

# saccade

## Abstract

Right now this is just copied over from the overview given after prospectus (and the follow up).

Sections correspond to topics rather than paper sections (for now)

## 1 Introduction

Spoken words create analog signals that are processed by the brain in real time. That is, as the spoken word unfolds, a cohort of possible resolutions are considered 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 the word “elephant” as “el-e-phant”. At the onset of “el”, a listener may activate a cohort of potential resolutions such as “elephant”, “electricity”, or “elder”, all of which may be considered competitors. With the subsequent “el-e”, words consistent with the received signal, such as “elephant” and “electricity” remain active competitors, while incompatible words, such as “elder”, are 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) (Tannenhaus 1995)[?]. 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 and fixations are recorded as this process unfolds, with the location of the participants’ eyes serving as proxy for which words/images are being considered. Any particular trial could have anywhere from 3 to 10 eye movements, known as saccades, followed by fixations lasting between 200ms-100ms. An illustration of these terms is given in Figure 1.

**Fixation curve and relation to TRACE:** How speech is perceived and understood has been a subject of much debate for a significant portion of psycholinguistic’s history. Starting in the 1980s and persisting today, most researchers subscribe to what is known as the connectionist model of speech perception. Briefly, this model posits that speech perception is best understood as a hierarchical dynamical system in which aspects of the model are either self reinforcing or self inhibiting with feedforward and feedback mechanisms. For example, hearing the phoneme \h\ as in “hit” will “feedforward”, cognitively activating words that begin with the \h\ sound. These activated words then “feedback” to the phoneme letter, inhibiting activation for competing phonemes such as \b\ or \t\. In 1986, McClelland and Elman implemented the TRACE model (TRACE doesn’t stand for anything – the name is a reference to “the trace”, a network structure for dynamically processing things in memory) implementing theoretical considerations into a computer model [?].

It was against simulated TRACE data that Allopenna (1998) found a tractable way of analyzing eye tracking data [?]. By coding the period of a fixation as a 0 or 1, depending on the referent and taking the average of fixations towards a referent at each time point (originally measured at intervals of 33ms, today measured at intervals of 4ms), Allopenna was able to create a “fixation proportion” curve that largely reflected the shape and competitive dynamics of word activation suggested by TRACE. Since this publication, it has been this transformation of the data that has guided the last 25 years of VWP research. An illustration depicting these curves, along with an example of a VWP trial, is presented in Figure 2.

**Research goals:** Some more context is helpful here to try and understand what exactly it is that researchers are trying to learn from this data. As mentioned previously, we are largely concerned with “activation”. In particular, though, we are often interested in how the activation of competing words compare. Specifically, we might ask, “at what point in the audio signal does the subject begin to identify the target word, relevant to competitors.” This is of special interest in studying language development in typically developed children and those with cochlear implants, where atypical children may require more of a signal before they are able to disambiguate what they are hearing. As such, it is often of interest to ask when and how activation differs both between target words and competitors, as well as between different subjects. It is largely these last two areas that have dominated much of the VWP research.

**A (mostly) irrelevant history:** Decades of papers chart the pursuit of relevant metrics that can be meaningfully asked of the data. These have included metrics such as AUC (area under the curve), mixed and hierarchical models, Metropolis-Hastings simulations and more. Typical of these attempts is either a lack of specificity (AUC can tell us *that* curves are different, but not *when*) or over specificity (contributing to large researcher degrees-of-freedom). One notable attempt was made by McMurray, where parameterized functional forms of the data were posited [?]. This allowed non-linear curve fitting algorithms

to generate meaningful parameter estimates of the data, such as the four-parameter logistic model which included parameters such as “crossover point” and the slope at the point of inflection. While this served to give functional form to the observed data, it was still subject to the same shortcomings of previous research in terms of specificity. In 2022, McMurray simulated eye tracking data generated by these parametric forms and showed that the curve fitting algorithms resulted in highly biased estimates [?]. In particular, he sought to have “true” data-generating curves that can create a distribution of possible subject curves. These curves, and the variance in the distribution, were generated empirically from previous studies (although these values were based on fixation assumption). A major weakness in this paper is the assumption of data-generating mechanisms, of which he performed four. It was noted in the paper that this was a weakness, but the argument was made that if recovery could not be successfully performed under moderate assumptions of data generation, then attempts with far more complex “real data” would be hopeless.

**The algorithm:** To understand what follows, it’s helpful to have a brief idea of the algorithm suggested by McMurray. Effectively, the algorithm for generating “looks” was this: At some time  $t$ , a value is drawn from a binomial distribution (or a curve at time,  $t$ , representing the trajectory of probability) and asks: what is the probability that a saccade launched at this time would land on the target object. Depending on this outcome, a duration of fixation would be randomly drawn next: if the saccade was determined to land on the “target”, the fixation would be long; if not, it would be shorter. At the end of each fixation, another saccade was launched. Critically, the time  $t$  as which this is determined does not occur at the END of the present fixation, but rather at the beginning. This is to account, in part, with oculomotor delay shown in Figure 1. In other words, once an eye moves, it is already premeditating where to move next. Additionally, once that decision is made cognitively, it is usually another 200ms before it is launched.

**The insight:** In short, we begin by recognizing that the actual object of interest is the underlying activation mediated through some cognitive mechanism upstream from eye movements. If we are interested in this mechanism as it unfolds over time, we should sample from this mechanism. As such, the only unit of observation that represents a sample in time is that of the launch of a saccade: past that point, and until the end of the fixation, we are receiving no more additional information about the underlying activation, with the exception of whatever can be gleaned by the length of the fixation. At any rate, it is at least incorrect to treat data launched at the saccade and the ongoing fixation as the same information relying the underlying cognitive mechanism. The solution to this, then, is to simply use the saccade data, omitting information collected from the fixation period. While this seems obvious now, a visualisation of saccade information would have been nothing more than a discrete collection of 0s and 1s, having no visual similarity to the TRACE simulations. *However*, taking the assumption presented by McMurray in 2010, we can still fit non-linear curves to this data, resulting in estimates of probabilities that match those predicted by TRACE.

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\**.

[insert picture here]

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.

[insert picture]

We must also consider the phenomenon of oculomotor (OM) delay; whereas a particular physiological response (such as eye movement) may be recorded at some time,  $t$ , the latent cognitive mechanism guiding this decision occurs at 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.

### 3 Looks

The figure above represents a more detailed composition of what we are calling a \*look\*, made up of the saccade and subsequent fixation. We see that we begin with some latent cognitive activation informing the decision to direct the eyes to a particular location. The probability that this direction is towards the target at a particular point in time is what we are referring to as our activation,  $f_{\theta}(t)$ .

Once this decision has been made, there is oculomotor delay. We don't know what it is, but the previous document illustrated scenarios in which it was known, fixed, and unknown/random. As the cognitive action is the particular instance that we are interested in modeling, the observed bias will be the time between this mechanism and the subsequent launch of the saccade. This period of time is  $\rho(t)$ .

As Bob pointed out, the saccade is ballistic and averages periods around 30-70ms. Given the ballistic nature of the saccade, it's intended destination is determined at onset and remains the same for the duration of the saccade. As such, the information received at the offset of the saccade (where it lands) is the same as it was at onset; it can't be changed. We will be more consistent in our measurement if we take the onset to represent our discrete instance of the saccade.

To see this notationally, let  $t_j$  be the time the time of cognitive action for look  $j$ . If  $\rho_j$  represents the oculomotor delay for this look, the onset of the saccade will be observed at  $t_j + \rho_j$ . If  $r_j$  represents the refractory period for the  $j$ th saccade, the saccade will end at  $t_j + \rho_j + r_j$ . Here, both  $\rho_j$  and  $r_j$  are random variables. As our indicator for fixation on target is the same at  $t_j + \rho_j$  as it is at  $t_j + \rho_j + r_j$ , by taking our measurement at  $t_j + \rho_j$ , we remove the additional variability introduced by  $r_j$ .

## 4 Data use methods

Somewhere should have something about relationship to TRACE – which sort of begs that I figure out how to compare my method with TRACE (Allopenna 1998)

### 4.1 Proportion/Averaging

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.,  $t = 0, 4, \dots, 2000$ ), 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.

## 4.2 Saccade

If we are to consider eyetracking data samples from some probabilistic 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 mechanism 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

$$s_j \sim \text{Bern}(f_\theta(t_j)).$$

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.

### 4.3 “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.

## 5 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:

[insert image]

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

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

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.

## 6 Looks in action

We can see how these looks play out on the observed plot for the unknown/random delay scenario. The underlying curve,  $f_{\theta}(t)$  determines probabilistically if a saccade launched at  $t$  will fixate on the target. Between this cognitive action and the physiological movement, there is a delay, represented by the horizontal shift at each time point between  $g_{\theta}(t)$  and  $f_{\theta}(t)$ . As we can see above, this horizontal shift is not equal across all points, with some areas showing larger degrees of oculomotor delay.

The duration of the saccade itself, marked in pink, occurs after the saccade has been launched. As laid out here, I don't think that the duration of this movement is of much significance. That is, I suspect that the factors dictating the duration of the saccade are more influenced by external factors (length of distance from previous location to next, individual variation) rather than by any measure of underlying activation.

Based on this, I believe that choosing as our observed data the moment of onset as a discrete measure of the saccade is appropriate, as it conveys the same relevant information as the full movement, with the additional benefit of being more consistent across observations.

As a follow up to Bob's other comment, it does seem to make some sense to refer to  $f_{\theta}(t)$  the latent activation curve (as that's what it is), while saving "saccade curve" for  $g_{\theta}(t)$  (which is what that is). The relationship between the two can then be described as  $g_{\theta}(t) = f_{\theta}(t - \rho(t))$ .



## 7 Simulations

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.

[image]

### 7.1 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 oculomotor delay.

### 7.2 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.

### 7.3 Unknown Random Delay

The final scenario for consideration involves a situation in 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 interrogate 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.

## 8 Discussion

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 oculomotor 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.

## 9 Conclusion