case_study_RT_accuracy

Jiaxuan Li

2023-06-10

Experiment design

Describe the design of the experiment in your own words. The study investigate whether adaptation is exposure-specific. Participants are randomly assigned into two groups: Control and Accent. Both groups will be first hearing native accented speech from Talker 1, and go through three blocks of exposure, and will finally be tested on Mandarin accented speech. In the control group, subject will exposed to Native accent speech from a talker different from the practice stage. In the Accent group, subject will be exposed to Mandarin-accented speech from the same talker as they will be tested.

We predict that performance will improve during exposure phase for conditions; and the performance will be better during the test phase in the Accent condition than Control condition.

Data cleaning

1. Examine RT distribution

Examine the distribution of RT (subjMean_BlockRT) across subjects. Does it make sense?

The RT distribution roughly follows normal distribution but it heavily right-skewed – probably suggesting that some participants did not pay attention or did not have a good equipment and should be excluded.

2. Data exclusion

Describe the procedure you take to exclude outliers (subjects, trials, etc.).

Exclusion by subject

Describe your exclusion criteria based on a subject's performance.

e.g., We want to identify and remove subjects who consistently registered slow response times because they did not perform the task faithfully (e.g., multi-tasking) or because their computer equipment did not provide reliable recording of RTs over the web.

Based on the plotting, there are a lot of trials with extremely long RTs. We therefore exclude subjects with mean block RT greater than 9000ms.

```
## Number of excluded trials: 12
```

```
## Excluded trials by Condition:
```

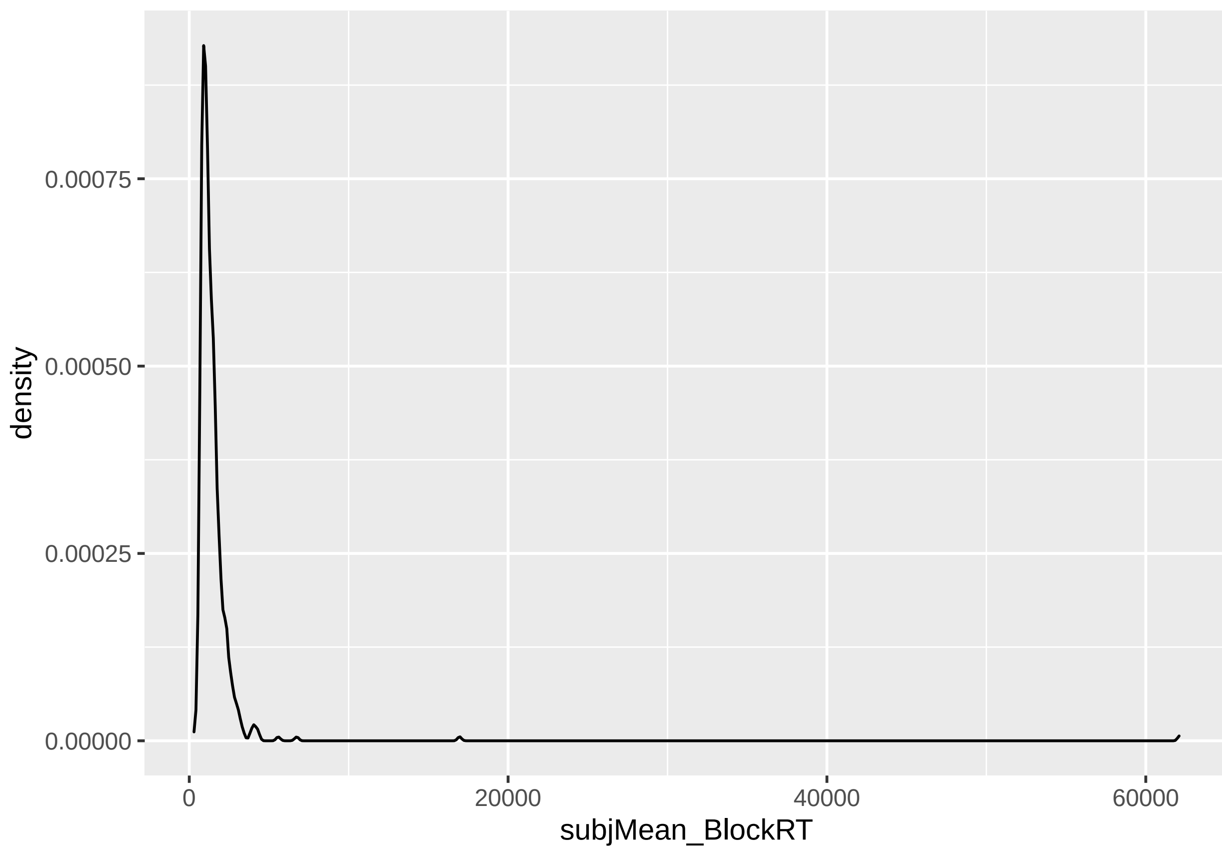


Figure 1: Distribution of subjects mean RTs by Block and Condition, prior to outlier exclusions.

```
##                            Condition Excluded_Trials
## 1          Control (native accent)            1764
## 2 Mandarin-accented English (same)            1836
```

Re-examine RT distribution after subject exclusion.

After data exclusion, the distribution is normally distributed.

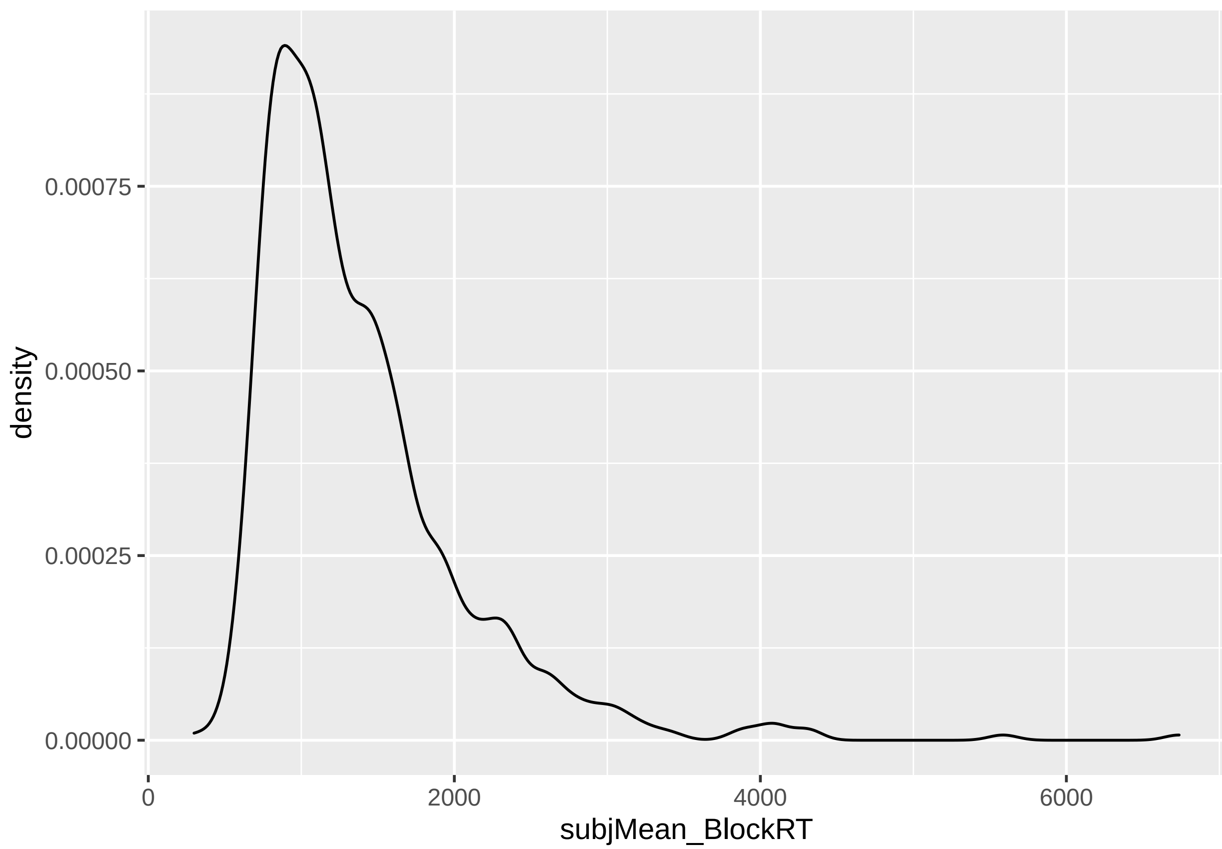
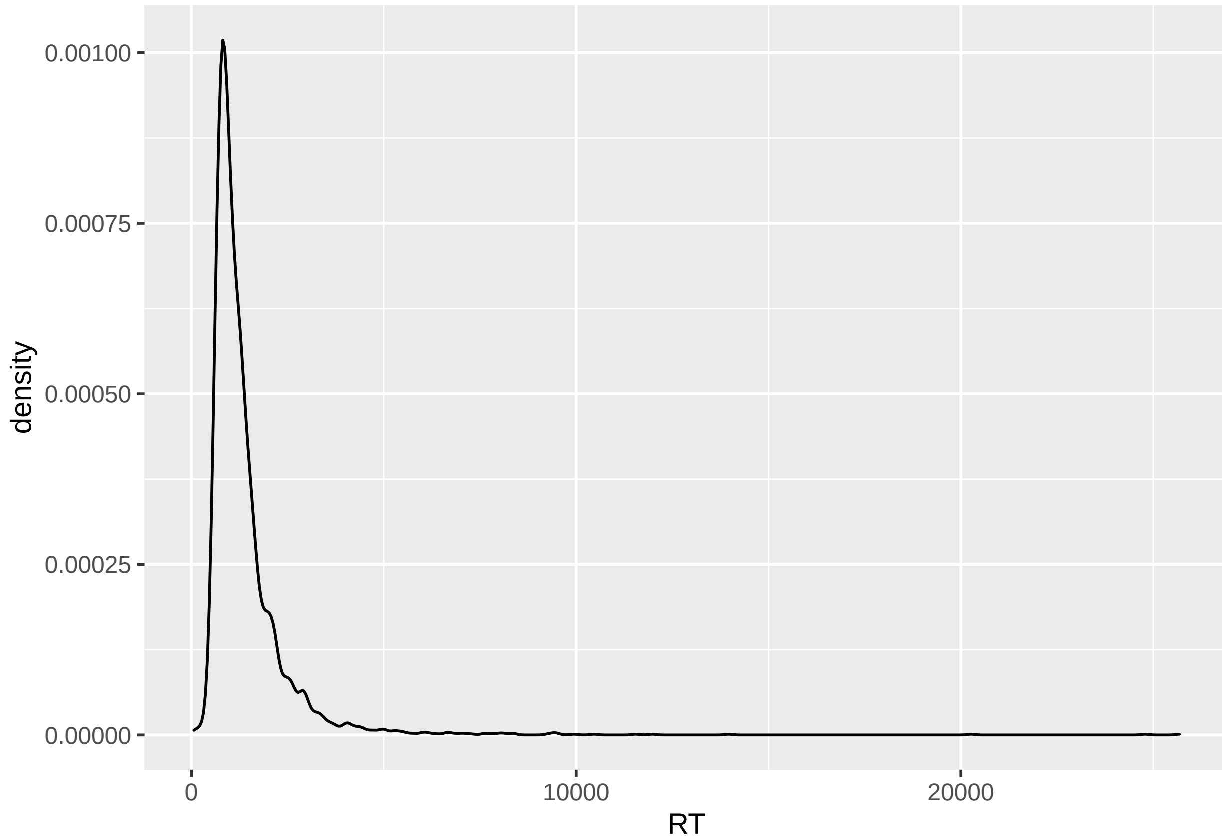


Figure 2: ...

Exclusion by trial with extreme RTs

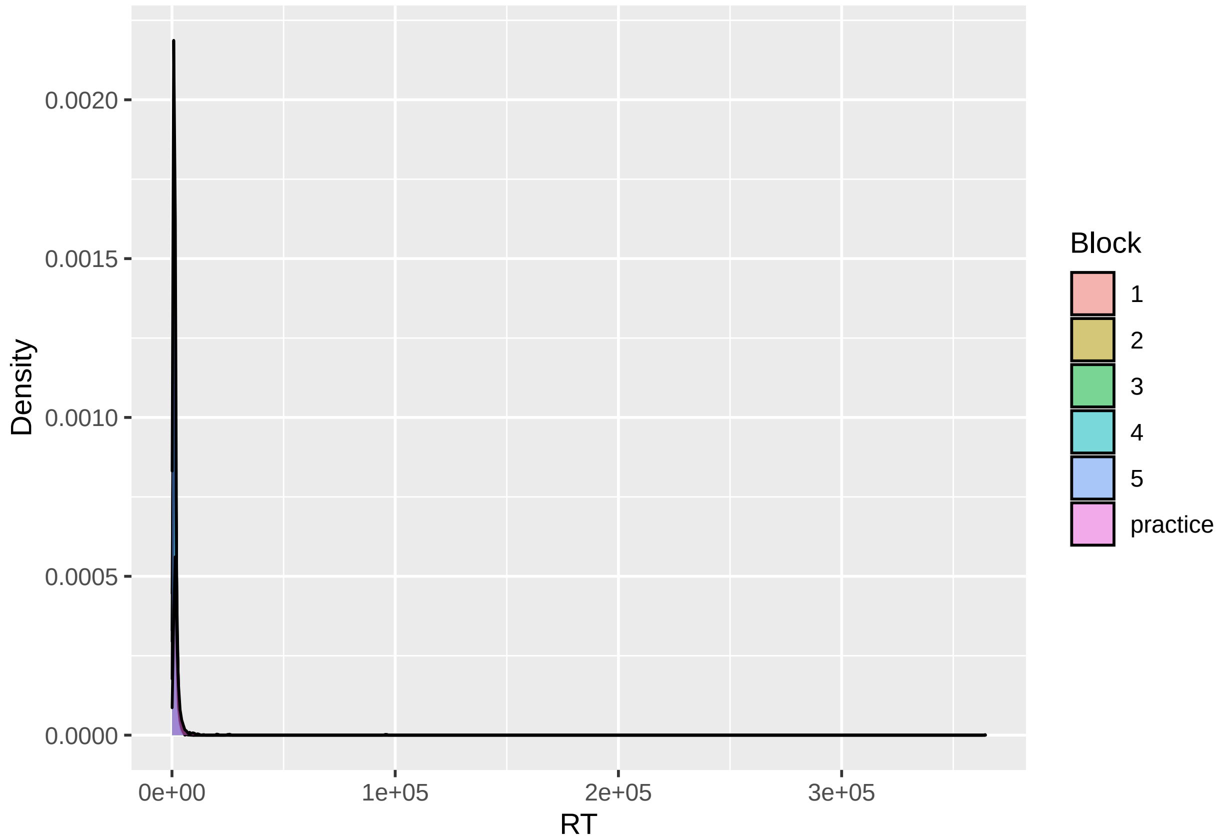
The second step of outlier removal was to exclude trials with atypical RTs. Describe your exclusion criteria by trial and do a second round of exclusion.

Q: Did trial-wise outlier exclusion disproportionately affect any experimental Conditions? A: The trial-wise outlier exclusion did not disproportionally affect experimental conditions.

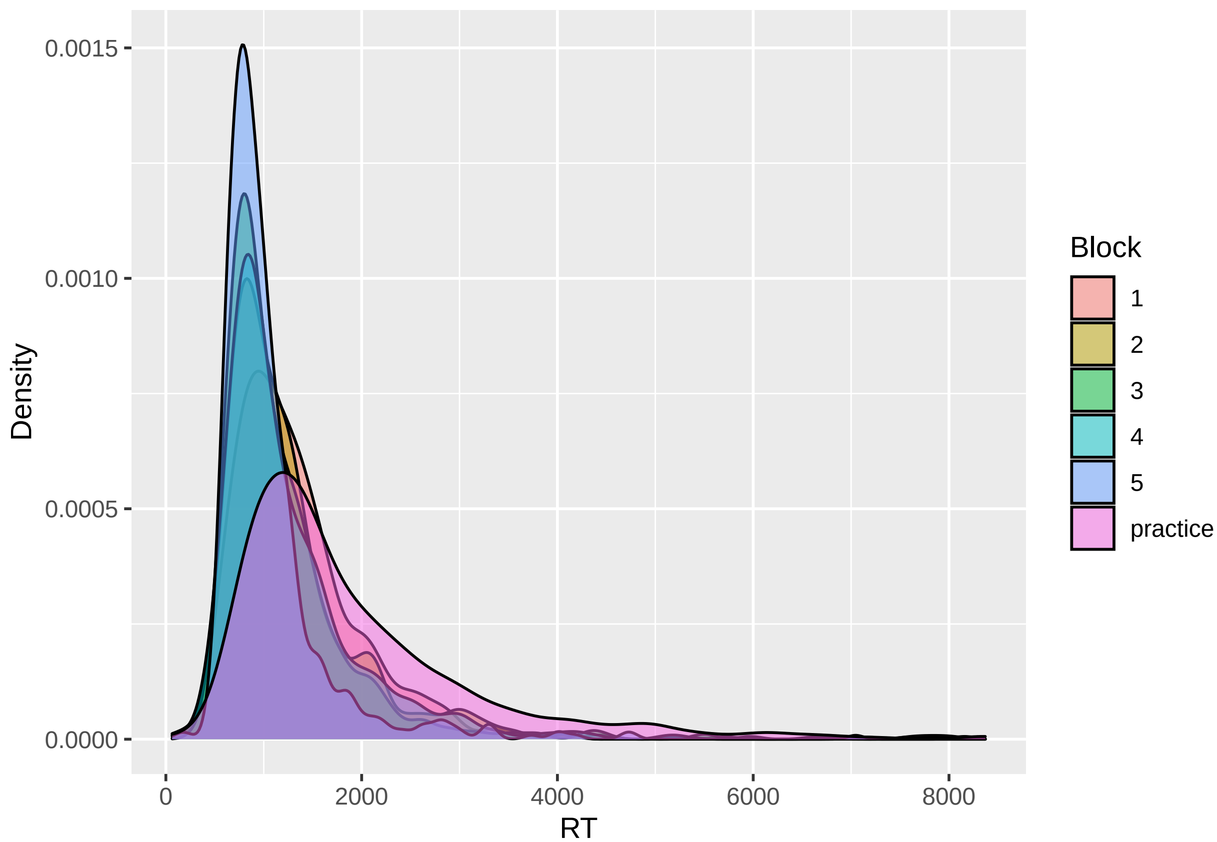



```
## Number of excluded trials: 12

## Excluded trials by Condition:

##                            Condition Excluded_Trials
## 1           Control (native accent)            1764
## 2 Mandarin-accented English (same)            1824
```

Q: Examine the mean RTs by block. Do they vary a lot before and after trial exclusion? Describe the effects.
A: Before exclusion, this distribution is heavily right skewed, but after exclusion the RT distribution follows a normal distribution.



Density
0.0020
0.0015
0.0010
0.0005
0.0000
0e+00
1e+05
2e+05
3e+05
RT
Block
1
2
3
4
5
practice

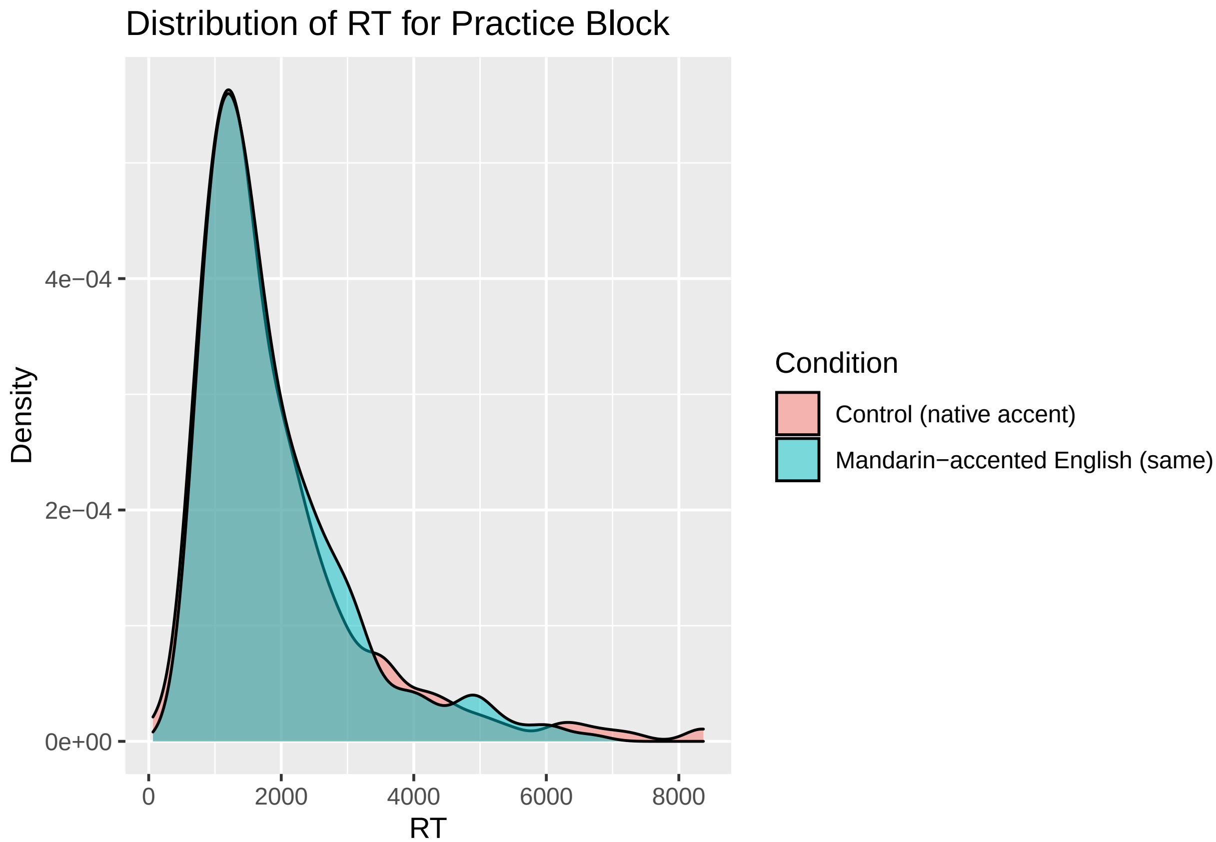


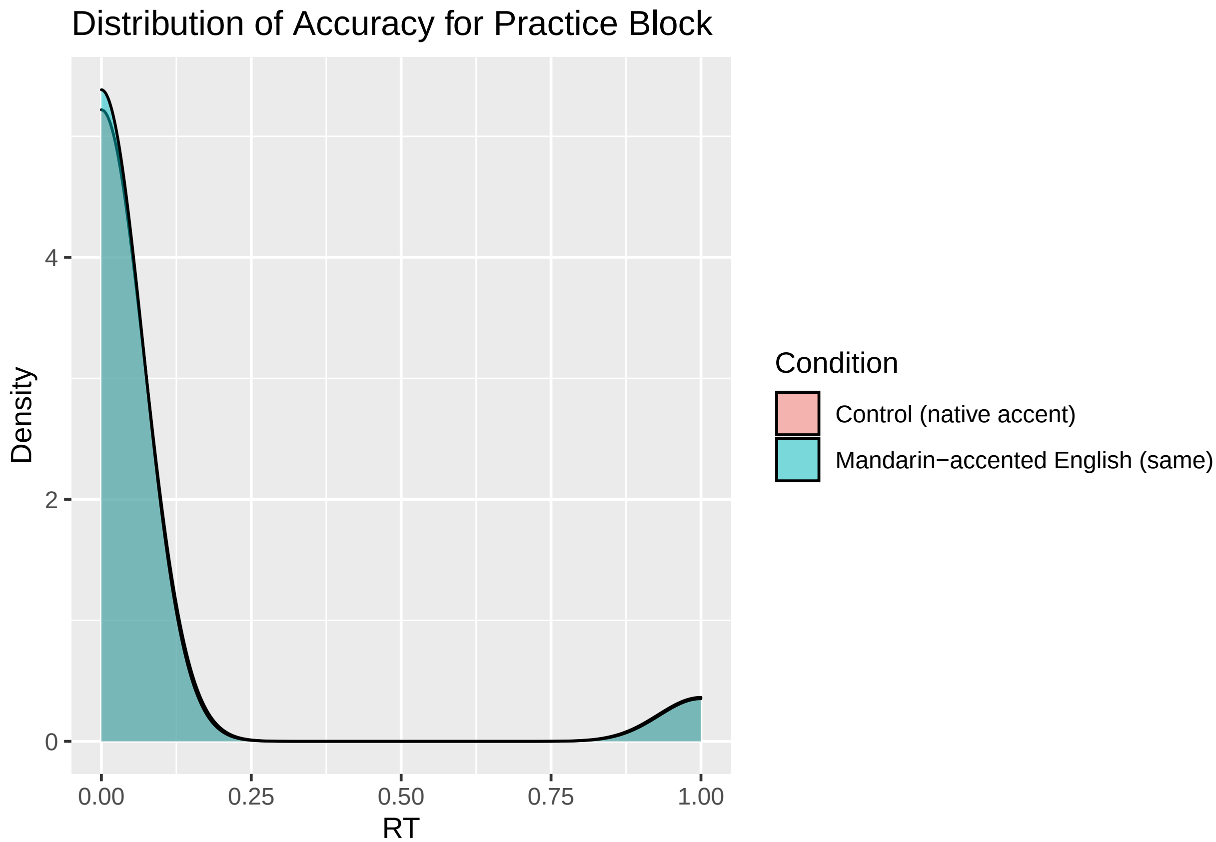
Examine RTs and Accuracy during practice and baseline (after exclusion steps 1 and 2)

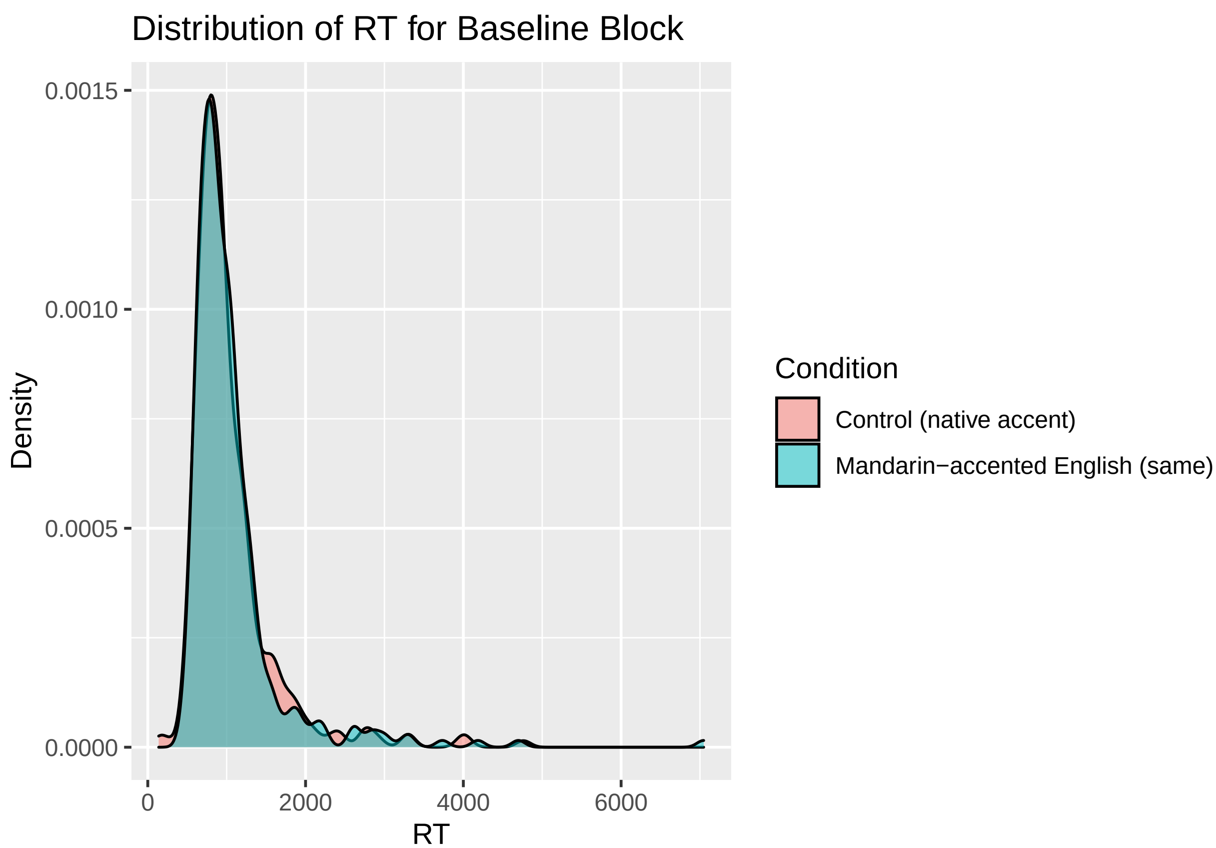
Now that we've excluded extreme subject and trial outliers, we can look at the practice and baseline data to assess our high-level predictions about how participants should perform on this web-based task.

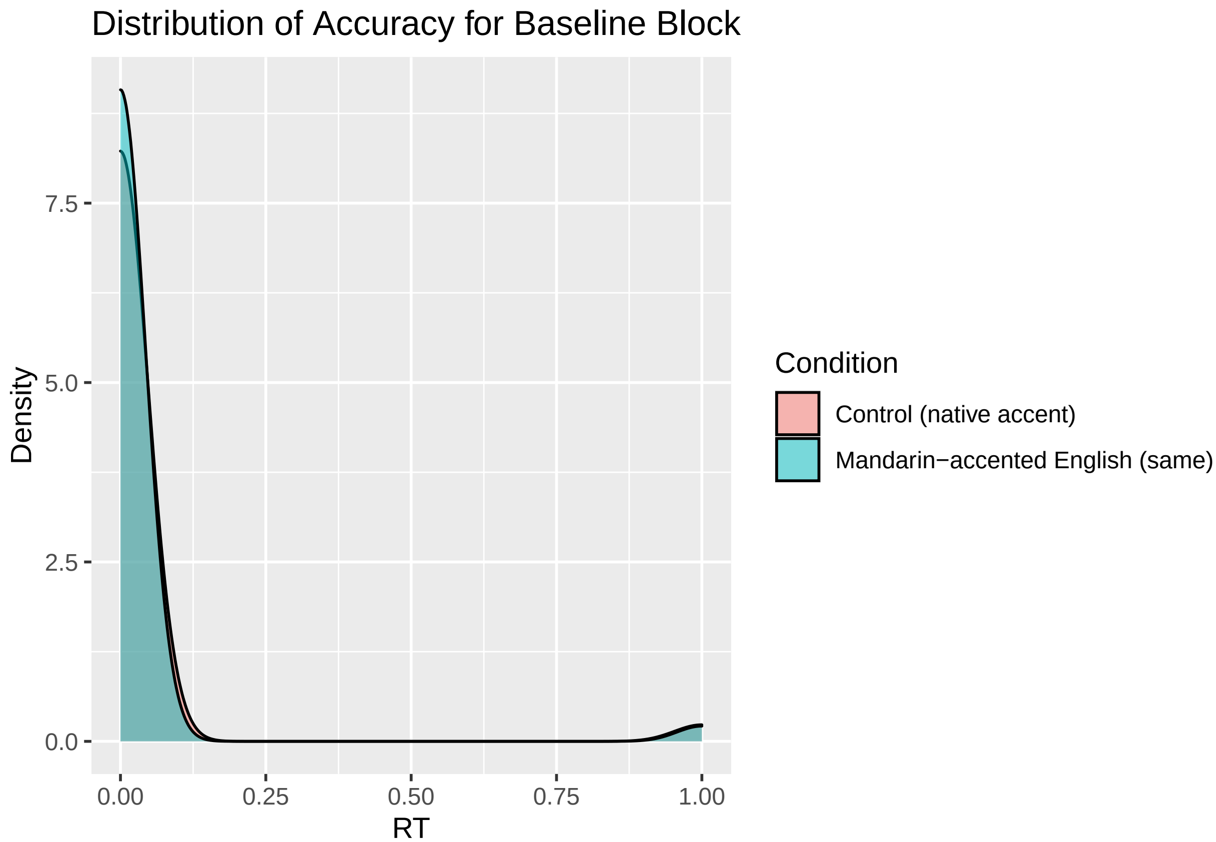
1. **One data pattern that we expect to find is that performance (both RTs and accuracy) in the practice and baseline blocks is comparable across experimental conditions.** We expect this because these blocks of the experiment were identical across conditions (i.e., native-accented stimuli presented in the clear).

   - … *if performance in the* ***practice block*** *differs substantially across conditions*, we would need to consider whether the subjects in each condition were sampled from the same underlying population (e.g., did we run all conditions at approximately the sme time of day?).
   - … *if performance in the* ***baseline block*** *differs substantially across conditions*, we would need to consider whether exposure to different types of speech during the main block of the experiment induced overall differences in task performance (in which case the baseline block doesn't provide a reliable condition-independent "baseline" for normalization purposes).

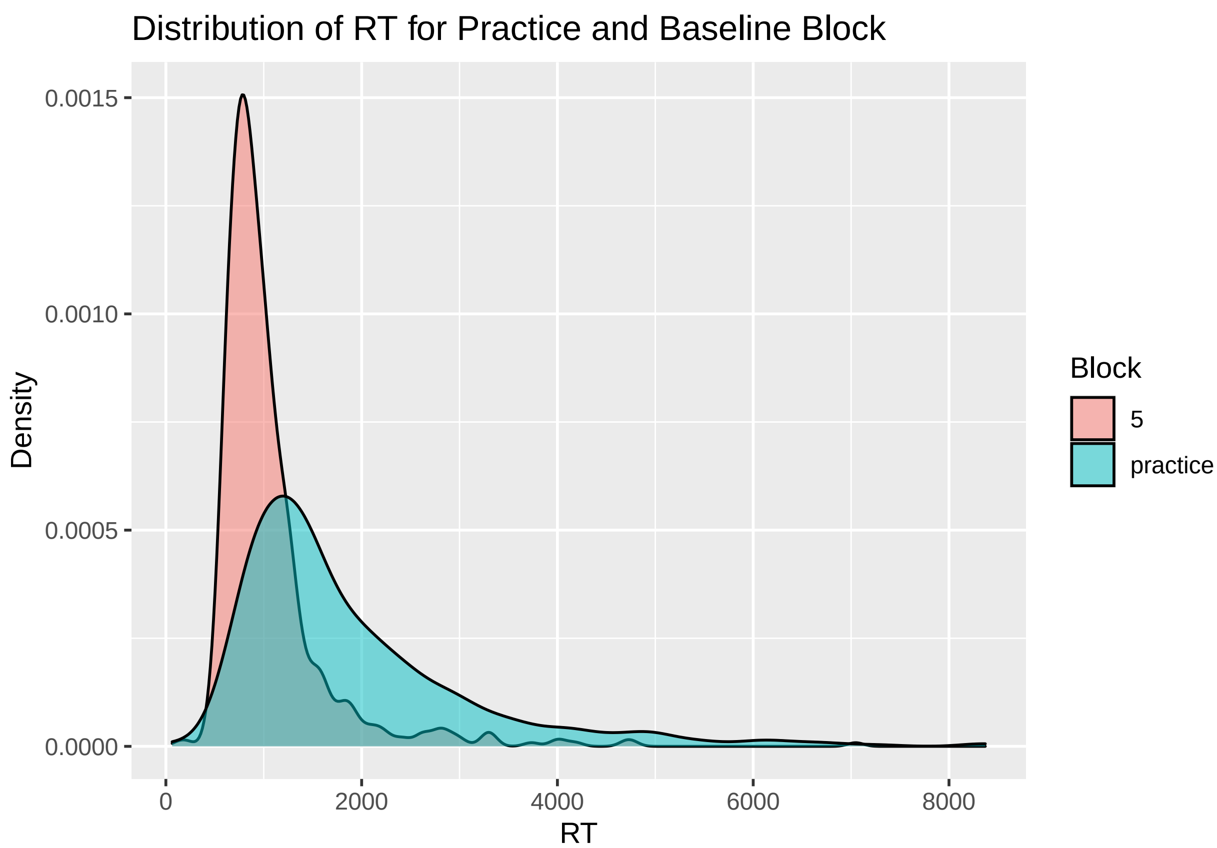

Distribution of RT for Practice Block
Density
4e−04
2e−04
0e+00
0
2000
4000
6000
8000
RT
Condition
Control (native accent)
Mandarin−accented English (same)

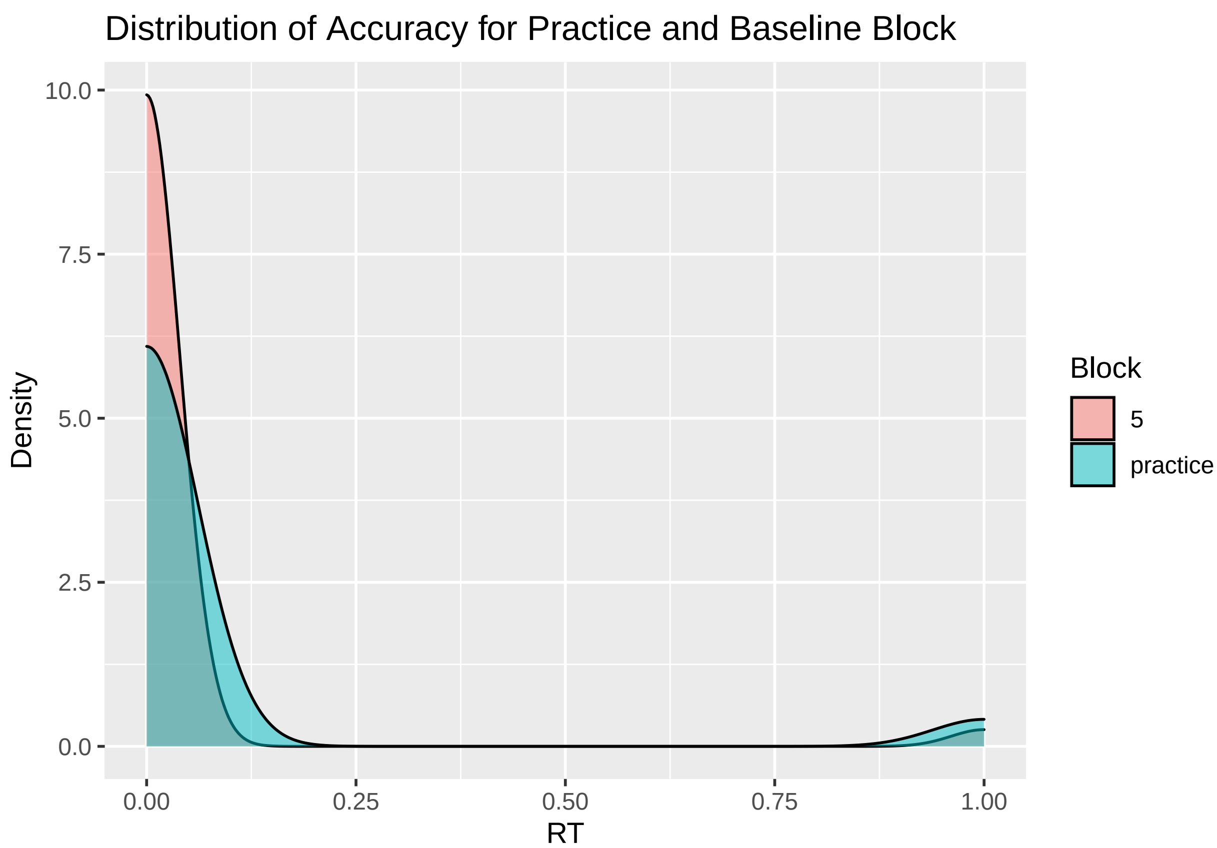

Distribution of Accuracy for Practice Block
Density
4
2
0
0.00
0.25
0.50
0.75
1.00
RT
Condition
Control (native accent)
Mandarin−accented English (same)


Distribution of RT for Baseline Block
0.0015
0.0010
0.0005
0.0000
Density
0
2000
4000
6000
RT
Condition
Control (native accent)
Mandarin−accented English (same)



2. **A second data pattern that we expect to find is evidence of improvement (adaptation) over the course of the task.** One way this would manifest is faster RTs and increased accuracy in the post-experiment baseline block, relative to the practice phase.


Distribution of RT for Practice and Baseline Block
0.0015
0.0010
0.0005
0.0000
Density
0
2000
4000
6000
8000
RT
Block
5
practice



Summary of exclusion criteria:

- Participant-level exclusions:
  - exclude subjects with mean response time over 9000ms
- Trial-level exclusions:
  - exclude trials with response time over 9000ms

We applied the same exclusion criteria across all RT and error analyses.

Normalize experimental RTs relative to baseline

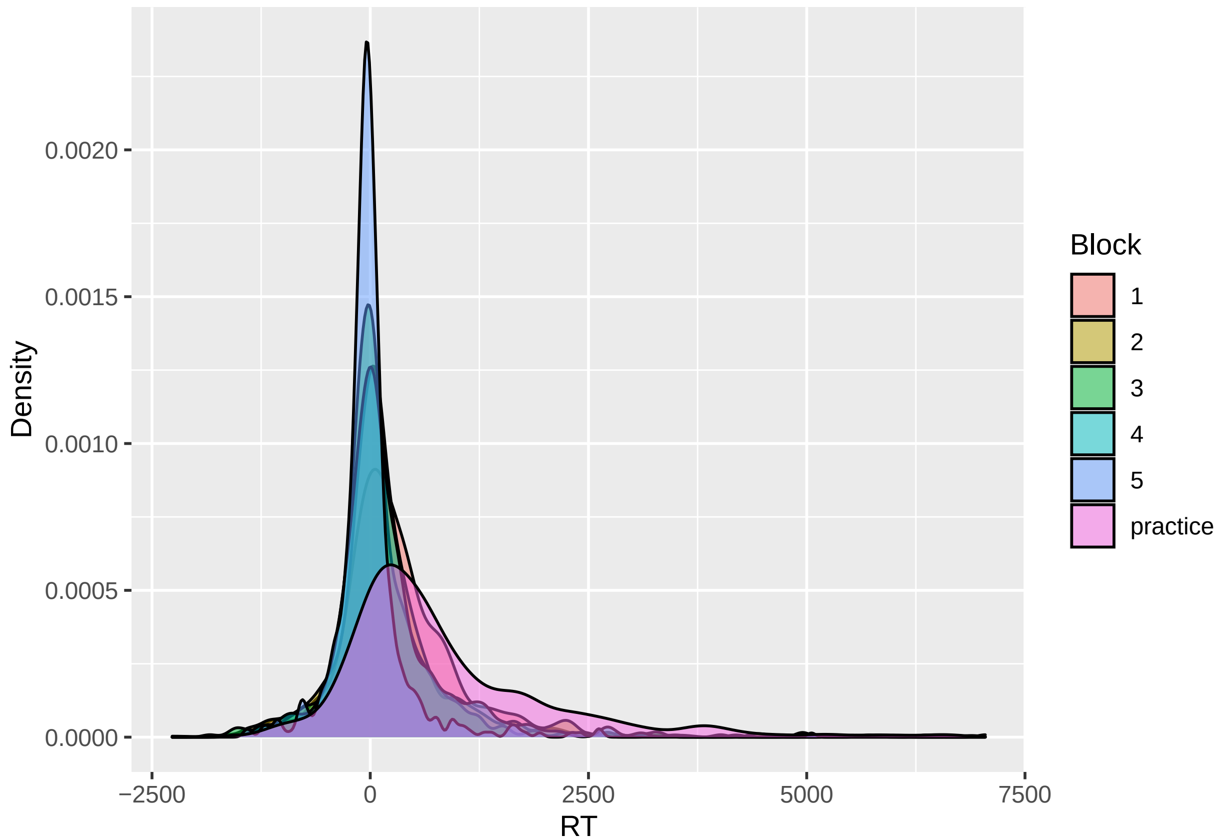
Now that we've completed all trial-wise RT exclusions, we can calculate *normalized* RTs that take into account each subject's baseline speed on this task. For this procedure, we adjust the RTs on each trial by subtracting out the corresponding subject's mean RT during the baseline phase. We refer to the resulting measure as *adjusted RTs*.

```
# calculate each subject's mean Baseline RT
# and subtract that value from experimental RTs
dat_out2 %<>%
  group_by(WorkerId) %>%
  mutate(
    # calculate subject-wise mean RTs during baseline block
```

```
    meanBaselineRT = mean(RT[PartOfExp == "baseline"]),

    # calculate normalized RTs
    AdjustedRT = RT - meanBaselineRT,

    # calculate subject-wise mean Adjusted RT across Blocks 1-4
    meanAdjustedRT = mean(AdjustedRT[PartOfExp == "main"])
  )
```

Now we want to check the distribution of adjuted RTs to make sure it seems reasonable, given our expectations about task performance.

Note that we expect baseline RTs to be faster on average than RTs during the experimental block, regardless of exposure condition. We expect this for two reasons. First, the baseline task occurred at the end of the experiment, after participants had adapted to the task. Second, *all* participants heard native accented speech during the baseline phase; hence, there was no need for accent adaptation during this phase.



Modeling strategy

Model building and assessment

RTs were analyzed using linear mixed effects regression, as implemented in the lme4 package (version 1.1-10: Bates, Maechler, Bolker, \& Walker, 2014) in R (R Core Team, 2014). Response accuracy (incorrect vs. correct response) was analyzed using mixed effects logistic regression (see Jaeger, 2008). All mixed effects models were specified with the maximal random effects structure justified by the experimental design: that is, by-subject and by-item random intercepts, by-subject random slopes for all design variables manipulated within subjects, and by-item random slopes for all design variables manipulated within items. If the definitionally maximal model failed to converge within ten thousand iterations, the model was systematically simplified in a step-wise fashion until the model converged. These steps involved removing correlations among random effects; dropping the random effects term with the least variance; and removing fixed effects that were inconsequential for the theory being tested (i.e., counterbalancing nuisance variables).

Variable coding

Unless otherwise specified, all numeric predictors were centered and categorical predictors were coded as sum contrasts, in order to reduce collinearity among predictors.

Experiment 1: Adaptation to Mandarin-accented English

Participants

Examine the number of participants per condition.

```
##                          Condition WorkerId
## 1           Control (native accent)       49
## 2 Mandarin-accented English (same)       51
```

Exp1 Response Times

Visualize the changes of RTs across blocks by condition.

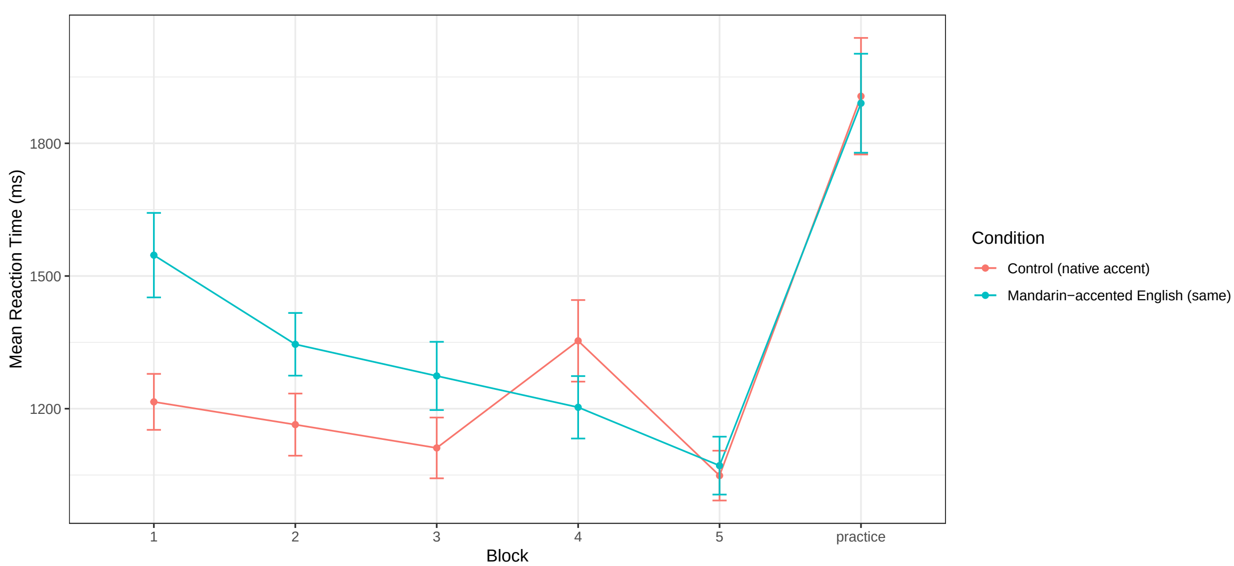


Figure 3: Average RTs by exposure condition in Experiment 1.

We assess the effect of exposure condition (Mandarin-accented English vs. control) on processing speed separately for RTs during the exposure phase and the test phase. To assess the *change* in RTs during the course of exposure, we split the 18-trial exposure phase into three blocks of 6 trials and use the resulting

Block variable as a categorical predictor of RTs. We use linear mixed-effects models to simultaneously model subject and item random effects.

Exposure A linear mixed effects model was fit to adjusted RTs for correct responses during the exposure phase.

Describe your fixed effects and random effects. Describe how each variable is coded.

In the exposure blocks, we applied a linear mixed effect model using maximal by-subject and by-item design (Eq. 1). The dependent variable is adjusted RT, and independent variables are 3-level factor exposure block using helmert coding and 2-level factor Condition using sum coding. Eq.1 AdjustedRT ~ Condition*Block + (1 + Block|WorkerId) + (1+Condition*Block|Word)

```
# Model specification:
# by-block analysis of RTs during EXPOSURE
dat_expo <-subset(exp1, (exp1$Block %in% c(1,2,3)) & (exp1$Error == 0))
m0 <- lmer(AdjustedRT ~ Condition*Block + (1 + Block|WorkerId) + (1+Condition*Block|Word),
                   data=dat_expo)
summary(m0)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: AdjustedRT ~ Condition * Block + (1 + Block | WorkerId) + (1 +
##     Condition * Block | Word)
##    Data: dat_expo
##
## REML criterion at convergence: 25118.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.2393 -0.4608 -0.1642  0.2344  9.4846
##
## Random effects:
##  Groups   Name                      Variance Std.Dev. Corr
##  WorkerId (Intercept)               118170   343.76
##           Block2                     39555   198.88   -0.34
##           Block3                     52711   229.59   -0.42  0.78
##  Word     (Intercept)                47707   218.42
##           ConditionAccented           6142    78.37   -0.59
##           Block2                     30789   175.47   -0.51  0.19
##           Block3                     16810   129.65   -0.73  0.73  0.67
##           ConditionAccented:Block2   22953   151.50    0.19 -0.32 -0.78 -0.41
##           ConditionAccented:Block3   11761   108.45    0.56 -0.30 -0.92 -0.84
##  Residual                           284327   533.22
##
##
##
##
##
##
##
##
##
##   0.57
```

```
##
## Number of obs: 1613, groups:  WorkerId, 100; Word, 24
##
## Fixed effects:
##                             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)                   283.23      61.11   46.34   4.635 2.93e-05 ***
## ConditionAccented            -122.65      44.73   70.65  -2.742 0.007726 **
## Block2                       -101.25      52.69   28.23  -1.922 0.064803 .
## Block3                       -175.36      48.36   35.62  -3.626 0.000892 ***
## ConditionAccented:Block2       48.21      49.47   29.78   0.975 0.337598
## ConditionAccented:Block3       77.45      46.10   41.87   1.680 0.100367
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) CndtnA Block2 Block3 CnA:B2
## CndtnAccntd -0.164
## Block2      -0.501  0.067
## Block3      -0.594  0.164  0.609
## CndtnAcc:B2  0.101 -0.428 -0.362 -0.158
## CndtnAcc:B3  0.210 -0.483 -0.319 -0.255  0.574
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

Test In the test blocks, we applied a linear mixed effect model using maximal by-subject and by-item design (Eq. 2). The dependent variable is adjusted RT, and independent variable is 2-level factor Condition using sum coding. Eq.2 AdjustedRT ~ Condition + (1 + Condition|WorkerId) + (1+Condition|Word)

```
# Model specification:
# by-block analysis of RTs during TEST
dat_test <-subset(exp1, (dat$Block ==4) & (dat$Error == 0))
m1 <- lmer(AdjustedRT ~ Condition + (1 + Condition|WorkerId) + (1+Condition|Word), data=dat_test)
summary(m1)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## AdjustedRT ~ Condition + (1 + Condition | WorkerId) + (1 + Condition |
##     Word)
##    Data: dat_test
##
## REML criterion at convergence: 8110.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8206 -0.4255 -0.1230  0.1886  9.5065
##
## Random effects:
##  Groups   Name              Variance Std.Dev. Corr
##  WorkerId (Intercept)        46194   214.93
##           ConditionAccented  40069   200.17   -0.02
##  Word     (Intercept)        32063   179.06
##           ConditionAccented   1609    40.12   1.00
```

```
##  Residual                    331017   575.34
## Number of obs: 516, groups:  WorkerId, 99; Word, 24
##
## Fixed effects:
##                    Estimate Std. Error     df t value Pr(>|t|)
## (Intercept)          180.97      53.85  38.45   3.361  0.00177 **
## ConditionAccented     82.23      40.04  57.01   2.054  0.04459 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## CndtnAccntd 0.157
## optimizer (nloptwrap) convergence code: 0 (OK)
## unable to evaluate scaled gradient
## Model failed to converge: degenerate  Hessian with 1 negative eigenvalues
```