acoustic_analysis

```r
cur_exp = "exp1"
features = c("duration", "meanIntensity", "meanpit")
# info = c('participant','verb','condition', 'word', 'word_num')
info = c('participant','item_id','location_condition', 'word', 'word_num')
bRemove_outliers = 0
```

This the analysis for exp1. The parameters of all exps can be seen at https: //github.com/Xinzhu-Fang/prosody_study_exp/blob/master/tAll_exps.csv.

The trial-by-trial design of this exp can be seen at https://github.com/ Xinzhu-Fang/prosody_study_exp/blob/master/exp1/tAll_trials.csv

```r
tAll_trials = read.csv(file.path('..', cur_exp, 'tAll_trials.csv'))
```

```r
df0 = read.csv(paste0('measure_', cur_exp, '.csv'), header = T)
df0$location_condition = NA
df0$item_id = NA

for (iR in 1:nrow(df0)){
  df0$location_condition[iR] = as.character(tAll_trials[tAll_trials$trial_id == df0$trialId[iR],'locat:
  df0$item_id[iR] = as.character(tAll_trials[tAll_trials$trial_id == df0$trialId[iR],'filler_or_item_ic
}

df1 = df0[startsWith(df0$item_id, "item"),]

# df0 = read.csv("measure_nonrhyming_84total_60No_24Yes_20181210.csv", header  = T)
# df0 = transform(df0,trialId=as.numeric(trialId))
# sort(df0$trialId, decreasing = FALSE)
# colnamesC(df1)

df2 = df1[df1$word != 'sp',]
# code for word_num
df2 <- df2 %>%
  dplyr::group_by(participant, trialId) %>%
  # dplyr::group_by(participant, question, trialId) %>%
  dplyr::mutate(word_num=1:dplyr::n()) %>%
  dplyr::select(c(info, features))
```

```
## Adding missing grouping variables: `trialId`
```

32 workers and 688 trials are included in this analysis.

```r
# write.csv(df2,'newdf.csv')
# code for getting Nth instance of question
# nthdf <- df1 %>%
#   group_by(participant,Verb, question, condition, word_num) %>%
#   mutate(Appearance=1:n())
#write.csv(nthdf,'nthdf.csv')

# subsetting it to relevant Nth appearance
```

```r
# workingdf <- nthdf %>%
#   filter (Appearance == 2)
#
# write.csv(workingdf,'workingdf2.csv')

normalize_data = function(df, remove_outliers){
  for(col_name in features){
    if(!is.numeric(df[[col_name]])){
      df[[col_name]] = as.numeric(df[[col_name]])
    }
    df[[col_name]] = scale(df[[col_name]])
    # there is surge of na after the first colling of the above line. tested by print(sum(is.na(df_Agen
    # print(sum(is.na(df_Agent)))

  }
  for(col_name in features){

    if(remove_outliers){
      df = df[df[[col_name]]>-2 & df[[col_name]]<2,]
      # print(sum(is.na(df_Agent)))

    }
  }
  return(df)
}

# process_data = function(file_name){
process_data = function(df){
  # df <- read.csv(file_name,header = TRUE, fileEncoding="UTF-8",na.strings=c("", "NA","--undefined--")
  # df <- na.omit(df)

  #df = df[df$wordlabel != 'sp']
  # df$verb = as.factor(df$verb)


  # df_Agent = df[(df$location_condition=='Agent' | df$location_condition=='Verb') & df$word_num=='3',]
  #
  # df_Verb = df[(df$location_condition=='Verb'| df$location_condition=='Patient') & df$word_num=='5',]
  #
  # df_Patient = df[(df$location_condition=='Patient'| df$location_condition=='Agent') & df$word_num==''

  df_Agent = df[(df$location_condition=='Agent' | df$location_condition=='Control') & df$word_num=='2',]
  # df_Agent inheri row hum from df

  df_Verb = df[(df$location_condition=='Verb'| df$location_condition=='Control') & df$word_num=='4',]

  df_Patient = df[(df$location_condition=='Patient'| df$location_condition=='Control') & df$word_num==''

  # print(sum(is.na(df_Agent)))

  # relevant_columns = c('participant','verb','condition','duration','meanIntensity','meanpit')
  # df_Agent = df_Agent[relevant_columns]
  # df_Verb = df_Verb[relevant_columns]
```

```r
  # df_Patient = df_Patient[relevant_columns]
  print(sum(is.na(df[df$word != 'sp',])))
  # df1[(df1$meanpit == '--undefined--') && (df1$word != 'sp'),]
  # it seems that the only undefined is meanpitch for sp

  # print(df_Verb)

  df_Verb = normalize_data(df_Verb, bRemove_outliers)
  df_Agent = normalize_data(df_Agent, bRemove_outliers)
  df_Patient = normalize_data(df_Patient, bRemove_outliers)
  # print(sum(is.na(df_Agent)))



  # return(list(df_Agent_duration, df_Agent_meanIntensity, df_Agent_meanpit, df_Patient_duration, df_Pa
  return(list(df_Verb, df_Agent, df_Patient))
}



#
# file_name = 'newdf.csv'

# c(df_Agent_duration, df_Agent_meanIntensity, df_Agent_meanpit, df_Patient_duration, df_Patient_meanIn
# c(df_Verb, df_Agent, df_Patient) %<-% process_data(file_name)
c(df_Verb, df_Agent, df_Patient) %<-% process_data(df2)
```

```
## [1] 0
```

```r
combine_datasets = function(Agent,Verb,Patient){
  Agent$condition = mapvalues(Agent$location_condition,c('Agent'),c('contrast'))
  Verb$condition = mapvalues(Verb$location_condition,c('Verb'),c('contrast'))
  Patient$condition = mapvalues(Patient$location_condition,c('Patient'),c('contrast'))

  Agent$Location = 'Agent'
  Verb$Location = 'Verb'
  Patient$Location = "Patient"

  return(rbind(Agent,Verb,Patient))
}

summarize_data = function(d, feature){
  # http://www.cookbook-r.com/Graphs/Plotting_means_and_error_bars_(ggplot2)/
  return(summarySE(d,measurevar=feature ,groupvars=c('Location','condition')))
}
plot_data = function(d,feature, title){
  print(ggplot(d, aes(x=Location, y=get(feature), fill=condition)) +
          geom_bar(position=position_dodge(), stat="identity") +
          geom_errorbar(aes(ymin=get(feature)-ci, ymax=get(feature)+ci),
                        width=.2,
                        position=position_dodge(.9))+
          xlab("Location") +
          ylab(paste0("normalized ", feature)) +
          scale_fill_hue(name="location_condition",
```

```
                        breaks=c("Control", "contrast"),
                        labels=c("NonContrastive", "Contrastive")) +
          ggtitle(title))
}
for (iF in features){
  print(iF)

  combined_dataset = combine_datasets(df_Agent, df_Verb, df_Patient)
  summarized_dataset= summarize_data(combined_dataset, iF)


  plot_data(summarized_dataset,iF, title= paste0('Effect of contrast on ', iF))
}
```
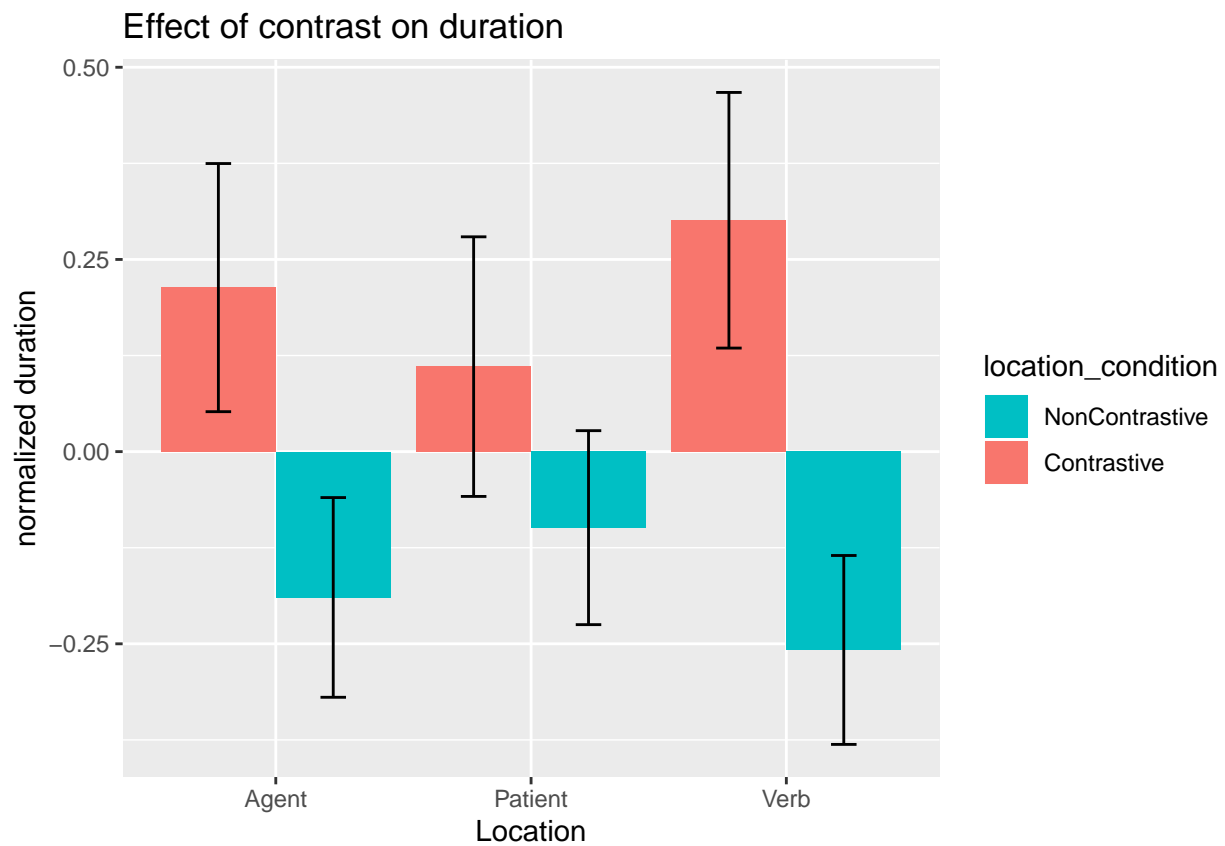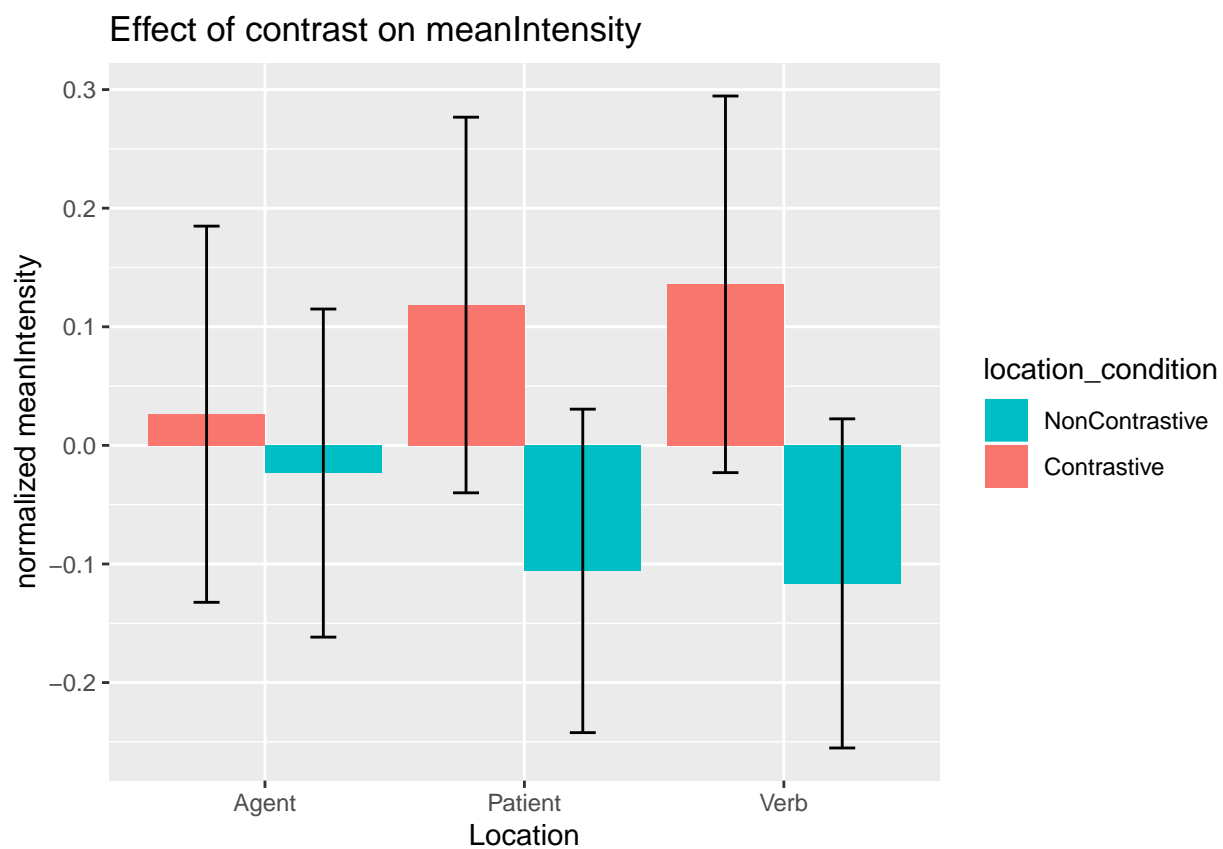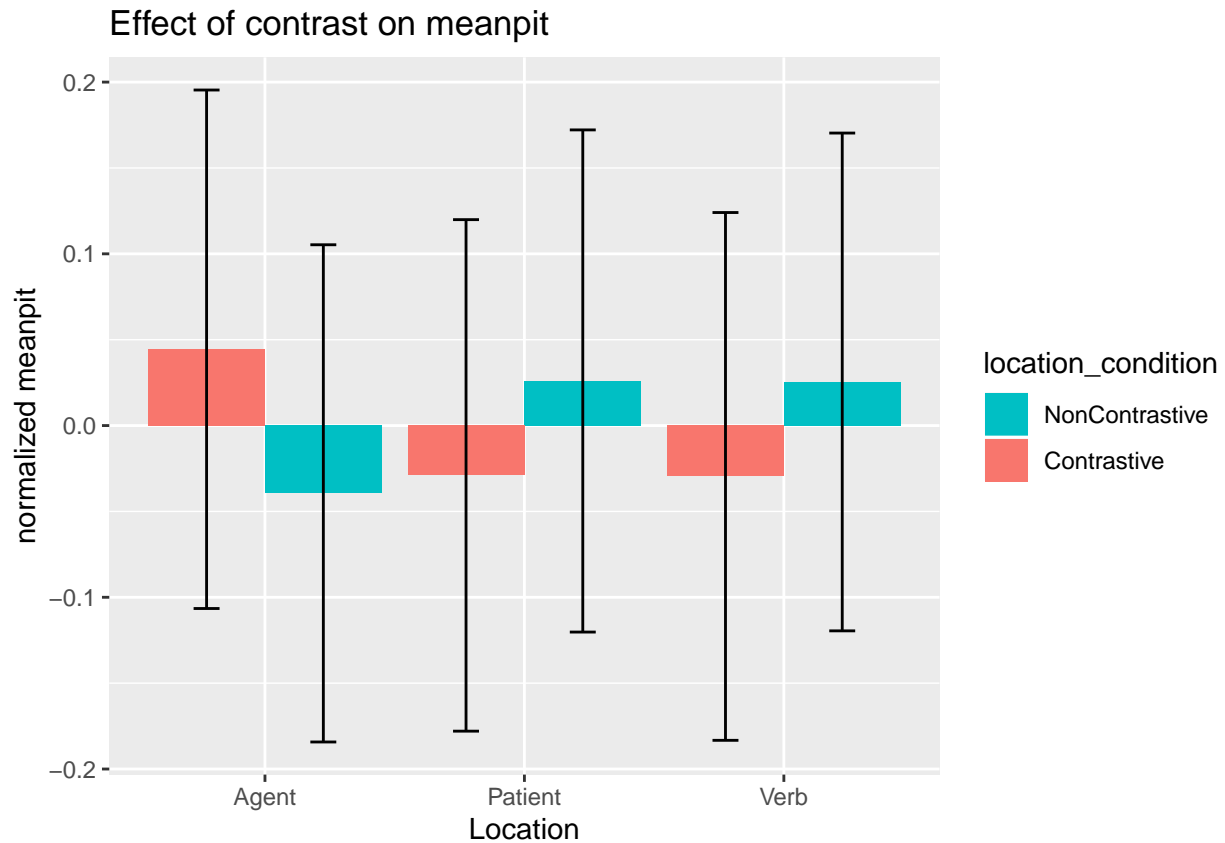
```
## [1] "duration"
```


Effect of contrast on duration

```
## [1] "meanIntensity"
```

Effect of contrast on meanIntensity

```
## [1] "meanpit"
```

Effect of contrast on meanpit


```
run_regression = function(location,observation){
  cat(" \n###", observation, "of", location, " \n")
  r = lmer(get(observation) ~ location_condition + (1 + location_condition|participant) + (1 + location_
  # r = lmer(get(observation) ~ location_condition  + (1 + location_condition | item_id), data=get(past
  print(summary(r))
  summary(r)
  cat(" \n")
}


for (iF in features){
  run_regression("Agent",iF)


  run_regression("Patient", iF)

  run_regression("Verb", iF)



}

##
## ### duration of Agent

## boundary (singular) fit: see ?isSingular
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## get(observation) ~ location_condition + (1 + location_condition |
##     participant) + (1 + location_condition | item_id)
##    Data: get(paste0("df_", location))
##
## REML criterion at convergence: 926.6
##
## Scaled residuals:
##     Min     1Q  Median     3Q     Max
## -2.2925 -0.4185 -0.0362  0.3110  8.9229
##
## Random effects:
##  Groups      Name                      Variance  Std.Dev. Corr
##  participant (Intercept)               5.016e-01 0.708267
##              location_conditionControl 7.613e-03 0.087255 -1.00
##  item_id     (Intercept)               2.529e-05 0.005029
##              location_conditionControl 6.709e-03 0.081911 -1.00
##  Residual                              6.461e-01 0.803792
## Number of obs: 357, groups:  participant, 31; item_id, 4
##
## Fixed effects:
##                           Estimate Std. Error t value
## (Intercept)                0.19262    0.14474   1.331
## location_conditionControl -0.42767    0.09773  -4.376
##
## Correlation of Fixed Effects:
##            (Intr)
## lctn_cndtnC -0.449
## convergence code: 0
## boundary (singular) fit: see ?isSingular
##
##
##
## ### duration of Patient

## boundary (singular) fit: see ?isSingular

## Linear mixed model fit by REML ['lmerMod']
## Formula:
## get(observation) ~ location_condition + (1 + location_condition |
##     participant) + (1 + location_condition | item_id)
##    Data: get(paste0("df_", location))
##
## REML criterion at convergence: 802.6
##
## Scaled residuals:
##     Min     1Q  Median     3Q     Max
## -6.3623 -0.3224 -0.0529  0.2501 11.5657
##
## Random effects:
##  Groups      Name                      Variance Std.Dev. Corr
##  participant (Intercept)               0.729082 0.85386
##              location_conditionPatient 0.009187 0.09585  -1.00
##  item_id     (Intercept)               0.036759 0.19173
```

```
##              location_conditionPatient 0.044654 0.21131  1.00
## Residual                                0.413349 0.64292
## Number of obs: 358, groups:  participant, 32; item_id, 4
##
## Fixed effects:
##                           Estimate Std. Error t value
## (Intercept)               -0.07905    0.18654  -0.424
## location_conditionPatient  0.22643    0.12762   1.774
##
## Correlation of Fixed Effects:
##            (Intr)
## lctn_cndtnP 0.220
## convergence code: 0
## boundary (singular) fit: see ?isSingular
##
##
##
## ### duration of Verb
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## get(observation) ~ location_condition + (1 + location_condition |
##      participant) + (1 + location_condition | item_id)
##    Data: get(paste0("df_", location))
##
## REML criterion at convergence: 865.3
##
## Scaled residuals:
##     Min     1Q  Median     3Q     Max
## -2.5808 -0.4390 -0.0649  0.2419  9.1578
##
## Random effects:
##  Groups      Name                   Variance Std.Dev. Corr
##  participant (Intercept)            0.31709  0.5631
##              location_conditionVerb 0.02352  0.1534   1.00
##  item_id     (Intercept)            0.02043  0.1429
##              location_conditionVerb 0.05280  0.2298   1.00
##  Residual                           0.55146  0.7426
## Number of obs: 351, groups:  participant, 32; item_id, 4
##
## Fixed effects:
##                        Estimate Std. Error t value
## (Intercept)             -0.2788     0.1359  -2.051
## location_conditionVerb   0.4851     0.1435   3.380
##
## Correlation of Fixed Effects:
##            (Intr)
## lctn_cndtnV 0.420
##
##
## ### meanIntensity of Agent

## boundary (singular) fit: see ?isSingular

## Linear mixed model fit by REML ['lmerMod']
## Formula:
```

```
## get(observation) ~ location_condition + (1 + location_condition |
##     participant) + (1 + location_condition | item_id)
##    Data: get(paste0("df_", location))
##
## REML criterion at convergence: 774.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.6109 -0.3692  0.0593  0.5648  2.3951
##
## Random effects:
##  Groups      Name                      Variance Std.Dev. Corr
##  participant (Intercept)               0.84438  0.9189
##              location_conditionControl 0.02925  0.1710   -0.20
##  item_id     (Intercept)               0.09765  0.3125
##              location_conditionControl 0.01142  0.1068   1.00
##  Residual                              0.36815  0.6067
## Number of obs: 357, groups:  participant, 31; item_id, 4
##
## Fixed effects:
##                           Estimate Std. Error t value
## (Intercept)               -0.02256    0.23368  -0.097
## location_conditionControl -0.12811    0.09099  -1.408
##
## Correlation of Fixed Effects:
##             (Intr)
## lctn_cndtnC 0.227
## convergence code: 0
## boundary (singular) fit: see ?isSingular
##
##
##
## ### meanIntensity of Patient

## boundary (singular) fit: see ?isSingular

## Linear mixed model fit by REML ['lmerMod']
## Formula:
## get(observation) ~ location_condition + (1 + location_condition |
##     participant) + (1 + location_condition | item_id)
##    Data: get(paste0("df_", location))
##
## REML criterion at convergence: 765.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4298 -0.5612 -0.0151  0.5428  3.5833
##
## Random effects:
##  Groups      Name                      Variance Std.Dev. Corr
##  participant (Intercept)               0.744297 0.8627
##              location_conditionPatient 0.007208 0.0849   1.00
##  item_id     (Intercept)               0.000000 0.0000
##              location_conditionPatient 0.016900 0.1300     NaN
##  Residual                              0.368845 0.6073
```

```
## Number of obs: 358, groups:  participant, 32; item_id, 4
##
## Fixed effects:
##                            Estimate Std. Error t value
## (Intercept)               -0.12829    0.16019  -0.801
## location_conditionPatient  0.21301    0.09379   2.271
##
## Correlation of Fixed Effects:
##            (Intr)
## lctn_cndtnP 0.020
## convergence code: 0
## boundary (singular) fit: see ?isSingular
##
##
##
## ### meanIntensity of Verb
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## get(observation) ~ location_condition + (1 + location_condition |
##     participant) + (1 + location_condition | item_id)
##    Data: get(paste0("df_", location))
##
## REML criterion at convergence: 736
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9560 -0.6146 -0.0408  0.6592  3.0914
##
## Random effects:
##  Groups      Name                   Variance Std.Dev. Corr
##  participant (Intercept)            0.77814  0.8821
##              location_conditionVerb 0.01384  0.1176   0.55
##  item_id     (Intercept)            0.12100  0.3478
##              location_conditionVerb 0.03953  0.1988   0.70
##  Residual                           0.33789  0.5813
## Number of obs: 351, groups:  participant, 32; item_id, 4
##
## Fixed effects:
##                         Estimate Std. Error t value
## (Intercept)              -0.1318     0.2387  -0.552
## location_conditionVerb    0.2505     0.1202   2.084
##
## Correlation of Fixed Effects:
##            (Intr)
## lctn_cndtnV 0.426
##
##
## ### meanpit of Agent

## boundary (singular) fit: see ?isSingular

## Linear mixed model fit by REML ['lmerMod']
## Formula:
## get(observation) ~ location_condition + (1 + location_condition |
##     participant) + (1 + location_condition | item_id)
```

```
##     Data: get(paste0("df_", location))
##
## REML criterion at convergence: 827.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.0120 -0.2566  0.0168  0.2016  3.6351
##
## Random effects:
##  Groups      Name                       Variance  Std.Dev.  Corr
##  participant (Intercept)                5.665e-01 7.527e-01
##              location_conditionControl  2.883e-01 5.369e-01 -0.22
##  item_id     (Intercept)                9.706e-09 9.852e-05
##              location_conditionControl  4.685e-09 6.845e-05 -0.77
##  Residual                               4.305e-01 6.561e-01
## Number of obs: 357, groups:  participant, 31; item_id, 4
##
## Fixed effects:
##                            Estimate Std. Error t value
## (Intercept)                -0.01107    0.14865  -0.074
## location_conditionControl  -0.06448    0.12384  -0.521
##
## Correlation of Fixed Effects:
##            (Intr)
## lctn_cndtnC -0.350
## convergence code: 0
## boundary (singular) fit: see ?isSingular
##
##
##
## ### meanpit of Patient

## boundary (singular) fit: see ?isSingular

## Linear mixed model fit by REML ['lmerMod']
## Formula:
## get(observation) ~ location_condition + (1 + location_condition |
##     participant) + (1 + location_condition | item_id)
##    Data: get(paste0("df_", location))
##
## REML criterion at convergence: 832.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8900 -0.2656  0.0307  0.3346  3.2988
##
## Random effects:
##  Groups      Name                       Variance Std.Dev. Corr
##  participant (Intercept)                0.670561 0.81888
##              location_conditionPatient  0.163581 0.40445  -0.37
##  item_id     (Intercept)                0.000000 0.00000
##              location_conditionPatient  0.009874 0.09937   NaN
##  Residual                               0.439300 0.66280
## Number of obs: 358, groups:  participant, 32; item_id, 4
##
```

```
## Fixed effects:
##                            Estimate Std. Error t value
## (Intercept)               -0.00109    0.15515  -0.007
## location_conditionPatient -0.07797    0.11542  -0.675
##
## Correlation of Fixed Effects:
##            (Intr)
## lctn_cndtnP -0.369
## convergence code: 0
## boundary (singular) fit: see ?isSingular
##
##
##
## ### meanpit of Verb
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## get(observation) ~ location_condition + (1 + location_condition |
##     participant) + (1 + location_condition | item_id)
##    Data: get(paste0("df_", location))
##
## REML criterion at convergence: 721.5
##
## Scaled residuals:
##     Min     1Q  Median      3Q     Max
## -4.4838 -0.2916  0.0288  0.3701  4.7927
##
## Random effects:
##  Groups      Name                   Variance Std.Dev. Corr
##  participant (Intercept)            0.735118 0.85739
##              location_conditionVerb 0.009631 0.09814  -0.73
##  item_id     (Intercept)            0.000820 0.02864
##              location_conditionVerb 0.012262 0.11073  -1.00
##  Residual                           0.341716 0.58456
## Number of obs: 351, groups:  participant, 32; item_id, 4
##
## Fixed effects:
##                        Estimate Std. Error t value
## (Intercept)             0.02086    0.15996   0.130
## location_conditionVerb -0.06517    0.08641  -0.754
##
## Correlation of Fixed Effects:
##            (Intr)
## lctn_cndtnV -0.331
##
# r = lmer(get(observation) ~ condition + (1 | participant) + (1 | verb), data=df)
```