# Methods

## Introduction and theory of Spivey experiment

The research by Spivey et al. (2005) was inspired by results from research into spoken-word recognition indicating that “partial activation of multiple lexical representations […] cascading to later stages of processing even just part of the way through hearing a word” (Spivey et al., 2005). Thus, they wanted to investigate whether the judgment of spoken-word recognitions was a dynamic process with information provided to later stages in process before a conclusion has been reached, or if the process is more discrete stage-based with each step having to be completed before information is passed on, e.g. passing information before or after the correct lexical item have been identified. So the experiment is basically an investigation into whether spoken-word processing is the result of a dual-system framework or dynamic framework.

The experiment emphasised the use of mouse-tracking over eye-tracking, since eye-tracking will only be a semicontinuous account of the underlying cognitive processes since it is based on several steady fixations on an object. Mouse-tracking “provides an unusually high-fidelity emission of the continuous cognitive dynamics inherent in real-time spoken-language processing” (Spivey et al, 2005).

The objective of the experiment is therefore to observe the pulling effect of a distractor object on the mouse-cursor movement, making the mouse-trajectories curve towards it in a graded fashion as opposed to a quick sharp angle like shape. This is done to show whether the underlying process follow a dynamic framework or dual-system framework. This is done through comparing two kinds of experimental trials: a control trial, where two target objects are phonologically dissimilar, and a cohort trial (the name given to the experimental trial by Spivey et al. (2005)) with two phonologically similar objects.

Notes:

This should be tidied up; a lot of the information comes twice.

Should we introduce a bit more about pronunciation of words and multiple parallel partial activation of lexical representations?

## Experiment – OpenSesame

The experimental setup and design have been approximated to the design of Spivey et al. (2005), but some details will differ. These details will be highlighted in the explanation.

The experiment was a repeated binary forced-choice selection task, where participants were presented with two visual stimuli on a screen (one in the top right corner and one in the top left) and one auditory stimulus in a pair of headphones. Their objective was to click with the computer mouse on the visual stimulus corresponding to the auditory stimulus, e.g. hearing the word “carrot” and clicking on the picture of a carrot. For us to replicate the original experiment as close as possible, the pictures of objects were provided by the original researchers (shout out to The Spiveyman) to make sure that the visual stimuli were the exact same as in the original. An example of one trial can be seen in figure X. The design details of each trial can be seen in figure X. Between each trial the mouse-cursor was automatically reset to the fixation dot on the screen positioned in the bottom centre (should we give exact coordinates?).

Participants received thorough instructions and went through 4 practice trials to familiarise themselves with the design. After this the 8 pairs of phonologically similar objects were used to create 32 cohort trails and 32 control trials with random order of trials for each participant.

Demographics (i.e. gender, age and level of English) were gathered from all participants. In the experimental trials x- and y-coordinates of the mouse cursor were gathered alongside their corresponding timestamps at a sampling rate of 100 Hz (corresponding to one sample each 10 ms). Other important measures gathered in each trial were response time, type of each trial, the presented stimuli and whether each response was correct or not.

Pilot studies showed that some of the stimuli used in the original experiment were unfamiliar to our group of participants, e.g. “beaker” and “racket”. Therefore, to familiarise participants with the words and corresponding pictures, all the visual stimuli alongside their name spelled out were presented to the participants before the experiment started. We have assessed that this will have no effect on the experimental trials (include reason? No practice effect?).

Graphical user interface, application

Description automatically generated

Graphical user interface

Description automatically generated

Our experimental design differs from Spivey et al. (2005) in two important aspects. In the original experiment participants had to click a “Start” button in the bottom centre to start each trial. In our experiment each trial started automatically after 500 ms of the fixation dot screen (do we have to give a reason for this?). The second aspect is related to the presenting of the auditory stimuli. In the original experiment the auditory stimulus was presented 500 ms after the visual stimuli, while in our experiment they were presented simultaneously (again, should we provide a reason?).

Also, we are not aware of any practice trials taking place in the original experiment, so this might also be a difference, however not that important.

The experiment was coded and developed with the use of the software OpenSesame (Mathôt et al., 2012) and mouse-tracking in OpenSesame was made possible by the Mousetrap plugin for OpenSesame (Kieslich & Henninger, 2017). As previously stated, all 16 visual stimuli were provided by a researcher of the original experiment to approximate the effect from the original experiment. All auditory stimuli were created using Amazon Polly to make a set of stimuli with an objective and homogenous pronunciation(reference?). Mean length of all auditory stimuli were 0.63 seconds ranging from 0.50 to 0.78 seconds. The screen size used for presentation of stimuli was 1024x768 pixels. A Logitech M510 mouse with a Dot Per Inch (DPI) of 1000 was used with mouse-acceleration turned on and mouse-sensitivity set to the fifth value from the left on the control panel for mouse settings on a computer running macOS Catalina (reporting of mouse settings has been done according to good standard practice suggested by Fischer & Hartmann (2014)).

Sampling of the x-, y- coordinates? (100 Hz sampling rate (one recording each 10 ms))

## Participants

In the experiment 13 people participated. 9 identified themselves as female and 4 as male. They were primarily university students studying Cognitive Science, but not exclusively. Their age ranged from 20 to 25 years old with a mean age of 22.17 years. Their self-evaluated level of English on a scale from 1 to 10 ranged from 5 to 9 with a mean of 7.07. From this we assessed all participants to have a good enough understanding of the English language to be included in the analysis. No one differed significantly from the others regarding accuracy, where the best performing participants had an accuracy of 100% (64 correct out of 64 trials), and the worst performing participants had an accuracy of 96.9% (62 correct out of 64 trials). Reaction times were normally distributed with no participant as a clear outlier. Consent was gathered from all participants.

## Analysis

### Replicating Spivey et al.

T-test

* Between percentage of incorrect trials
* between total response time
* between time\_initiation\_movement
* between duration\_of\_movement (x)
* between x-coordinates in cohort and control for specific time slices
* between proximity to target and distractor object (for both control and cohort)

Degree of curvature

* Bimodality coefficient
* Kolmogorov-Smirnov test

Area under the trajectory

* Kolmogorov-Smirnov test

To check for whether our experiment have produced similar data and results as Spivey et al. (2005) a range of the same tests and a visual inspection of the average trajectories have been done. The following statistical analyses have been done:

* Paired samples t-test on percentage of incorrect trials between control and cohort trials
* A linear mixed effects model with total response time of a trial predicted by type of trial (control vs. cohort)
* A linear mixed effects model with initiation time of a trial predicted by type of trial (control vs. cohort)
* A linear mixed effects model with duration time of actual movement predicted by type of trial (control vs. cohort)
* Calculation of the bimodality coefficient based on within participant standardised area under the curve scores for both control trials and cohort trials
* A visualisation of averaged trajectories for control and cohort trials

The results of each test are shown in figure X alongside the corresponding results from Spivey et al. (2005). A general comment to the results is that since the amount of participants of our study is lower compared to the original, the p-values of each analyses will not be reproduced exactly.

All analyses produced comparable results to the original experiment except the linear mixed effects model, where duration time of movement was predicted by trial condition. The mean of each trial type indicates the same tendency as seen in the original experiment, but it did not produce a significant p-value. This is possibly due to a lack of statistical power caused by too few participants in the replication study.

A special focus should be put on the bimodality coefficient. Since the coefficient found in our study is below 0.555, it still indicates unimodal distributed trajectories, reproducing the most important result from the study by Spivey et al. (2005).

In general, since all other types of analyses produced comparable results to the original experiment, we will conclude that the data produced is close enough to the original to be used for the purpose of this paper.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Test | Spivey control mean | Spivey cohort mean | Spivey p-value | Our control mean | Our cohort mean | Our p-value |
| Incorrect trials | 2.5% | 8.9% | P < 0.001 | 0.4% | 2.7% | P < 0.05 (p = 0.01184) |
| Total response time | 1717 ms | 1812 ms | P < 0.001 | 1197 ms | 1240 ms | Paired t-test: p < 0.01 (p = 0.003)  Mixed effects model: p < 0.05 (p = 0.0153) |
| Initiation time | 357 ms | 335 ms | P > 0.1 | 364 ms | 381 ms | Paired t-test: p > 0.1 (p = 0.49)  Mixed effects model: p > 0.1 (p = 0.346) |
| Duration time | 1360 ms | 1477 ms | P < 0.001 | 833 ms | 859 ms | Paired t-test: p > 0.05 (p = 0.3014)  Mixed effects model: p > 0.05 (p = 0.279) |
| Degree of curvature (possibly AUC) | -0.165 | 0.164 | Bimodality coefficient:  Control: 0.366  Cohort: 0.381 | -0.178 | 0.182 | Bimodality coefficient:  Control: 0.430  Cohort: 0.365 |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

### Classifier Performance

Preprocessing:

* Filter for practice and incorrect trials
* Velocity and acceleration calculation
* Common starting point?
* Normalising coordinates?
* Time normalisation to 101 bins
* Standardisation
* PCA

Notes:

Should only the analysis for replicating Spivey be included here? Or should the whole shebang about PCA and LDA also be in this section?

Notes:

Reported based on best-practice assessment by Schoemann et al., 2021.

What test should be used to evaluate reproduction of Spivey results?

Should we write that we encouraged participants to be quick to reduce participants waiting, finding the right answer and then move in a straight line?

Spivey et al. look at mouse movement initiation time. Should we do the same? It could probably tell us something about whether people waited to move the cursor until the word has been fully pronounced?

If some of the theory from either negation or Spivey experiment seems redundant, maybe just use “outside the scope of this paper”???