Eye-Tracking and Auditory Cognitive Learning

Xiaoyi Fei, Emma Friedman, Lindsey Yao

Department of Statistics, Carnegie Mellon University, Undergraduate Research Symposium 2007

Introduction

Have you ever wondered what subconscious thinking processes are occurring while you hear and consequently identify a sound? Identification of a target object is often evident via eye movements even before the target word is finished being spoken. Thus, utilizing a device that tracks eye positions and movement, a team from the Carnegie Mellon University cognitive psychology department, led by Dr. Kaori Idemaru and Dr. Lori Holt, is gathering information on this process from the path the eye takes as a subject focuses on a computer screen while deciding how to identify the auditory stimulus presented. The primary purpose of the team's investigation is to understand how to encourage more efficient auditory cognition and discover a new way to test the auditory capacity of hearing impaired persons. By running the data in the statistical software R, our analysis research group has created regression functions that model the data collected. These models estimate the likelihood that the experiment subject will look toward the appropriate response portal given a specific type of auditory stimulus.

Data Collection Method

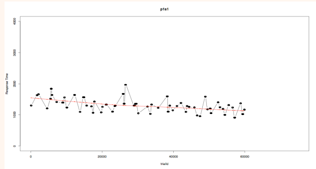
There are two square frames on the right and left sides of the computer screen, called the 'right portal' and 'left portal'. The left portal corresponds to a response of 'd', and the right portal corresponds to the 'g' response. In our study, the right portal corresponds to the /da/ sound, and the left portal corresponds to the /ga/ sound. The two portals never switch assignments.

At each trial, the subjects are presented with either a /da/-type sound or a /ga/-type sound. The degree to which they are distinctly /da/ or /ga/ varies along a five point spectrum. The strongest /ga/ sound is labeled as stimulus 1, and the strongest /da/ sound is labeled stimulus 9. There are three intermediary stimuli: 3, 5, and 9, representing sounds that are neither distinctly /da/ nor /ga/. The target stimulus number for each trial is recorded as *Stimulus*.

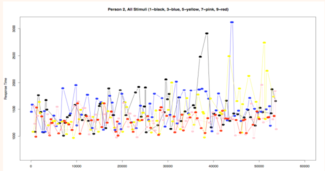
Over the course of the experiment, the subjects hear a sequence of 150 of these stimulus sounds and are instructed to categorize these sounds by clicking on one of two buttons, labeled "d" and "g". In between trials, the location of the subject's eyes is normalized by requiring the individual to fixate on a central point on the computer screen for two seconds. Once this has been accomplished, a new stimulus is presented and the trial continues. The data collected is a stream of (x,y) coordinates of the subject's left and right eyes. Data is collected every 20 milliseconds.

Preliminary Graphical Results

Our first goal during the preliminary graphical analysis stage was to determine if there was learning from trial to trial. Learning is defined as a shortening of the response time from the beginning of the experiment to the end per person per stimulus. A graph of the response times of every trial where stimulus one was presented, and an imposed trend (lowess) line, shows that learning did occur.

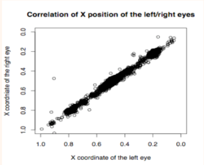


Although this is only a graph of stimulus one, you can see that learning did occur on all stimuli by plotting the response times of all stimuli: 1, 3, 5, 7, and 9.



Our second goal was to find a function that would determine, at any time, the probably of the subject looking at a specific portal. We chose the left portal, and wanted to find the probability, given the number of times the subject had already been presented with the stimulus (because learning *does* occur), that the subject would look towards the left portal.

We decided to determine this function based only the x-coordinate position of the left eye after finding a strong correlation between the left and right eye.



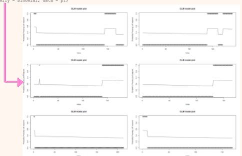
We decided to disregard the y-coordinates because they did not affect whether the subject is looking at the left or right portal.

Functional Analysis

Our functional analysis serves two goals: first, to predict whether a subject will respond with 'd' or 'g', and second, to predict the probability of looking left, at any time. The input variables in both functions are the current time, and the number of times the subject has been presented the stimulus of interest.

```
glm(formula = leftkey ~ left + trial_frac * as.factor(Stimulus) +
    num_pres * as.factor(Stimulus), family = binomial, data = p1)
```

This graph predicts the subjects response, 'd' or 'g', at the end of the trail given time



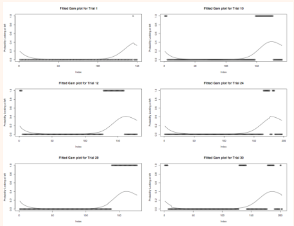
The above function predicts whether the subject will respond 'd' or 'g' at the end of the trial, while the below function predicts the probability that the subject is looking left at any time during the trial.

Stimulus 1

```
Call:
gam(formula = left ~ s(trial_frac) + num_pres, family =
binomial,
    data = p11)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.02825  -0.39246  -0.13911  -0.08622   3.48299

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)   -3.214768   0.123002 -26.136  < 2e-16 ***
s(trial_frac)  2.699893   0.135650  19.903  < 2e-16 ***
num_pres      -0.012102   0.002126  -5.694 1.24e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7002.0  on 11835  degrees of freedom
Residual deviance: 4920.4  on 11830  degrees of freedom
AIC: 4932.4

Number of Fisher Scoring iterations: 12
```

Graphs of Subject One's Stimulus One GAM Function

All variables are strongly statistically significant



Using the GAM function we are able to predict, at any time, the probability of the subject looking left, given the number of times that stimulus has been presented and the time.

Conclusions

This current study observes the participant responses to sound categories, specifically /da/ and /ga/ sounds. Following research on an individual's ability to discern sounds into categories based on audio qualities, it would be interesting to apply the techniques of this study to researching an individual's ability to discern sounds into categories based on visual qualities. For instance, testing two categories – animals and non-animals. The participant would hear variations of words sounding similar to frog/fog, bear/bar, ram/ran, etc. This study would further the understanding of cognitive learning by tying in another sensory stimulus. Another variation of this experiment could be to show two images, and train the participant to associate similar sounds to each of the images. This could study how the participant develops an anticipatory response to the sound being played and to which image this anticipatory response corresponds

In terms of the statistical research for this study, we were able to pick out distinct trends in behavior according to stimulus type. We were also able to derive models that can predict the probability, at any given time within the trial, that the subject will be looking towards the left portal (representing a 'd' response). From here, further research and calculations should be conducted to aggregate these functions. This would provide a model for predicting how likely a new subject is to look towards a specific portal, given different stimuli, and how often the subject has been presented with that stimulus.

Literature cited

Richardson, D., Dale, R., & Spivey, M. (in press). *Eye movements in language and cognition: A brief introduction.* In Gonzalez-Marquez, Coulson, Mittelberg, & Spivey (Eds.) Methods in Cognitive Linguistics. John Benjamins.

Acknowledgments

We would like to thank Lori Holt and Christi Gomez for allowing us to work on this project with them. We would also like to thank Brian Junker and Jim Delaney for all their time, patience and advice.

For further information

Please contact *lholt@andrew.cmu.edu.*