

Preliminary Data Screening

Kaleem Ullah

2024-03-15

Check the percentage of videos that were played

```
video_data <- read.csv("video_data_flattened.csv", header=TRUE)
num_ppn <- length(unique(video_data$ppn))
df_recog <- read.csv("emotion_recog.csv", header=TRUE)
df_iden <- read.csv("emotion_iden.csv", header=TRUE)
expected_played <- num_ppn*24
videos_played <- length(video_data$ppn)
```

Expected videos played: 4176 Total videos played: 3802 Percentage of videos played: 91.0440613

Check individually for each task

1. Recognition Task

Videos Played

```
zoneType <- df_recog$Zone.Type
last_iter = "response_button_text"
unplayed = 0
for (i in zoneType){
  if (last_iter == "response_button_text" & i == "response_button_text") {
    unplayed = unplayed + 1
  }
  last_iter = i
}
unplayed
```

```
## [1] 222
```

Reaction Time

```
reac_recog <- sum(video_data$reaction.recog < 1000, na.rm=TRUE)
```

Percentage of videos played less than 1 second: 52.7880063

2. Identification Task

```
reac_iden <- sum(video_data$reaction.iden < 1000, na.rm=TRUE)
```

Percentage of videos played less than 1 second: 53.3403472

Time Taken for each task

```
total_time_recog <- as.numeric(df_recog[df_recog$Trial.Number == "END TASK",]$Reaction.Time) / 1000
total_time_iden <- as.numeric(df_iden[df_iden$Trial.Number == "END TASK",]$Reaction.Time) / 1000
```

1. Recognition Task

```
summary(total_time_recog)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  25.42  170.66  219.18  264.14  296.95 1815.12
```

```
sort(total_time_recog)
```

```
##      [1]  25.422  31.380  45.046  91.040  96.196  98.448  98.596  99.272
##      [9] 104.617 107.342 111.155 111.775 115.758 119.523 121.568 122.775
##     [17] 128.081 128.367 129.331 129.646 135.577 137.616 138.455 138.739
##     [25] 142.918 143.230 150.596 150.994 151.013 151.856 152.826 153.897
##     [33] 155.782 158.933 161.435 163.218 163.511 164.371 165.455 166.600
##     [41] 168.411 168.509 168.671 169.105 170.249 171.065 171.897 173.488
##     [49] 175.949 178.825 180.430 184.543 184.918 185.819 186.579 187.301
##     [57] 188.599 189.077 190.421 190.748 191.345 191.644 192.511 194.451
##     [65] 194.517 197.281 198.404 199.389 200.454 200.693 201.052 201.348
##     [73] 201.997 202.299 202.639 203.629 204.323 204.339 206.078 207.159
##     [81] 207.639 208.773 209.117 212.359 212.609 213.867 217.691 217.871
##     [89] 217.996 219.175 219.479 219.566 220.208 220.897 222.238 222.494
##     [97] 226.693 227.377 227.993 230.694 232.241 234.132 236.941 238.428
##    [105] 238.732 243.662 244.315 245.208 245.882 246.494 246.524 247.029
##    [113] 247.033 247.822 248.763 250.711 251.103 251.666 252.270 253.500
##    [121] 257.019 258.191 262.869 263.864 265.258 265.827 268.261 269.695
##    [129] 271.605 280.823 281.466 284.930 292.694 296.209 297.698 301.737
##    [137] 304.419 305.480 306.155 308.144 316.431 316.729 319.790 322.622
##    [145] 326.810 334.347 334.647 334.760 337.818 338.576 339.439 341.709
##    [153] 347.387 358.188 361.520 361.950 375.457 389.208 389.507 392.982
##    [161] 396.023 397.512 400.795 412.057 414.828 415.081 430.047 469.989
##    [169] 472.437 473.420 484.781 496.328 526.559 537.140 607.232 760.812
##   [177] 1584.184 1684.825 1815.116
```

2. Identification Task

```
summary(total_time_iden)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   15.38  141.25  187.58  220.41  238.95 1102.60
```

```
sort(total_time_iden)
```

```
## [1] 15.375 32.619 33.482 49.708 51.679 62.428 77.437 78.599
## [9] 79.975 83.865 87.290 88.349 88.696 90.020 90.913 92.681
## [17] 93.115 94.540 100.328 101.144 101.585 102.332 105.150 107.472
## [25] 107.538 108.882 112.977 118.110 118.580 118.722 119.439 119.469
## [33] 119.485 120.541 123.124 126.182 126.550 127.969 128.030 128.552
## [41] 131.204 134.952 136.407 138.213 139.260 143.242 143.643 144.493
## [49] 145.434 145.664 146.034 149.096 150.942 151.285 151.500 153.553
## [57] 155.474 156.478 157.020 161.188 161.749 163.131 163.386 164.003
## [65] 164.024 166.810 167.132 168.770 169.548 170.579 172.627 174.414
## [73] 174.442 174.596 175.458 175.808 176.246 176.940 177.509 177.972
## [81] 179.644 180.493 180.495 180.820 182.026 182.118 184.496 184.825
## [89] 186.491 187.579 188.820 188.830 190.553 190.751 191.234 192.893
## [97] 193.784 194.745 198.627 200.524 202.230 202.506 205.985 207.269
## [105] 207.977 208.197 209.474 209.697 210.527 210.916 213.676 215.794
## [113] 217.136 218.429 219.431 219.506 219.987 221.210 221.734 223.637
## [121] 225.605 225.657 225.987 226.131 228.561 230.500 234.159 234.215
## [129] 234.458 236.278 236.313 237.938 238.245 238.417 239.480 239.485
## [137] 239.685 239.925 240.694 245.426 248.366 251.326 253.223 262.647
## [145] 262.782 267.613 268.929 271.070 281.173 282.005 282.442 285.649
## [153] 289.841 291.008 315.167 323.961 328.524 331.971 335.538 351.566
## [161] 355.338 378.736 381.506 386.006 390.017 395.798 422.499 422.718
## [169] 453.505 459.580 460.937 470.820 520.955 606.520 615.076 908.653
## [177] 1009.096 1022.720 1102.595
```

Preliminary Analyses

```
# This R file contains code to read and analyze preprocessed data from Gorilla
require(lme4)
```

```
## Loading required package: lme4
```

```
## Loading required package: Matrix
```

```
# Read files
setwd("D:/RA - Spring 2024/Gorilla Data")
df <- na.omit(read.csv("preprocessed_adoptee.csv"))

df$culture <- as.factor(df$culture)
# df$ppn <- as.factor(df$ppn)

# Hypothesis 1
accuracy_fit_1 <- glmer(accuracy ~ culture + (1 | ppn1),
  data = df,
  family = binomial,
  control = glmerControl(optimizer = "bobyqa"),
  nAGQ = 10)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
accuracy_fit_2 <- glmer(accuracy ~ culture + (1 | ppn1/emotion),
  data = df,
  family = binomial,
  control = glmerControl(optimizer = "bobyqa"))
```

```
## boundary (singular) fit: see help('isSingular')
```

```
summary(accuracy_fit_1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 10) [glmerMod]
## Family: binomial ( logit )
## Formula: accuracy ~ culture + (1 | ppn1)
## Data: df
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##  3905.2   3930.4 -1948.6   3897.2     4025
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -0.5206 -0.5206 -0.4676 -0.4595  2.1763
##
## Random effects:
##   Groups Name            Variance Std.Dev.
##   ppn1    (Intercept)  0          0
## Number of obs: 4029, groups: ppn1, 173
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.52047    0.07096 -21.427  <2e-16 ***
## cultureCN    0.21481    0.09752   2.203   0.0276 *
## cultureNL   -0.03476    0.10094  -0.344   0.7306
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) cltrCN
## cultureCN -0.728
## cultureNL -0.703  0.512
## optimizer (bobyqa) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
summary(accuracy_fit_2)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: accuracy ~ culture + (1 | ppn1/emotion)
## Data: df
## Control: glmerControl(optimizer = "bobyqa")
```

```
##
##      AIC      BIC   logLik deviance df.resid
## 3688.3 3719.8 -1839.1 3678.3 4024
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -0.9636 -0.2904 -0.2510 -0.2461  2.0044
##
## Random effects:
## Groups      Name      Variance Std.Dev.
## emotion:ppn1 (Intercept) 2.657    1.63
## ppn1         (Intercept) 0.000    0.00
## Number of obs: 4029, groups: emotion:ppn1, 1377; ppn1, 173
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.24920    0.12151 -18.510 <2e-16 ***
## cultureCN    0.29170    0.11414  2.556 0.0106 *
## cultureNL   -0.03903    0.11707  -0.333 0.7388
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) cltrCN
## cultureCN -0.531
## cultureNL -0.474 0.507
## optimizer (bobyqa) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
anova(accuracy_fit_1, accuracy_fit_2)
```

```
## Data: df
## Models:
## accuracy_fit_1: accuracy ~ culture + (1 | ppn1)
## accuracy_fit_2: accuracy ~ culture + (1 | ppn1/emotion)
##              npar      AIC      BIC logLik deviance Chisq Df Pr(>Chisq)
## accuracy_fit_1    4 3905.2 3930.4 -1948.6 3897.2
## accuracy_fit_2    5 3688.3 3719.8 -1839.1 3678.3 218.91 1 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Hypothesis 2
iden_fit <- lmer(likert1 ~ culture + (1 | ppn1),
                data=df)
summary(iden_fit)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: likert1 ~ culture + (1 | ppn1)
## Data: df
##
## REML criterion at convergence: 18135.4
##
## Scaled residuals:
```

```
##      Min      1Q  Median      3Q      Max
## -3.4312 -0.6364 -0.0764  0.6167  3.2295
##
## Random effects:
##   Groups   Name      Variance Std.Dev.
##   ppn1      (Intercept) 2.608    1.615
##   Residual                4.708    2.170
## Number of obs: 4029, groups: ppn1, 173
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  4.09109    0.13637  29.999
## cultureCN    0.03728    0.08387   0.445
## cultureNL    0.22590    0.08360   2.702
##
## Correlation of Fixed Effects:
##              (Intr) cltrCN
## cultureCN -0.305
## cultureNL -0.307  0.498
```

Hypothesis 3

```
glm_fit <- glmer(accuracy ~ culture + likert1 + (1 | ppn1),
  data = df,
  family = binomial,
  control = glmerControl(optimizer = "bobyqa"),
  nAGQ = 10)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
glm_fit_2 <- glmer(accuracy ~ culture + likert1 + (1 | ppn1),
  data = df,
  family = binomial)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
summary(glm_fit)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 10) [glmerMod]
##   Family: binomial ( logit )
## Formula: accuracy ~ culture + likert1 + (1 | ppn1)
##   Data: df
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC    logLik deviance df.resid
##  3896.6   3928.1  -1943.3   3886.6     4024
##
## Scaled residuals:
##      Min      1Q  Median      3Q      Max
## -0.5991 -0.5004 -0.4659 -0.4220  2.4245
##
## Random effects:
```

```
## Groups Name      Variance Std.Dev.
## ppn1 (Intercept) 5.666e-16 2.38e-08
## Number of obs: 4029, groups: ppn1, 173
##
## Fixed effects:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.72567    0.09620 -17.938 < 2e-16 ***
## cultureCN    0.21395    0.09765   2.191 0.02846 *
## cultureNL   -0.04561    0.10113  -0.451 0.65197
## likert1      0.04872    0.01499   3.249 0.00116 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr) cltrCN cltrNL
## cultureCN -0.537
## cultureNL -0.496 0.511
## likert1   -0.674 0.000 -0.034
## optimizer (bobyqa) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
summary(glm_fit_2)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: accuracy ~ culture + likert1 + (1 | ppn1)
## Data: df
##
##           AIC      BIC   logLik deviance df.resid
##    3896.6    3928.1  -1943.3   3886.6      4024
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -0.5991 -0.5004 -0.4659 -0.4220  2.4245
##
## Random effects:
## Groups Name      Variance Std.Dev.
## ppn1 (Intercept) 0          0
## Number of obs: 4029, groups: ppn1, 173
##
## Fixed effects:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.72567    0.09620 -17.938 < 2e-16 ***
## cultureCN    0.21395    0.09765   2.191 0.02846 *
## cultureNL   -0.04561    0.10113  -0.451 0.65197
## likert1      0.04872    0.01499   3.249 0.00116 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr) cltrCN cltrNL
## cultureCN -0.537
## cultureNL -0.496 0.511
```

```
## likert1    -0.674  0.000 -0.034  
## optimizer (Nelder_Mead) convergence code: 0 (OK)  
## boundary (singular) fit: see help('isSingular')
```