclic Sequential Sampling Model

Sequential sampling models are successful and widely-used accounts of human decision-making. In these models, the decision-maker is assumed to sample information from a stimulus, until some critical level of total evidence has been obtained internally, and an overt behavioral response is triggered. In this way, sequential sampling models provide a detailed account of the time course of decision-making, and make predictions about a range of experimentally observable measures, including decision accuracy, response time, and response confidence (e.g., Busemeyer & Rapoport, 1988; Ratcliff, 1978; Vickers, 1979)

Our cyclic sequential sampling model is a modification of the standard approach, suited to a bistable perception task rather than a general two-choice decision task, and is directly inspired by the model proposed by Vickers (1972). A schematic representation is shown in Figure 1, which shows four different possible sample paths moving from left to right. Each of paths moves according to evidence sampled from a stimulus that can be perceived in two states. The two boundaries, shown by the solid lines, correspond to these states. We label these 'high-low' and 'low-high', corresponding to the auditory bistability stimuli that are the focus of this study.

Unlike a standard sequential sampling model, each sampled paths only terminates when it reaches the upper boundary. This is because the participant currently maintains one of the two bistable perceptions, and an overt response is only triggered when sufficient evidence is gathered for the alternative perception. In Figure 1, the participant begins with the 'low-high' percept corresponding to the lower boundary, and registers a change-in-percept response to 'high-low' only when the upper boundary is reached. The arrivals of the sample paths

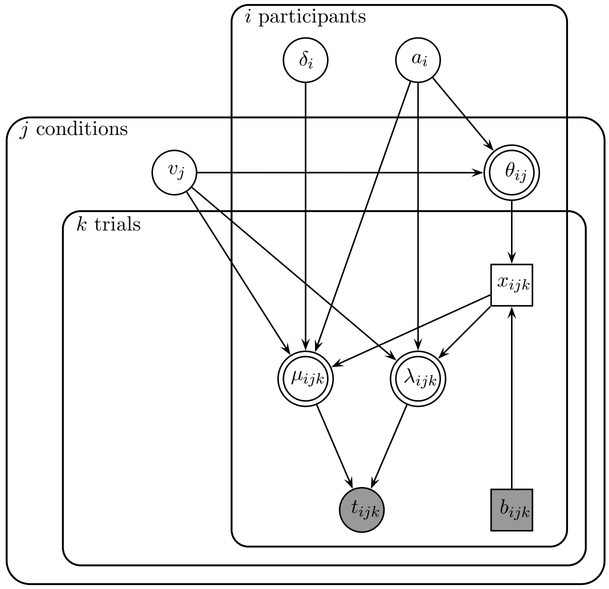

Figure 2: Graphical model implementation.

give rise to the distribution of perceptual reversal times shown by the histogram above the upper boundary.

Notice that if a sample path reaches the lower boundary corresponding to the current percept, it is reset to the starting evidence value between the two boundaries, and the evidence accumulation process continues. This was a basic insight provided by Vickers (1972), and is why the model is called a cyclic variant of the sequential sampling approach.

As shown in Figure 1, there are three parameters in the model. The drift rate v is the mean of the Gaussian distribution from which evidence values are sampled, and corresponds to the relative level of evidence the stimulus provides for one perceptual interpretation over the other. The boundary separation a is the difference between the two decision boundaries, and corresponds to the level of evidence a participant needs to reach a perceptual interpretation. Finally, the offset δ captures the component of perceptual reversal time observed experimentally that is not due to the workings of the internal decision process, but rather to other factors such as movement time to record a response.

Implementation as a Graphical Model

We implement our cyclic sequential sampling model using the formalism provided by graphical modeling (see Lee, 2008; Shiffrin, Lee, Kim, & Wagenmakers, 2008, for recent tutorials aimed at cognitive scientists), implemented in WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000). This allows us to perform fully Bayesian inference on our model using experimental data.

The graphical model is shown in Figure 2. Each node in the graph corresponds to a variable, and their dependencies are captured by the graph structure, with children depending only on their parents. Following Lee (2008), circular and square nodes denote continuous and discrete variables; unfilled and filled nodes denote unobserved (i.e., latent parameters) and observed (i.e., data) variables; encompassing plates represent independent replications of a part of the graph within the model, corresponding to numbers of participants doing numbers of experimental conditions; and double-bordered nodes are deterministic functions of other nodes.

Our graphical model implementation only approximates the response time distributions described by the cyclic decision model. The basis of the approximation comes from observing that the response time distribution at each boundary is a mixture of response time distributions over paths that reached the boundary after $0,1,\ldots$ resets. We assume each of these mixture components has a Gaussian distribution, which makes the model very tractable.

Formally, in the graphical model there is boundary separation a_i and offset δ_i for the ith participant, and the drift rate v_j for the jth condition. A standard result (e.g., Wagenmakers, van der Maas, & Grasman, 2007) is that the probability of reaching the upper boundary for the ith participant in the jth condition is $\theta_{ij} = 1/\left(1+\exp\left(a_i v_j\right)\right)$. The number of resets needed before the boundary is reached on the kth trial is therefore sampled as $x_{ijk} \sim \text{Geometric}\left(\theta_{ij}\right)$ for the upper boundary and $x_{ijk} \sim \text{Geometric}\left(1-\theta_{ij}\right)$ for the lower boundary. In the graphical model, the appropriate boundary is given by an indicator variable b_{ijk} for the $i$the participant in the jth condition at the kth trial, which is a known part of the experimental data.

For the Gaussian approximation, we rely on recent analytic results (Wagenmakers et al., 2007) giving the mean

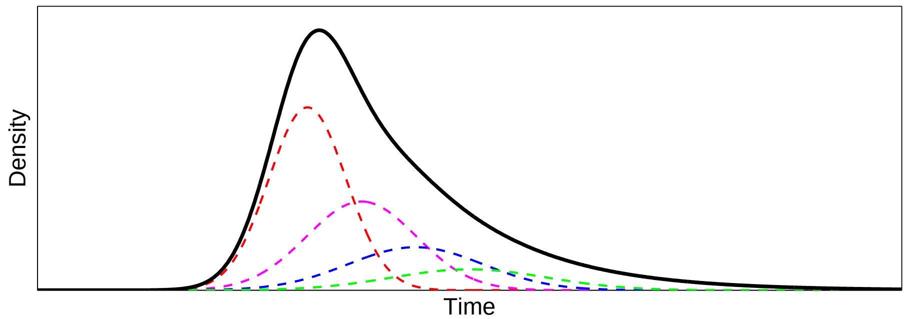

Figure 3: Approximation to the response time distribution by a mixture of progressively less weighted, increasing mean and increasing variance Gaussians.

and variance for the distribution of time taken to reach a boundary from the starting point half way between. The mean is given by

$$\frac{a_i\left(1-\exp\left(-a_i v_j\right)\right)}{2v_j\left(1+\exp\left(-a_i v_j\right)\right)},$$

and the variance by

$$\frac{2v_j^3\left(\exp\left(a_i v_j\right)+1\right)^2}{a_i\left(2a_i v_j \exp\left(-a_i v_j\right)+\exp\left(-2a_i v_j\right)-1\right)}.$$

Accordingly, we assume a reversal time coming from x_{ijk} resets, and after a fixed offset δ_i, has mean

$$\mu_{ijk}=\delta_i+x_{ijk}\frac{a_i\left(1-\exp\left(-a_i v_j\right)\right)}{2v_j\left(1+\exp\left(-a_i v_j\right)\right)},$$

and variance

$$\lambda_{ijk}=x_{ijk}\frac{2v_j^3\left(\exp\left(a_i v_j\right)+1\right)^2}{a_i\left(2a_i v_j \exp\left(-a_i v_j\right)+\exp\left(-2a_i v_j\right)-1\right)}.$$

So, finally, our graphical model assumes the observed reversal time t_{ijk} on the kth reversal of the ith participant in the jth condition, is distributed as $t_{ijk} \sim \text{Gaussian}\left(\mu_{ijk},\lambda_{ijk}\right)$.

We emphasize this does not correspond to assuming the total reversal time distribution has a Gaussian form, because we are mixing a series of Gaussians, and this mixture has an appropriate negatively skewed shape. This approach to approximation is shown in Figure 3, using the drift rate $v = 0.06$ and boundary separation $a = 1$ as a concrete example. This gives a probability of $\theta = .515$. The broken lines correspond to the Gaussian distributions for $k = 1$, 2, 3 and 4 resets (using the same color coding as Figure 1), which covers more than 97% of the total probability according to the Geometric distribution. The means and variances of these Gaussians are given by the approximations in the graphical model, and the relative probability of each Gaussian in determining the overall mixture is given by the Geometric distribution. The solid black line is the weighted sum of the Gaussian components, and has the characteristic negative skew of a response time distribution.

Experiment

Participants

The pilot data collected to develop and evaluate our model were collected from four naive participants.

Stimuli

Stimuli were 60 s sequences, of three types, in which tones alternated in amplitude. The first type was unbiased, with 250ms gaps between both tones. The second and third types were biased—towards HL and LH respectively—by alternating between 150ms and 350ms gaps, with the shorter gap corresponding to a bias towards that percept.

The low tone was a 440 Hz pure tone and the high tone was a 660 Hz pure tone (10 ms rise/fall) with a duration of 250 ms. A total of 60 trials were presented to each participant, in a counter-balanced order. Stimuli were generated in MATLAB, at CD quality (44.1 kHz sample rate) and were presented via earphones to both ears. Sound levels were not measured, but were verified to be easily audible to all participants.

Procedure

The participants were familiarized with the experiment by hearing one example sequence. The experimental sequences were then presented in random order. The observer was asked to indicate their perceived grouping by pressing the corresponding button (HL versus LH) on the screen using a mouse. All the instructions were given verbally in English.

Modeling Analysis

Our modeling results are based on 10,000 posterior samples, collected after a burn-in of 10,000 samples, and using multiple MCMC chains to assess convergence. Figure 4 show the ability of the model to fit the data. Each panel corresponds to an individual participant, experiencing a specific type of reversal (i.e., either HL changing to LH, or vice versa), in a specific experimental condition (i.e., the three types of stimuli). The panels are arranged in groups of four, corresponding to the four

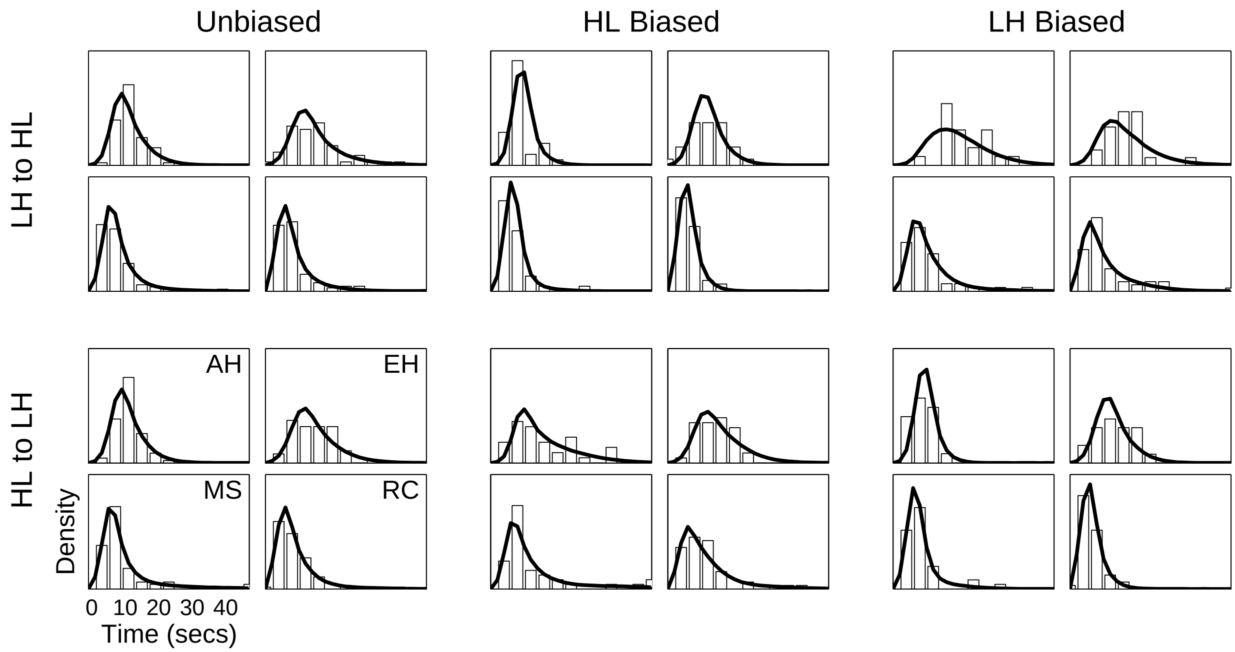

Figure 4: Posterior predictive fit of model to all data.

participants, and in major rows corresponding to reversal type, and columns corresponding to stimulus type.

Within each panel, the histograms represent the observed response times, and the solid line shows the posterior predictive distribution of the model. The response times are longer when the stimulus is biased against a percept, shorter when it is biased towards the percept, and intermediate when there is no bias. The model clearly captures these patterns. Different participants also have consistently different distributions across the conditions—such as EH having longer inter-reversal times—and the model also captures these patterns.

It is important to note that the posterior prediction used in Figure 4 is not a maximized fit, as typically seen in tests of sequential sampling models (e.g., Ratcliff & Smith, 2004), but rather an averaged fit, taken over the entire posterior parameter space, and so automatically takes into account model complexity. This means the ability of the model to fit all of the raw data well, as seen in Figure 4, provides strong evidence that it has a basic level of descriptive adequacy.

Figure 5 show the marginal posterior distributions for the drift rate and boundary separation and parameters. There is systematic variation in the drift rates over conditions, with the unbiased condition drift rate posterior being centered on zero, while the biased conditions show drift rates above and below zero, as expected. There is essentially no overlap between the distributions, and it is clear that the experimental bias manipulation had the theoretically expected effect on drift rates. There is also evidence of individual differences in the inferred boundary separation parameters, with, for example, participant EH being inferred as requiring greater levels of evidence before a perceptual reversal decision is made. Taken together, these parameter inferences demonstrative a selective influence property (e.g., Voss, Rothermund, & Voss, 2004) for our model, meaning that drift rate changes when the stimulus changes, and boundary rates change across participants.

Discussion

Our initial modeling results suggest that the cyclic sequential sampling approach can provide a good descriptive account of the distribution of inter-reversal times in bistable perception, and can infer meaningful parameter values. This means the model promises to provide a mechanism for furthering our understanding and ability to measure bistable perceptual phenomena.

The two key parameters in the model are the drift rate and boundary separation. The drift rate is a property of the stimulus, and measures the evidence the stimulus provides for each possible ambiguous interpretation. Our model allows this measure to be taken from behavioral data, and introduces the possibility of developing theoretical accounts of how physical properties of the stimulus (e.g., the inter-stimulus intervals) relate to its psychological properties (i.e., its evidence for a percept). The boundary separation is a property of the observer, and measures the level of evidence required to alternate between percepts. The ability of our model to infer these values introduces the possibility of exploring individual differences in bistable perception.

More generally, the initial success of our model supports using sequential sampling models as theoretical ac-

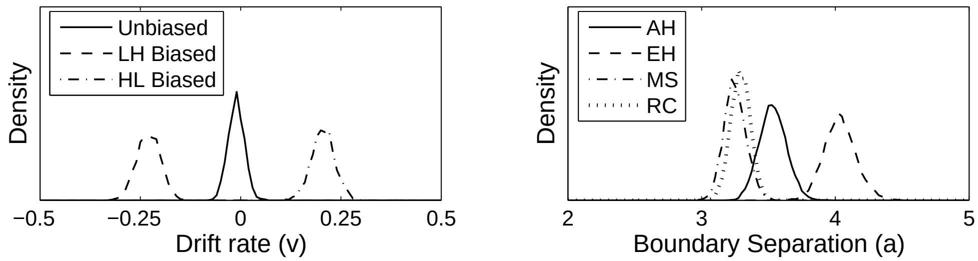

Figure 5: Posterior distributions for the drift rate and boundary separation parameters.

counts of the time course of perceptual organization. A particularly interesting finding is that the little-explored cyclic sequential sampling mechanism seems to work well. This mechanism assumes that decision-making continually resets itself, to continue searching for an alternative understanding of available information, and so formalizes a simple model of world change. It would be interesting to explore whether cyclic sequential sampling models can be applied beyond the niche of bistable perception, to more general and ubiquitous decision-making tasks in which the external environment continually changes and needs reinterpretation.

References

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, *89*, 244-249.

Busemeyer, J. R., & Rapoport, A. (1988). Psychological models of deferred decision making. *Journal of Mathematical Psychology*, *32*(2), 91–134.

Cusack, R. (2005). The intraparietal sulcus and perceptual organization. *Journal of Cognitive Neuroscience*, *17*, 641-651.

Gutschalk, A., Micheyl, C., Melcher, J. R., Rupp, A., Scherg, M., & Oxenham, A. J. (2005). Neuromagnetic correlates of streaming in human auditory cortex. *Journal of Neuroscience*, *25*, 5382-5388.

Helmholtz, H. V. (1925). *Treatise on physiological optics* (Vol. 1). Dover.

Kim, Y. J., Grabowecky, M., & Suzuki, S. (2006). Stochastic resonance in binocular rivalry. *Vision Research*, *46*, 392–406.

Koehler, W., & Wallach, H. (1944). Figural after-effects: An investigation of visual processes. *Journal of the American Philosophical Society*, *88*, 269–357.

Lee, M. D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin & Review*, *15*(1), 1–15.

Leopold, D. A., & Logothetis, N. K. (1999). Multistable phenomena: Changing views in perception. *Trends in cognitive sciences*, *3*(7), 254–264.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS: A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, *10*, 325–337.

Necker, L. A. (1832). Observations on some remarkable phenomena seen in Switzerland; and an optical phenomenon which occurs on viewing of a crystal or geometric solid. *Philosophy Magazine*, *3*, 329-337.

Pressnitzer, D., & Hupé, J. (2006). Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Current Biology*, *16*(13), 1351–1357.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two–choice reaction time. *Psychological Review*, *111*, 333–367.

Shiffrin, R. M., Lee, M. D., Kim, W.-J., & Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, *32*(8), 1248–1284.

Snyder, J. S., Alain, C., & Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *Journal of Cognitive Neuroscience*, *18*, 1–13.

Vickers, D. (1972). A cyclic decision model of perceptual alternation. *Perception*, *1*(1), 31–48.

Vickers, D. (1979). *Decision processes in visual perception*. New York, NY: Academic Press.

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, *32*, 1206–1220.

Wagenmakers, E.-J., van der Maas, H. J. L., & Grasman, R. P. P. P. (2007). An EZ–diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, *14*, 3–22.

Winkler, I., Takegata, R., & Sussman, E. (2005). Event-related brain potentials reveal multiple stages in the perceptual organization of sound. *Brain Research*, *25*, 291–299.