

Description of Inter-Item Timings (IRTs)

Tyler Peckenpaugh

3/29/2019

Contents

<i>Inter-item timing</i>	1
<i>Description and cleanup</i>	2
<i>Distribution of IRTs, all participants</i>	2
<i>Missing data and attrition</i>	3
<i>Group sizes after attrition</i>	3
<i>Distribution of experimental item IRT after attrition</i>	3
<i>Mean and SD of winsorized IRT by condition</i>	5
<i>Item and subject variation</i>	5
<i>Number of participants who show predicted pattern</i>	5
<i>Number of items that show predicted pattern</i>	6
<i>Analyses</i>	6
<i>Regression analyses</i>	6
<i>GEE Analyses</i>	8
<i>Excluding items</i>	9
<i>GEE Analyses</i>	9

Inter-item timing

Subjects were asked to read each sentence twice, once with no preview at all, and then again after unlimited preview. Inter-reading time (IRT) is a measure of the amount of time between when a subject stops speaking after a cold reading and when they begin speaking for a previewed reading.

IRT = delay after the end of a cold reading and before the start of a previewed reading

Practically, this was done over 1515 recordings (33 participants, 48 items = 1584 pairs, with 69 missing data). This was measured using Google's WebRTC Voice Activity Detection (VAD) over .wav files that had been subjected to a band-pass filter with a low threshold of 200Hz¹ and a high threshold of 3100Hz.

¹ a low hum in the room needed to be accounted for

Description and cleanup

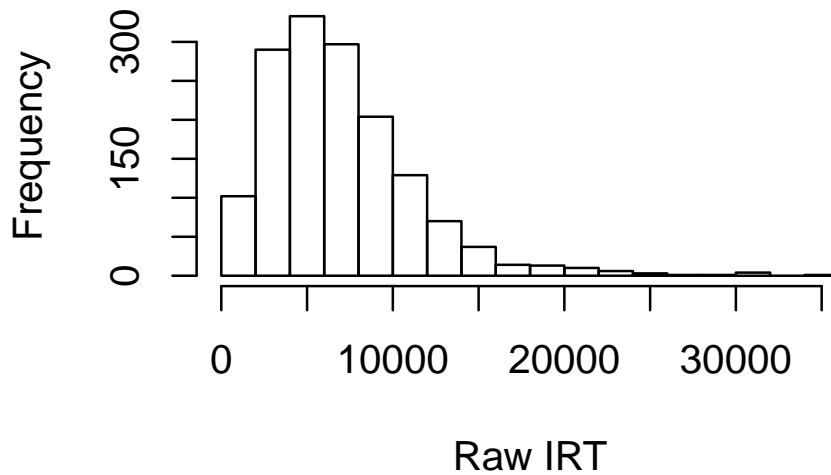
The following section details the IRT data and the outlier removal and resulting participant attrition.

Distribution of IRTs, all participants

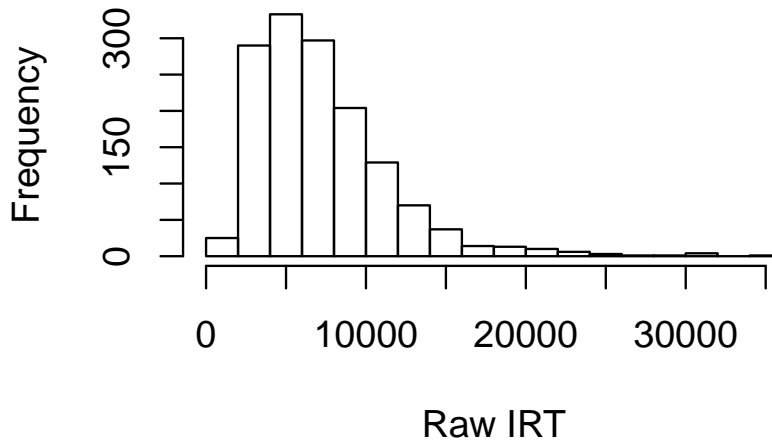
The overall mean IRT for all participants, all items (including fillers), and all conditions is 6982.88ms (sd = 4351.97). The highest IRT was 35792ms.

The following histograms show the distribution of IRT across all items and all participants. In the second graph, overly short IRTs (shorter than 1749ms; 75² such data) are excluded. In the third, overly long (longer than 14804.7; 76 such data) and overly short IRTs are excluded.

² This is 5% of the 1515 total data

Raw IRT, all Parts

Raw IRT, all Parts, short excluded



Missing data and attrition

Due to noise in recordings and/or technical difficulties during data collection, a number of IRTs are missing for experimental items in the data. The following table shows which participants are missing how many IRTs; ideally each would have 48 IRTs and 16 experimental IRTs.

The 8 participants missing more than 3 experimental IRTs (1, 4, 5, 6, 8, 11, 19, 203) are excluded.

Group sizes after attrition

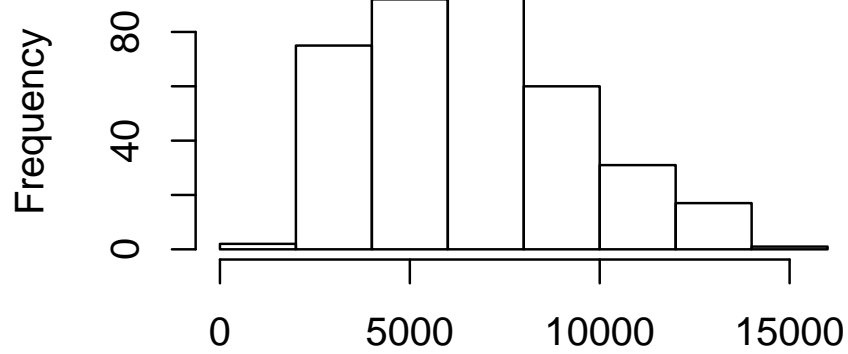
The following table³ shows how the participants are distributed across groups after attrition. Ideally, there would be 4 per group-order cell, but because of attrition the cells are uneven. Because regression is able to account for uneven groups, this defect will hopefully not play an important role in the analyses that follow.

³ There is 5 in Group 1, Split BA because I ran four participants per group-order, and then one extra who happened to be assigned to group 1, split BA; and by happenstance, none of the participants from that cell needed to be excluded.

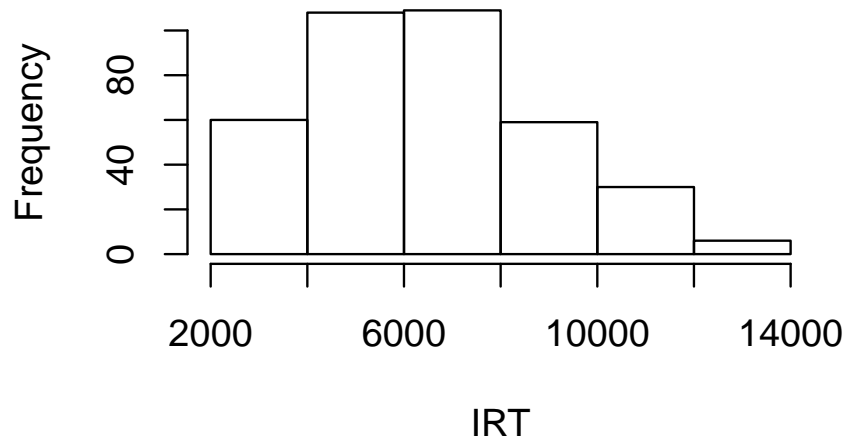
Distribution of experimental item IRT after attrition

The following histograms show the distribution of experimental item IRTs after attrition, and then the Winsorized IRTs, and finally the common log of winsorized IRTs, which are the shape of the data most suited to regression analyses.

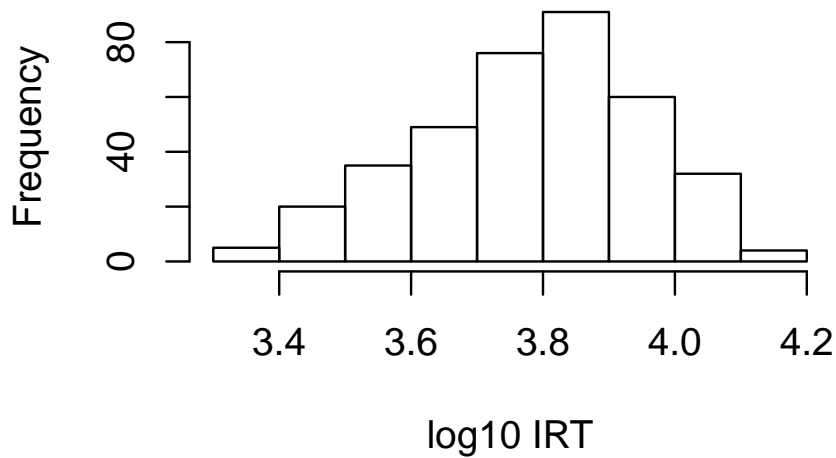
Raw IRT



Winsorized IRT



Common log of winsorized IRT



Mean and SD of winsorized IRT by condition

If we assume that interrogative PP-attachment garden paths are easier to process as an interrogative than in the declarative, and that IRT represents how difficult a sentence is to process, we would expect the difference in mean IRT to be larger for declarative garden paths compared to declarative controls than for the same comparison of interrogatives.

The means of the Winsorized IRT by condition indeed show this pattern.

Difference for declaratives: 388.88; for interrogatives: 222.96. While a difference across the interrogative/declarative condition of 165.92 is not huge, it does match the hypothesized pattern.

Item and subject variation

There is variation across participants in terms of whether or not they show this pattern.

Number of participants who show predicted pattern

In the analyzed data, 17 of 25 participants show the expected pattern. A one-tailed binomial allows against the null hypothesis that the true probability of the pattern holding across participants is less than 50% is significant ($p = 0.02$).

Number of items that show predicted pattern

For items, 11 of 16 show the pattern. A one-tailed binomial against the null hypothesis that the true probability of the pattern holding across items is less than 50% is statistically significant ($p = 0.04$).

Analyses

The following models explore the effect of garden path (\pm GP) and interrogativeness (\pm Q) on IRT.

Regression analyses

A full model with random slopes for GP, Q, and their interaction for both error terms fails to converge. A model with random slopes for just GP + Q likewise fails to converge. Simple models with no random slopes are shown below.

```

inter reading log 10
inter reading log 10
Predictors
Estimates
CI
p
Estimates
CI
p
(Intercept)
3.77
3.71 – 3.83
<0.001
3.77
3.71 – 3.83
<0.001
GP_conditionTRUE
0.02
-0.00 – 0.05
0.060
0.02
0.00 – 0.04
0.023
Q_conditionTRUE
0.02
-0.00 – 0.05
0.091

```

0.02
 0.00 – 0.04
 0.045
 GP_conditionTRUE:Q_conditionTRUE
 -0.01
 -0.04 – 0.03
 0.706
 Random Effects
 2
 0.01
 0.01
 00
 0.02 participant
 0.02 participant
 0.00 item
 0.00 item
 ICC
 0.71 participant
 0.71 participant
 0.00 item
 0.00 item
 Observations
 372
 372
 Marginal R2 / Conditional R2
 0.007 / 0.721
 0.007 / 0.721

The interaction model does not represent a statistically significant improvement in fit ($X^2 = 0.143$, $p < 0.71$), and so does not support the primary hypothesis.

The simplified model still does not show a statistically significant interaction.

inter reading log 10
 Predictors
 Estimates
 CI
 p
 (Intercept)
 3.77
 3.71 – 3.82
 <0.001
 GP_conditionTRUE
 0.02

-0.00 – 0.05
 0.058
 Q_conditionTRUE
 0.02
 -0.00 – 0.05
 0.090
 GP_conditionTRUE:Q_conditionTRUE
 -0.01
 -0.04 – 0.03
 0.701
 Random Effects
 2
 0.01
 00 participant
 0.02
 ICC participant
 0.71
 Observations
 372
 Marginal R2 / Conditional R2
 0.007 / 0.717

Removing the random effect of item does not degrade the model in a statically significant way ($X^2 = 0.75$, $p = 0.39$), but removing the random effect of participant does ($X^2 = 372.29$, $p = 0$).

GEE Analyses

Generalized Estimating Equation (GEE) analyses with participant as the error term show much the same results.

inter reading log 10
 inter reading log 10
 Predictors
 Estimates
 CI
 p
 Estimates
 CI
 p
 (Intercept)
 3.76
 3.70 – 3.82
 <0.001
 3.76

3.71 – 3.82
 <0.001
 GP_conditionTRUE
 0.03
 0.01 – 0.05
 0.012
 0.02
 0.00 – 0.04
 0.016
 Q_conditionTRUE
 0.03
 0.00 – 0.06
 0.027
 0.02
 0.00 – 0.04
 0.025
 GP_conditionTRUE:Q_conditionTRUE
 -0.02
 -0.06 – 0.02
 0.324
 Observations
 372
 372

The interaction model does not represent a statistically significant improvement in fit ($X = 0.973$, $p < 0.32$), and so does not support the primary hypothesis.

Excluding items

If we exclude items where the IRT for the control condition (-Q, -GP) is longer than 7s, we get the following:

Difference for declaratives: 427.58; for interrogatives: 287.87; difference of differences: 139.71.

GEE Analyses

Generalized Estimating Equation (GEE) analyses with participant as the error term show much the same results.

Dependent variable
 Dependent variable
 Predictors
 Estimates
 CI

p
 Estimates
 CI
 p
 (Intercept)
 3.74
 3.68 – 3.80
 <0.001
 3.75
 3.69 – 3.80
 <0.001
 GP_conditionTRUE
 0.04
 0.01 – 0.08
 0.017
 0.03
 0.01 – 0.05
 0.015
 Q_conditionTRUE
 0.04
 -0.01 – 0.08
 0.116
 0.02
 -0.01 – 0.05
 0.138
 GP_conditionTRUE:Q_conditionTRUE
 -0.03
 -0.10 – 0.03
 0.362
 Observations
 451
 451

 Dependent variable
 Dependent variable
 Predictors
 Estimates
 CI
 p
 Estimates
 CI
 p
 (Intercept)
 3.74

3.70 – 3.78
 <0.001
 3.75
 3.71 – 3.78
 <0.001
 GP_conditionTRUE
 0.04
 -0.01 – 0.10
 0.114
 0.03
 -0.01 – 0.07
 0.137
 Q_conditionTRUE
 0.04
 -0.02 – 0.09
 0.211
 0.02
 -0.02 – 0.06
 0.285
 GP_conditionTRUE:Q_conditionTRUE
 -0.03
 -0.11 – 0.05
 0.459
 Observations
 451
 451

The interaction model does not represent a statistically significant improvement in fit ($X = 0.83$, $p < 0.36$), and so does not support the primary hypothesis.

	Missing IRTs	Available percentage of IRTs	Missing experimental IRTs	Available percentage of experimental IRTs
1	1	97.92%	6	62.5%
2	3	93.75%	2	87.5%
3	3	93.75%	3	81.25%
4	6	87.5%	4	75%
5	25	47.92%	9	43.75%
6	1	97.92%	5	68.75%
7	0	100%	0	100%
8	10	79.17%	4	75%
9	0	100%	0	100%
10	0	100%	0	100%
11	13	72.92%	6	62.5%
12	2	95.83%	2	87.5%
13	1	97.92%	3	81.25%
14	0	100%	3	81.25%
15	0	100%	3	81.25%
16	0	100%	2	87.5%
17	0	100%	0	100%
19	1	97.92%	9	43.75%
20	0	100%	0	100%
21	1	97.92%	3	81.25%
22	0	100%	0	100%
201	0	100%	0	100%
203	0	100%	6	62.5%
204	0	100%	0	100%
205	0	100%	2	87.5%
206	0	100%	1	93.75%
207	0	100%	0	100%
208	0	100%	0	100%
209	1	97.92%	1	93.75%
210	0	100%	0	100%
212	0	100%	0	100%
214	1	97.92%	3	81.25%
215	0	100%	0	100%

	Split AB	Split BA	Group Total
Group 1	1	5	6
Group 2	3	4	7
Group 3	3	4	7
Group 4	2	3	5
Split Total	9	16	25

Table 2: Group/order totals after attrition

	Mean	SD
-Q -GP	6190.815	2324.193
-Q +GP	6579.692	2281.459
+Q -GP	6595.294	2389.589
+Q +GP	6818.253	2647.520

Table 3: Condition means

participant	-Q -GP	-Q +GP	+Q -GP	+Q +GP	pattern
2	6253.333	7516.550	4695.500	6598.867	TRUE
3	5753.200	7075.000	6724.667	6992.500	TRUE
7	5128.500	5479.250	6059.000	5901.500	TRUE
9	6095.750	6277.250	5911.250	6167.250	TRUE
10	9019.000	9265.000	8878.250	9488.750	TRUE
12	8041.400	8320.000	8826.867	8617.200	TRUE
13	8174.600	9811.600	9763.467	10350.850	TRUE
14	7172.867	8716.867	8424.650	10978.000	TRUE
15	7762.533	7808.250	8265.467	7402.933	TRUE
16	10678.667	10063.850	10220.300	13420.200	FALSE
17	4074.500	6155.750	6341.500	5058.750	TRUE
20	10209.500	9248.000	11517.250	10013.250	FALSE
21	5651.333	6414.500	5663.150	6120.200	TRUE
22	5572.500	5161.000	5262.750	6758.500	FALSE
201	4065.000	5036.500	4970.750	4403.250	TRUE
204	5460.750	6220.000	6605.250	6288.000	TRUE
205	9241.500	9281.667	7570.500	9986.333	TRUE
206	5121.250	4633.800	5291.750	4974.400	FALSE
207	6656.000	6541.750	6685.000	5464.250	FALSE
208	3331.750	3490.500	3007.750	3343.250	TRUE
209	3458.700	3747.600	3250.267	3701.150	TRUE
210	6600.500	6142.250	7252.250	6344.500	FALSE
212	5592.250	5152.500	5305.500	7267.500	FALSE
214	3339.600	3987.533	3517.667	3299.667	TRUE
215	4452.500	3746.000	5065.750	4746.250	FALSE

Table 4: Mean IRT by condition and participant

item	-Q -GP	-Q +GP	+Q -GP	+Q +GP	pattern
1	6218.333	5892.400	7567.933	6861.857	FALSE
2	7472.667	6696.800	4865.133	5837.200	FALSE
3	7420.833	6333.571	6480.000	7184.200	FALSE
4	5827.880	7715.167	6640.960	5169.800	TRUE
5	6614.933	6522.371	6645.257	5629.200	FALSE
6	6134.771	6826.533	6575.560	7036.343	TRUE
7	5905.650	7216.467	6726.171	8695.200	TRUE
8	5571.320	6956.400	6091.500	6415.720	TRUE
9	5844.714	6607.600	7332.033	7508.440	TRUE
10	6705.600	6813.033	6611.743	4408.667	TRUE
11	4905.680	5177.000	5735.429	8334.440	TRUE
12	5120.914	5876.600	7570.167	6927.714	TRUE
13	7048.160	6479.667	6570.486	7973.033	FALSE
14	5657.143	6564.867	6403.560	6621.343	TRUE
15	5858.367	6143.600	7656.800	6470.833	TRUE
16	6743.886	7223.333	6036.333	6648.886	TRUE

Table 5: Mean IRT by condition and item

	Mean	SD
-Q -GP	6292.379	3202.923
-Q +GP	6719.957	2770.712
+Q -GP	6690.195	3046.030
+Q +GP	6978.065	3278.521

Table 6: Condition means