2022-06-28

Overview

Attempt to better characterize dissertation proposal

1 Introduction

Spoken words create audio signals that are processed by the brain in real time. That is, as the spoken word unfolds, a cohort of possible resolutions are activated until the target word is recognized. The degree to which a particular candidate word is recognized is known as activation.

An important part of this process involves not only correctly identifying the word but also eliminating competitors. For example, we might consider a discrete unfolding of a word such as "elephant" as "el-e-phant". At the onset of "el", a listener may activate a cohort of words such as "elephant", "electricity", or "elder", all of which may be considered competitors. With the subsequent "el-e", "elephant" and "electricity" remain active competitors, both being consistent with the received signal, while "elder" is eliminated. Such is a rough description of this process, continuing until the ambiguity is resolved and a single word remains.

Our interest is in measuring the degree of activation of a target relative to competitors. Activation, however, is not measured directly, and we instead rely on what can be observed with eyetracking data, collected in the context of the Visual World Paradigm (VWP). To briefly illustrate, the VWP is an experimental design in which participants undergo a series of trials to identify a spoken word. Typically, each trial has a single target word, along with multiple competitors. The target word is spoken, and participants are asked to identify and select an image on screen associated with the spoken word. Eye movements are recorded as this process unfolds, with the location of the participants eyes serving as proxy for which words/images are being considered. Complicating matters is the fact that any physiological movement is downstream the cognitive decision to do so, with some unknown interval of time separating the decision and subsequent action. As our interest ultimately lies with this cognitive process, rather than eye movements proper, we must take into account this delay.

This document serves to briefly illustrate current methods of using eyetracking data to gain insight on the underlying activation of spoken words, along with a presentation of a proposed method. We argue that the proposed method is both closer to the mental process we wish to recover in addition to being a more mathematically defensible model of the data collected.

2 Eyetracking data

To define our terms, a *saccade* represents the physical movement of the eye, typically lasting between 20-200ms. To remove ambiguity, a saccade is marked at the particular instance that the movement begins, rather than the duration of time it is in motion. A *fixation* is characterized by a lack of movement in which the eye is fixated on a particular location. Fixations also tend to be more variable in terms of duration. Together, a saccade, followed by a subsequent fixation, is known as a *look*.
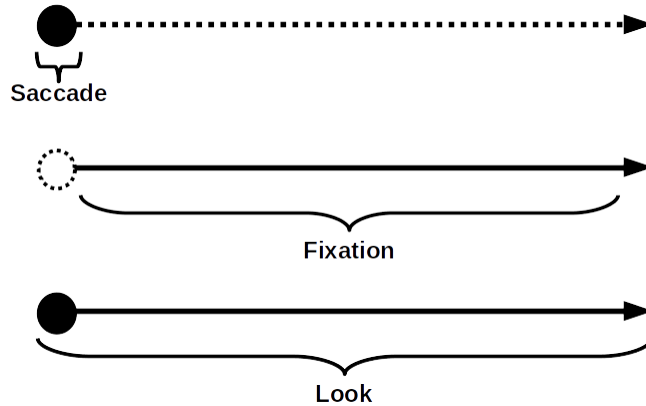

Figure 1: Types of eyetracking data

The length of a single vwp trial is commonly around 2000ms, during which time the trajectory of the eye is captured. Absent the few milliseconds each trial that the eye is in motion, each time point captures either a saccade or a fixation at a specific location. A typical trial contains around three to eight looks, recorded at discrete intervals (often 4ms) and coded with a zero or one, depending on the object of fixation.
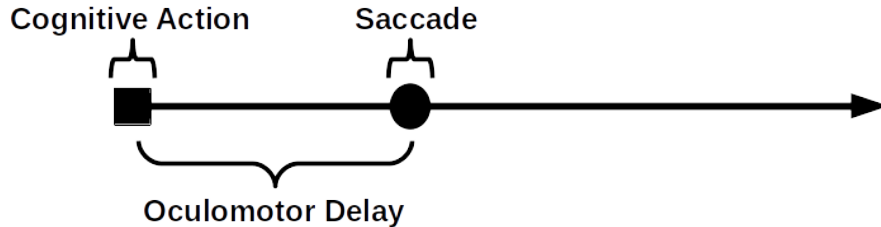

Figure 2: Oculomotor delay

We must also consider the phenomenon of oculomotor (OM) delay; whereas a particular phsyiological response (such as eye movement) may be recorded at some time, t, the latent cognitive mechanism guiding this decision occurs as some unknown point prior. Accounting for this delay will be critical in creating an unbiased recovering of the underlying functional mechanism.

Ultimately, the goal of eyetracking data is to provide a proxy measure for activation; by associating saccades or fixations to the target object with activation, we are able to use empirically collected data to emulate an activation curve as a function of time. We assume here that this curve is parameteric, notated $f_\theta(t)$. Both of the following methods are attempting to recover this same parametric curve, differing only in the data collected and how it is used. Further, both methods are burdened with the same problem of oculomotor delay. Finally, both methods result in data structures that are compatible with `bdots`, allowing subsequent analyses to be the same.

2.1 Proportion/Averaging method

One method employed to use this data involves measuring intervals of fixation to a target over a series of trials. For each trial i and time point t (typically sampled at intervals of 4ms, i.e., $t = 0, 4, \ldots, 2000$), we collect a sample of z_{it}, an indicator of whether a participant is fixated on the target object at that point in time. Averaging over the collection of trials, we construct an estimate of $f_\theta(t)$,

$$y_t = \frac{1}{N} \sum_i z_{it}.$$

In other words, it is implicitly assumed that the trajectory of the eye follows the trajectory of activation, where the average proportion of fixations at a particular time is a direct estimate of activation. As each individual trial is only made up of a few ballistic movements, the aggregation across trials allows for these otherwise discrete measurements to more closely represent a continuous curve. Curve fitting methods, such as those employed by `bdots`, are then used to construct estimates of function parameters fitted to this curve.

One of the primary benefits of this method is that it captures the duration of fixations, with longer times being associated with stronger activations. This becomes important when differentiating fixations associated with searching patterns (i.e., what images exist on screen?) against those associated with consideration (is this the image I've just heard?). A shortcoming, however, is that it conflates two distinct types of data, generated via different mechanisms, the fixation and saccade.

2.2 Saccade method

If we are to consider eyetracking data samples from some probabalistic curve, it becomes necessary to differentiate between the two types. A saccade launched at some time, t, can be considered a sample from a data-generating mechanism at t. The duration of time between a given saccade and the one following follows a different mechansim altogether. By clearly delineating the mechanism from which we are sampling, we are able to reduce observed bias in the reconstruction of the activation curve.

In light of this, and in contrast to the proportion method, we propose estimating the activation curve with the saccade data alone. The primary benefit of this is two-fold. First, as suggested above, by decoupling two different types of data we are better able to reconstruct the generating mechanism for the object of interest. Second, by proximity to the underlying cognitive mechanism, the saccade gives a better indication of activation at a specific moment in time.

An important difference between these two methods is in the structure of the data itself. Whereas the former collects an array of data, with an observation for each time point in each trial, the saccade method is sparse, with the observed data indicating the outcome of the saccade, as well as the time observed. It is best represented as a set of ordered pairs, $\mathcal{S} = \{(s_j, t_j)\}$, with j indexing each of the observed saccades, and with

$$s_j \sim Bern(f_\theta(t_j)).$$

A value of $s_j = 1$ indicates a saccade resulting in a fixation on the target.

As with the proportion method, the observed data can be used as input for `bdots` to construct estimates of generating parameters.

### 2.2.1 A note on nomenclature ("notemenclature")

With regards to naming this proposed curve, we are in a bit of a gray area: calling it an "activation" curve obscures the fact that this is not a direct measure of this otherwise latent process; calling it a "saccade" curve appears to place more emphasis on the ocular mechanics. For now, we will refer to it as a saccade curve, as it is understood literally to be the curve generating saccade data, though we actively remain open to other suggestions.

3 Oculomotor delay

We begin with an assumption that the curve of interest can be represented parametrically. For example, the four parameter logistic, defined as

$$f(t|\theta) = \frac{h - b}{1 + \exp\left(4 \cdot \frac{s}{h-b}(x - t)\right)} + b,$$

is often used to describe the trajectory of probability of a subject launching a saccade and fixating on the target location while simultaneously used as a proxy for word activation. To illustrate, a subject with the depicted fixation curve may initiate a saccade beginning at time $t = 970$, with a probability of $p = 0.5$ of subsequently resting on the target:

Logistic Fixation Curve
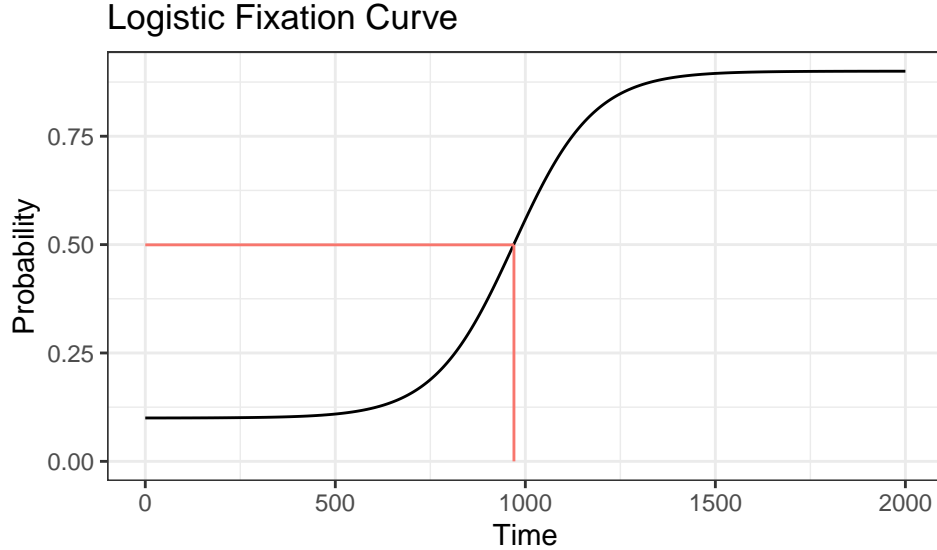
Mentioned previously, this same parametric curve is used in both the proportion and saccade methods.

As saccades are easily gathered from available eyetracking data, we are, in principle, able to collect samples directly from this curve. This goal is complicated, however, by oculomotor delay. That is, an observed saccade at t_j is likely a sample from the fixation curve $f_\theta(t)$ at some point prior to t_j. The degree to which this delay occurs, as well as the between and within subject variability of this delay, is a matter of active investigation. Most generally, we may consider an observation s_j at time t_j to be distributed

$$s_j \sim Bern[f_\theta(t_j - \rho(t_j))],$$

where $\rho(t)$ represents oculomotor delay. As written, we may consider circumstances in which:

1. $\rho(t)$ is a constant function (including 0)
2. $\rho(t)$ is a random variable, independent of the value of t_j
3. $\rho(t)$ is a random variable, dependent on t_j and possibly other aspects of the trial

To differentiate between the underlying data-generating mechanism and what is observed, we let

$$g_\theta(t) = f_\theta(t - \rho(t)),$$

where $g_\theta(t)$ is what is *observed* at time t. A saccade planned at $t = 300ms$ with an oculomotor delay of $\rho = 200ms$ will be observed at $t = 500ms$. That is, $g_\theta(500) = f_\theta(500 - 200) = f_\theta(300)$.

At present, it is common under the proportion method to account for this delay via a 200ms shift of the entire constructed proportion curve. We will propose instead a method whereby each saccade may be shifted individually and less homogenously. Reasons and implications for this will be presented in the next section.

We now consider a variety of scenarios for oculomotor delay and the subsequent impacts on the recovery of the underlying fixation curve from the observed data.

4   Simulation

We begin with the implicit assumption that there is an underlying activation curve that may be described parametrically. The following simulations will generate data according to three scenarios:

1. A situation in which the oculomotor delay is known
2. A situation in which the oculomotor delay is unknown, but of a fixed quantity (here, 200ms)
3. A situation in which the oculomotor delay is an unknown random variable, independent of time

Each simulation will be conducted with $N = 300$ trials, sampled from the same data generating function for each, with the attempted recovery of the generating curve done using the `bdots` package.

Note that these are short toy simulations to illustrate the saccade method and are not intended to be comprehensive.
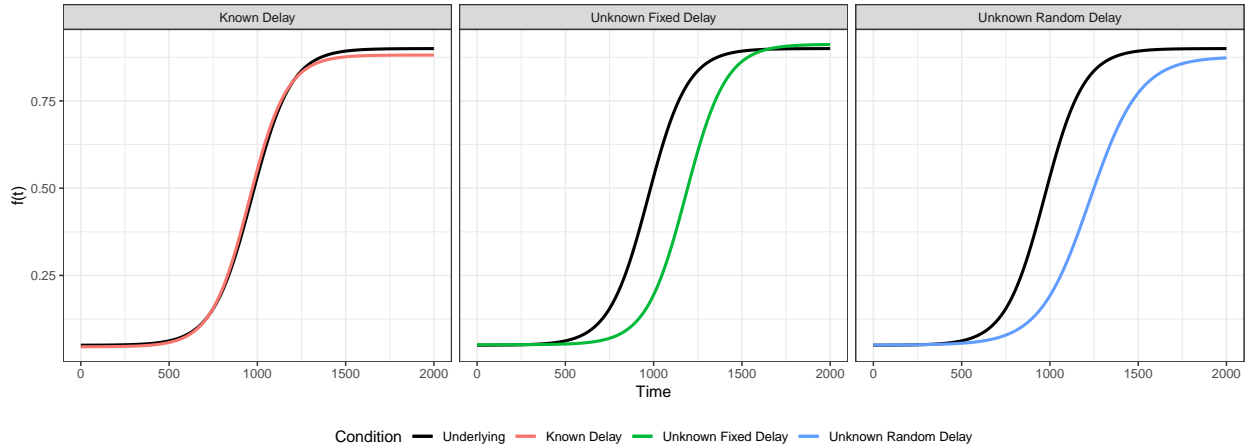
4.1   Known delay

In the case in which the oculomotor delay is known, an unbiased recovery of the data generating curve is not an issue – we simply horizontally shift each observed saccade by its known oculmotor delay.

4.2 Unknown fixed delay

The simulation was conducted using a fixed oculomotor delay of $\rho = 200ms$. Although the resulting recovered curve is biased, this bias simply results in a horizontal shift, $g(t) = f(t - \rho)$. This is especially relevant in a situation in which we are interested in comparing the data generating curve between two groups.

For example, one method of analyzing VWP data (which inspired the `bdots` package) was to determine on which intervals $I = \cup_k I_k$ two data generating curves were statistically different. Suppose, for simplicity, that there is an interval $I = [t_1, t_2]$ on which the difference between two curves, $f(t|\theta_1) - f(t|\theta_2)$, is statistically significant. Given that we observe $g_i(t) = f(t - \rho|\theta_i)$, we would simply find that a significant difference occurs at $I + \{\rho\} = [t_1 + \rho, t_2 + \rho]$, a horizontal shift resulting from the oculomotor delay.

In other words, the size of the interval would remain the same, and the relative differences between curves would be preserved under a horizontal shift.

4.3 Unknown random delay

The final scenario for consideration involves a situation is which the OM delay is unknown and random. Here, the bias will not be resolved with a simple horizontal shift, and the shape of the curve itself may be different between the one generating the data and the one observed. This has largest implications when comparing estimated curves between two groups.

We will interogate a number of potential methods for dealing with this issue, though we do feel confident that, even with this known bias, our proposed method will still be preferable to existing ones.

5 Future Directions

With regards to the curve described, there are a number of avenues seemingly worthy of investigation. The most pressing of these appears to be methods to minimize the amount of bias present in scenario three, which presents the largest obstacle in the functional recovery of the data generating mechanism. Of special note here is the fact that the particular intervals in which this bias occurs can have a large effect on the overall bias, over and above that introduced by the occulomotor delay.

For example, consider the plot above in the situation in which there is an unknown random delay (blue curve). We may observe at 500ms the value of the data generating mechanism at 300ms (that is, we observed $g_\theta(500) = f_\theta(300)$), while $g_\theta(500) \approx f_\theta(500)$. In other words, the bias over this area is small if we make no correction.

In contrast, an observation at $t = 1000$ results in a highly biased estimate, as $g_\theta(1000) \ll f_\theta(1000)$. Accordingly, we note that the amount of bias at an observed point is a function of the derivative of the data generating function in a neighborhood of that point. Whether or not this observation proves profitable remains to be seen.

There also seems to be value in finding a way to incorporate the length of fixations into the modeling process. For example, we might consider the impact of weighting each saccade by the duration of its subsequent fixation, as it seems intuitive that saccades resulting in longer fixation periods are more likely initiated by activation rather than a searching pattern.