The Effects of Visual and Auditory Cues on Human Reaction Time

By: Dylan Demo, Jess Joblin, and Josh Lapish

STS 3250: Design / Analysis of Experiment

1. INTRODUCTION

In this experiment an online computer program was used to compare reaction time in response to various methods of stimulus (visual, auditory, and a combination of the two). Participants were administered tests in a randomized order and their reaction times were recorded to find the mean reaction time across each method of stimulus. The response measured in the experiment was the time in milliseconds it took for the participant to execute a click on a mousepad after being exposed to the appropriate stimulus for the given test. The experiment utilizes a Complete Block Design (CB[1]), with the blocks being each of the twelve participants and the experimental factor being the three levels of stimulus used to test reaction time. This design was most appropriate for the goal of this experiment because reusing the participants across all treatments reduced variation by turning potential nuisances into a factor of the experiment, so that any variation in individual response time was controlled by using the same participant for each level of the treatment factor. The experiment aims to look at how auditory and visual stimuli affect mean reaction time, and the question being investigated is which method of stimulation results in the fastest mean reaction time across all subjects.

2. METHODS

2.1. Experimental Design

Elements of the Experiment

In this Complete Block Design (CB[1]) experiment, there are a variation of elements that were defined to successfully run and analyze the tests. 12 participants were used in this experiment to test their reaction time and each was provided a different order of the three tests to take. For example, one participant's order of tests would be visual, auditory, combined and another participant's order of tests would be auditory, combined, visual.

Experimental Units	Participant's Time Slot (36)
Blocks	The participants (12)
Factors	Type of Test (1)
Levels of Factors	Visual, Auditory, Combined (3)
Type of Design	CB[1]
Response	Mean Response Time
Number of Participants	12
Units	Milliseconds

When recording our data, we had three time slots (visual, auditory, and combined) and had twelve participants. With this information in hand, we know we have 36 experimental units, and they are representative of the time slots that were in the experiment. There is one experimental factor in this experiment, which is the type of test the participant will participate in. There are three levels to this factor, which are visual, auditory, and combined. All three of these levels are used to test the participants reaction time when using different senses. The blocks in this experiment are the individual participants. In order to remove possible biases or variability from this block, all participants will be reused. This insinuates that each participant will take each of the three tests, which are our levels. To remove additional biases, the order of the levels for the experiment will be randomly assigned to each participant. Our goal is to test how the mean reaction time differs between the three levels of our factor. The simulation generated the output in seconds, however, we converted all numbers to milliseconds.

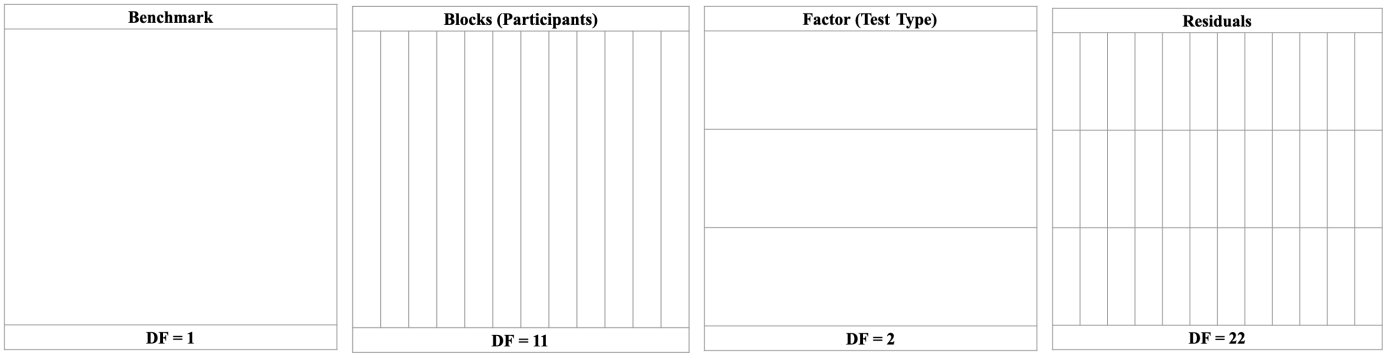
Pilot Study Impact

The pilot study allowed us to assess the pros and cons of our initial protocol. While the website's simulation worked well and provided us with the data we needed, we decided to completely alter the way in which we set up our experiment and how we actually tested each participant. After conducting the pilot study, we realized that a lot of bias was introduced because each participant used a different laptop, they were all in different environments, and they completed the experiment at different times in the day. Our goal was to reduce the possibility of biases and variability as much as possible, so we changed quite a few aspects to our protocol to make the experiment run more smoothly. First, we made sure that all conditions among the participants were equal. This means that only one of our laptops was used to collect our results. To decrease variability, we made

sure the brightness and volume of the laptop remained the same for each participant and we removed all outside noises by going to a separate, empty classroom. Additionally, we ensured that each participant had their finger resting on the mousepad to remove any variability of hovering over the trackpack.

The biggest adjustment we decided to make to this experiment was how we determined the order of the tests. Rather than using a coin, we wanted to find a method that would help randomize the 12 participants into 6 total categories of level orders, ensuring that each group of level orders had exactly 2 participants. After considering several options, we decided to put 12 slips of paper in a bag, all representing a different level order. For example, one slip may say "visual, auditory, combined", whereas another slip could say "combined, visual, auditory". When each participant entered the experiment room, they selected a piece of paper from the bag and that would be the order in which they would take the test. We made sure that the participant did not put the paper back in the bag to be reused. This will ensure that every participant does not take the combined test last because we want all levels to be in a random order.

Factor Diagram



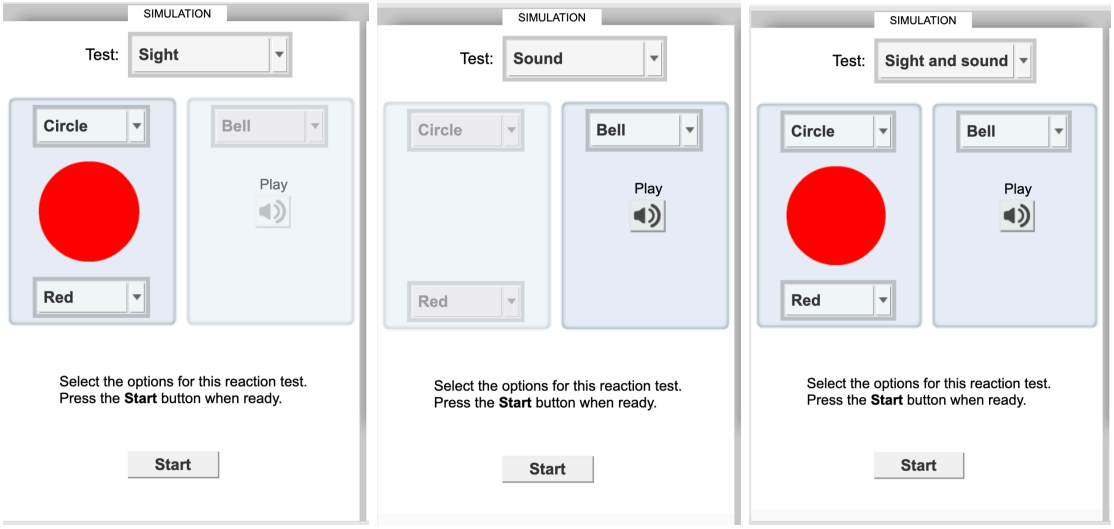
2.2. Materials and/or Participants Information

Materials	Participants
Jess's laptop Order of tests written on slips of paper (each order repeated twice for a total of 12 slips) a. 1 2 3 b. 1 3 2 c. 2 1 3 d. 2 3 1 e. 3 1 2 f. 3 2 1 Bag to put slips of paper in Website: https://gizmos.explorelearning.com/index.cfm?method=cResource.dspView&ResourceID=43	12 college students enrolled at Elon University 1. Gabl Sable 2. Brandon Gottlieb 3. Miles Dinou 4. Jack Lippman 5. Mara Waskiewicz 6. Cayla Bellamy 7. Alex Campbell 8. Lindsay Sawyer 9. TJ Remillard 10. Jillian Mendoza 11. Caleb Soo Hoo 12. Olivia Archer

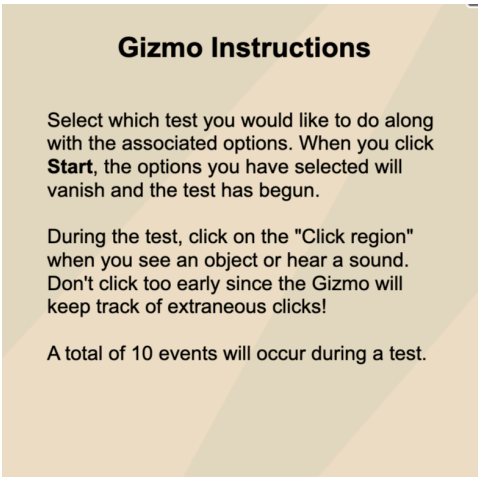
2.3. Procedure

For the experiment, each participant was randomly assigned an order in which to perform the three tests. The participants then entered a room with the experimenter, with the environment being controlled by having the lights on and no outside noises or distractions. The same laptop and settings (screen brightness and volume) were used when testing each participant. Each participant was then given the same instructions by their experimenter and then directed to take each of the three tests in the randomly assigned order previously determined. The reaction times for each attempt were recorded as well as the average and standard deviation for each of the three test types for each participant.

Test Type Screenshots From Website



Gizmo Instructions for Test Simulation



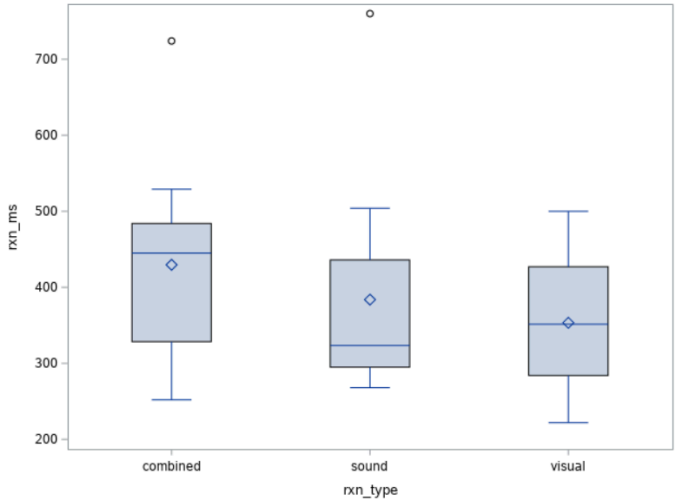
3. RESULTS

3.1. Numerical and Graphical Descriptive Statistics

Summary Statistics					
TEST TYPE	**N**	**MEAN**	**STD DEV**	**MIN**	**MAX**
Visual	12	353.33	85.67	222	500
Auditory	12	383.83	141.36	268	760
Combined	12	429.67	129.03	252	724

Summary Statistics - Grand Mean				
N	**MEAN**	**STD DEV**	**MIN**	**MAX**
36	388.94	121.78	222	760

Test Type vs Reaction Time Box Plots



**Blue diamond represents sample mean*

The tables shown above depict the summary statistics for the experiment. As we can see, the mean for the combined test is much greater than the visual and auditory means, with the visual having the smallest average. Despite this, the auditory test type has the greatest standard deviation, while the visual test type has the smallest standard deviation. Simply looking at these results, we can see that the mean reaction time is the smallest for those who take the visual test.

However, when we look at the box plots, all three look relatively the same and have similar spreads. It is important to note that the box plots account for two outliers, one in the combined test (724) and the other in the sound test (760). Because of these two outliers in the experiment, our summary statistics are partially skewed since those numbers were added into the average. Looking at the boxplots, it is much more difficult to find a difference in reaction times between the three tests.

3.2. Decomposition of the Data

Benchmark
388.94
"
"
DF = 1

Blocks (Participants)										
181.7	33.4	-66.3	-80.3	-26.6	151.7	44.4	-29.6	-98.6	-114.9	3.1	2.1
"	"	"	"	"	"	"	"	"	"	"	"
"	"	"	"	"	"	"	"	"	"	"	"
DF = 11											

Factor (Test Type)
-35.61
-106.11
40.73
DF = 2

Residuals											
34.06	-68.94	-141.94	-166.94	-61.94	5.06	111.06	-12.94	-105.94	-103.94	42.06	43.06
371.06	94.06	-9.94	0.06	-54.94	115.06	-83.94	-78.94	-75.94	-103.94	-112.94	-120.94
140.06	75.06	-46.94	-73.94	37.06	335.06	106.06	3.06	-113.94	-136.94	80.06	84.06
DF = 22											

Interpretation of Top Left Residual
For participant #1, the visual test measured 432 ms. This observation is 34.06 ms higher than the grand mean reaction time of 388.94 ms. This difference is how much variability we attribute to random chance for this particular observation.

Interpreting Effects
Generally speaking, factor diagrams are used to measure variability across multiple elements of an experiment. After collecting results, a factor diagram provides estimates for treatment and block effects by comparing their averages with the overall average of the response. For example, in order to interpret the results of individual participants, we can look at the "Residuals" factor diagram. Each section represents an individual participant and the number represents how much their reaction time differs from the grand mean, with a negative number insinuating it was below average, and a positive number displaying the individual was above average.

3.3. Formal Analysis

3.3.1. ANOVA Table

SOURCE	SS	df	MS	F	F*	P
Benchmark	5445875.65	1	-	-	-	
Participants	283499.89	11	25772.72	2.83	2.258	.0181
Test Type	35430.89	2	17715.44	1.95	3.443	.1665
Residuals	200159.11	22	9098.1414			
TOTAL	5964965.54	36				

Factor: Test Type

Hypotheses	Test Statistic	P - Value
$H_0 : \mu_1 = \mu_2 = \mu_3$ H_a: $\mu_i \neq \mu_j$ for at least one pair i, j	1.95	0.1665

For the experiment, we tested the null hypothesis that all test types would have the same mean reaction time against the alternative hypothesis that at least two test types result in different mean reaction times. Our F-Test statistic of 1.95 tells us that the average variability for test type is 1.95 times greater than the average variability in the residuals. With a p-value of 0.1665, there is insufficient evidence to suggest a test type effect on mean reaction time.

Block: Participants
With a p-value of 0.181, there is sufficient evidence to suggest a block effect on mean reaction time. What this means for the population is that despite different methods of stimulus not having an effect on reaction time, reaction time for individuals across all types of reaction time testing is significantly different.

3.3.2. Confidence Intervals

Test Type Comparison	Simultaneous 95% Confidence Intervals
Combined vs Sound	(-55.07, 146.74)
Combined vs Visual	(-24.57, 177.24)
Sound vs Visual	(-70.40, 131.40)

**Order of subtraction from left to right; ex: "combined vs sounds" = combined mean - sound mean*

All confidence interval comparisons contain 0, therefore we do not have sufficient evidence of a difference in mean reaction time between any of our levels. To reduce the chances of a type I error, we used Bonferroni's adjustment method to start at a higher family wise confidence level; this is calculated by taking our target confidence level for each comparison (95%) and raising it to power of 1/I, in which I is the number of comparisons among our test types. In this case, we started at a family wise confidence level of $.95^{\wedge(⅓)}$ or 98.30%.

For the blocks, we used the less-conservative Tukey adjustment method due to the high number of pairwise comparisons. There totaled 66 pairwise 95% confidence intervals (12 choose 2), and surprisingly only one of which turned out to be significant:

Participant # Comparison	Simultaneous 95% Confidence Intervals
1 vs 2	(13.37, 579.96)

**1 vs 2 = participant 1 mean - participant 2 mean*

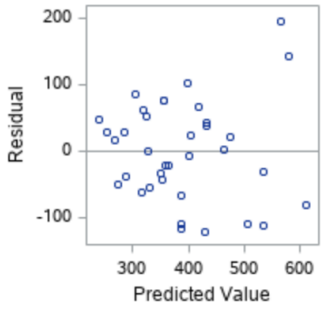
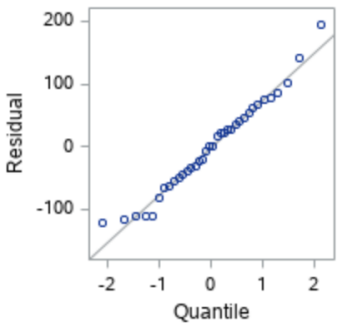
Since this interval does not contain 0, we have sufficient evidence of a difference in mean reaction time between participants 1 and 2 by between 13.37 and 579.96 milliseconds.

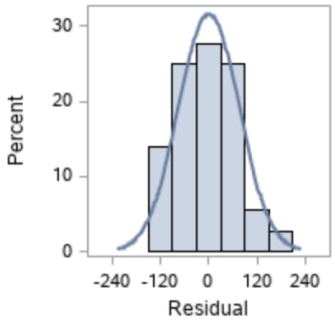
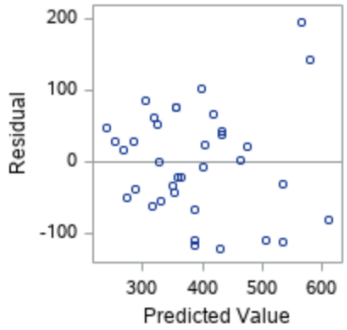
3.3.3. Contrast

Contrast	$\frac{1}{2}\mu_V + \frac{1}{2}\mu_S - \mu_C$
Point Estimate	-61.08
Confidence Interval	(-131.02, 8.85)
Hypotheses	H_O: $\frac{1}{2}\mu_V + \frac{1}{2}\mu_S - \mu_C = 1$ H_A: $\frac{1}{2}\mu_V + \frac{1}{2}\mu_S - \mu_C \neq 1$
F-Test Statistic	3.28
P-Value	0.0838

One contrast we performed compares the mean reaction time of sound and visual to combined (sound and visual mean - combined mean). The 95% confidence interval for this comparison is (-131.02, 8.85). Since our interval contains 0, we have insufficient evidence of a difference between the mean reaction time of sound and visual vs the mean reaction time of combined. This is also indicated by the large p value of .0838, which is larger than our alpha (.05).

3.3.4. Assumptions - Jess

<table>
<tr><th colspan="2">Assumption</th><th>Graph</th><th>Explanation / Conclusion</th></tr>
<tr><td>S</td><td>Errors have the same standard deviation</td><td> 
</td><td>Using the residual plot to compare the spread of residuals versus the predicted value, we can see that the points are equally spread above and below the x-axis.</td></tr>
<tr><td>I</td><td>Errors are independence</td><td colspan="2">Because we randomized the order in which the participants took the tests, time order cannot play a role in creating dependencies. Therefore, we have no reason to believe this assumption is not met.</td></tr>
<tr><td>N</td><td>Errors follow a normal distribution</td><td colspan="2">Since the sampling distribution for our test statistic relies on the errors following a normal distribution, we need to check this assumption in SAS. This can be done by looking at two graphs in the “Fit Diagnostics”. The residual vs. quantile plot approximately a 45° line. The histogram of residuals is approximately normal. Because of these two observations, we have enough evidence to assume the errors follow a normal distribution.
 


</td></tr>
<tr><td>A</td><td>Effects are additive</td><td rowspan="2"> 
</td><td>We will look at the same graph as we did for the equal variance assumption. Since there is not a pattern (ex. curve, line, u-shape), we do not have evidence of a violation of the additive assumption. We want the effects to get combined by adding because ANOVA models can only be used to estimate effects that are additive.</td></tr>
<tr><td>C</td><td>Unknown effects are constant</td><td>Since ANOVA models can only be used to estimate effects that are constant, we need to check this assumption. Looking at the same residual vs. predicted value plot, we do not see a pattern, which gives enough evidence to say the unknown true values are constant.</td></tr>
</table>

Since all of our assumptions are met, we have evidence that the results we collected in our experiment and the conclusions we made are valid.

4. DISCUSSION

Summary of Experiment

Overall, most of our hypothesis testing did not lead to significant results. In our ANOVA, our blocks (participants did contain significant results (p value = 0.0181), which indicates that we do have sufficient evidence of participants affecting the mean reaction time. Logically this makes sense; reaction time across participants is bound to experience variability. Though in our simultaneous 95% confidence intervals, only one pairwise interval out of 66 experienced significant results. The rest of the hypothesis testing did not indicate any significant results, which seemed rather surprising. Especially regarding the "visual vs sound" comparison, we hypothesized that participants would react faster to sound than visual, as this phenomenon has shown true in other experiments on reaction time (Shelton 2010).

Reflection

After reflecting upon our experiment, we believe our protocol ran very smoothly. The pilot study helped us tremendously because we were able to detect any flaws in our experiment that would increase the biases and variability. Because of this, we changed the way in which we actually collected our results, which ensured that all participants were under the same conditions. There were very few limitations we encountered throughout the experiment. After discussion, the only possible limitation we listed was the students that were selected to participate in the experiment. Since we used a convenience sample, we could only ensure that all students attended Elon University. It would have been interesting to use a sample of students that all were similar. For example, if they were all the same gender, same age, same major, or same GPA, the results may have differed or we could have seen less variability.

Future Suggestions

More improvements in our study could have better accounted for more nuisance variables, particularly regarding the structure of our procedure. One nuisance variable we could have eliminated was potentially the nervousness of the participant. Though the participant was not pressured in any way prior to taking the reaction test, the presence of the experimenter in the room could have affected their reaction time. To avoid this, we could have the experimenter leave the room before the participant takes the test. Another nuisance variable we could eliminate was simply the participant's understanding of the directions. Though explained thoroughly in a paragraph, the participant would likely have experienced better reaction times if they tested out a tutorial of each reaction test type before starting the actual tests.

5. REFERENCES

Shelton, Jose. Gippsland Physiotherapy Group, Melbourne, Australia; 2VIT University, Vellore, India. 2010. Retrieved December 2, 2021, from Comparison between Auditory and Visual Simple Reaction Times (scirp.org)

Sight vs. sound reactions Gizmo. Gizmos ExploreLearning. (n.d.). Retrieved December 1, 2021, from https://gizmos.explorelearning.com/index.cfm?method=cResource.dspView&ResourceID=43.

6. APPENDICES

Appendix 1 - Raw Data

Collected Data

ID #	Participant Name	Order of Tests	1. Visual Mean Response Time	2. Auditory Mean Response Time	3. Combined Mean Response Time
1	Gabl Sable	1 2 3	423	760	529
2	Brandon Gottlieb	1 3 2	320	483	464
3	Miles Dinou	2 1 3	247	379	342
4	Jack Lippman	2 3 1	222	389	315
5	Mara Waskiewicz	3 1 2	327	334	426
6	Cayla Bellamy	3 2 1	394	504	724
7	Alex Campbell	1 2 3	500	305	495
8	Lindsey Sawyer	1 3 2	376	310	392
9	TJ Remillard	2 1 3	283	313	275
10	Jillian Mendoza	2 3 1	285	285	252
11	Caleb Soo Hoo	3 1 2	431	276	469
12	Olivia Archer	3 2 1	432	268	473

**All response times are recorded in milliseconds*

Summary Statistics for Participants

Participant ID #	N	Mean	Std Dev	Minimum	Maximum
1	3	570.6666667	172.3204379	423	760
2	3	422.3333333	89.1309897	320	483
3	3	322.6666667	68.0906259	247	379
4	3	308.6666667	83.6799458	222	389
5	3	362.3333333	55.2479261	327	426
6	3	540.6666667	168.0277755	394	724
7	3	433.3333333	111.1680410	305	500
8	3	359.3333333	43.4664622	310	392
9	3	290.3333333	20.0333056	275	313
10	3	274.0000000	19.0525589	252	285
11	3	392.0000000	102.2399139	276	469
12	3	391.0000000	108.4758038	268	473

Appendix 2 - SAS Program

```
/* Proof of Data Collection and Descriptive Statistics */
/* The Effects of Visual and Auditory Cues on Human Reaction Time */
/* Dylan Demo, Jess Joblin, and Josh Lapish */

/* Reading in the data */
data reaction;
input participant $ rxn_type $ rxn_ms;
cards;
1 visual 423
1 sound 760
1 combined 529
2 visual 320
2 sound 483
2 combined 464
3 visual 247
3 sound 379
3 combined 342
4 visual 222
4 sound 389
4 combined 315
5 visual 327
5 sound 334
5 combined 426
6 visual 394
6 sound 504
6 combined 724
7 visual 500
7 sound 305
7 combined 495
8 visual 376
8 sound 310
8 combined 392
9 visual 283
9 sound 313
9 combined 275
10 visual 285
10 sound 285
10 combined 252
11 visual 431
11 sound 276
11 combined 469
12 visual 432
12 sound 268
12 combined 473
;

/*Grand Mean*/
proc means data=reaction;
run;

/*Factor Means*/
proc means data=reaction;
    class rxn_type;

run;

/*Block Means*/
proc means data=reaction;
    class participant;
run;

/*Box Plots*/
proc sgplot data=reaction;
    vbox rxn_ms / category = rxn_type;
run;

/*ANOVA, CI, Contrast, Assumptions*/
proc glm PLOTS=(DIAGNOSTICS RESIDUALS) data=reaction;
    class participant rxn_type;
    model rxn_ms = participant rxn_type / clparm;
    lsmeans participant rxn_type / adjust=bon cl pdiff;
    contrast 'Visual and Sound vs. Combined' rxn_type -1 .5 .5;
    estimate 'Visual and Sound vs. Combined' rxn_type -1 .5 .5;
run;
```

Appendix 3 - Protocol

1.) Take a convenience sample of 12 college students.

2.) Randomly assign each participant the order of the three treatment tests and ensure that three participants are assigned to each of the order categories. This will be done by putting three of each test level orders in a bag and having each participant select one without putting their selection back in the bag. Make sure each participant is informed of their test order.
   - There are six ways to order the levels, so each of the following will be in the bag two times:
     i.) Visual, Auditory, Combined
     ii.) Visual, Combined, Auditory
     iii.) Auditory, Visual, Combined
     iv.) Auditory, Combined, Visual
     v.) Combined, Visual, Auditory
     vi.) Combined, Auditory, Visual

3.) Have the website with the test pulled up on the laptop for the participant to take:
   - https://gizmos.explorelearning.com/index.cfm?method=cResource.dspView&ResourceID=43
     i.) This source will automatically time out after five minutes, so you will need to make an account in order to use this without any issues.

4.) Have the first participant enter an empty classroom with the experimenter and ensure all conditions are accounted for
   - Aspects to hold constant:
     i.) Turn all lights in the room on
     ii.) Ensure the same laptop is being used with the brightness and volume turned all the way up
     iii.) Ensure there are no other people in the room to eliminate outside noise and distractions
     iv.) On the website, use these settings:
       (1) Visual test: Sight = “red circle”
       (2) Auditory test: Sound = “bell”
       (3) Combined test: Sight = “red circle”, Sound = “bell”

5.) The experimenter will read the following instructions to make sure all participants are provided with the same information:
   - “There are three tests that will be conducted to test your reaction time: visual, auditory, and combined visual and auditory. For each reaction test, there will be 10 attempts that will be averaged. For the visual test click the touchpad in the blue area when the red circle appears. For the auditory test, click the touchpad in the blue area when a bell sound goes off. For the visual and auditory test, click the area when either a red circle appears or the bell sound goes off. This test will randomly switch ques between the red circle and bell; they will not be queued at the same time. Before you begin the first test, place your finger on the keypad to show you are ready.”

6.) Direct the participant to take the test.
   - They will take a total of three tests in the order they were previously given in the second step.

7.) Collect the results.
   - In an excel sheet, track the participant’s name along with the order in which they took the three tests
   - Record the average score, standard deviation, and ten reaction times for each of the three tests.