

acoustic_analysis

```
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/Xinzh-Fang/prosody_study_exp/blob/master/tAll_exps.csv.

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

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

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$trialId[iR], 'location_condition'])
  df0$item_id[iR] = as.character(tAll_trials[tAll_trials$trial_id == df0$trialId[iR], 'filler_or_item_id'])
  # df0$present_num[iR] = as.numeric(rownames(tAll_trials[tAll_trials$trial_id == df0$trialId[iR],]))
}

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`
```

30 workers and 820 trials are included in this analysis.

```
# 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
# 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_Agent)))
    # 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=='1',]

  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=='1',]

  # 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]
# 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

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)) +

```

```

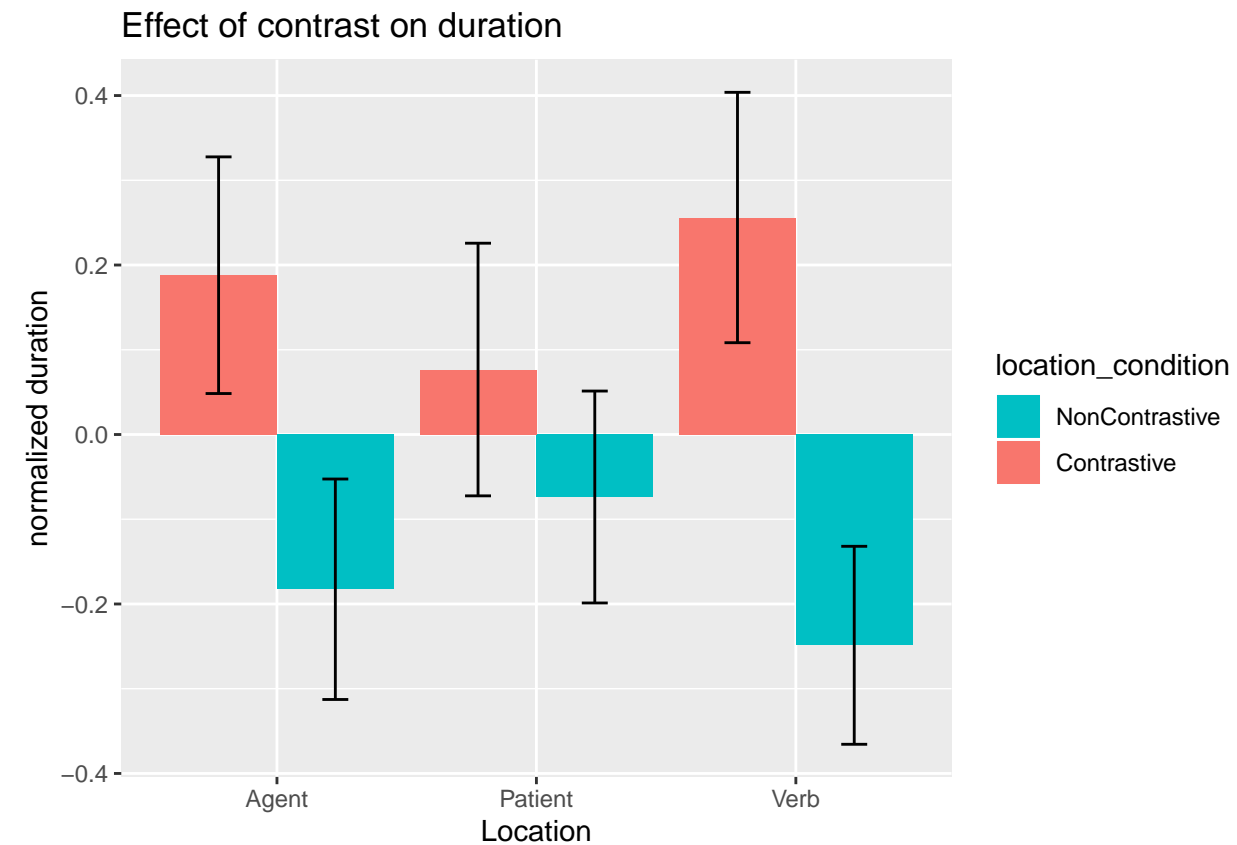
    scale_fill_hue(name="location_condition",
                   breaks=c("Control", "contrast"),
                   labels=c("NonContrastive", "Contrastive")) +
    ggtitle(title))
}
for (iF in features){
  print(iF)

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

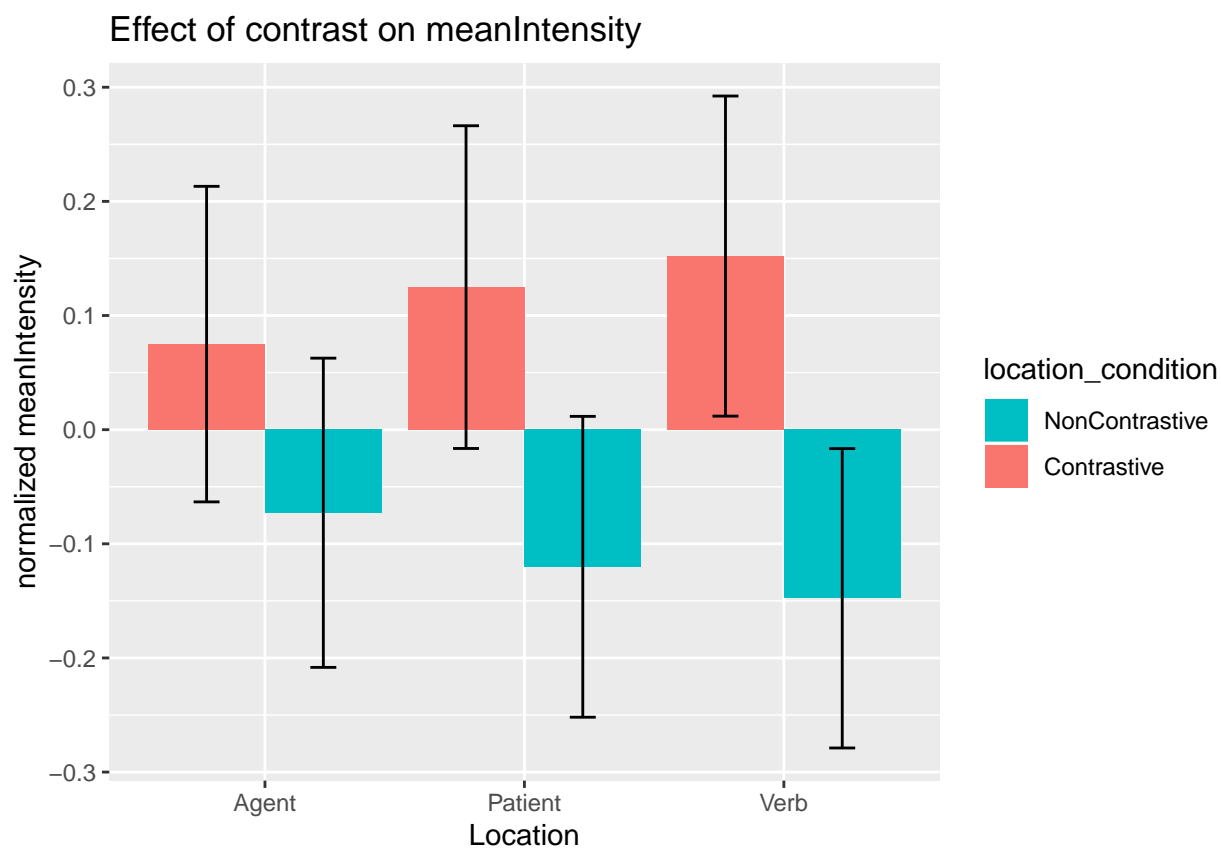
  plot_data(summarized_dataset,iF, title= paste0('Effect of contrast on ', iF))
}

```

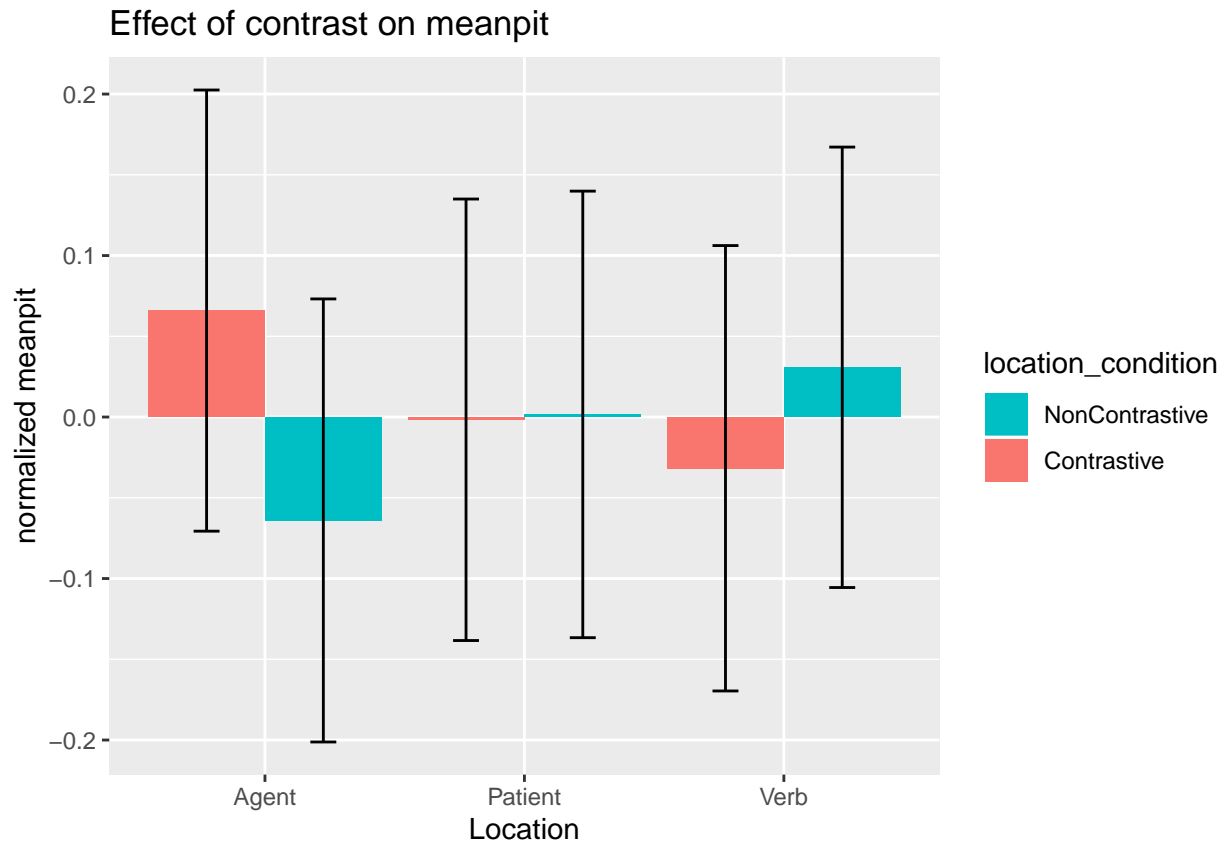
```
## [1] "duration"
```



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



```
## [1] "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: 1058.5
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.6275 -0.4178 -0.0442  0.3113  8.6997
##
## Random effects:
##   Groups      Name                Variance Std.Dev. Corr
##   participant (Intercept)          4.688e-01 0.684693
##               location_conditionControl 2.657e-04 0.016300 1.00
##   item_id     (Intercept)          8.895e-05 0.009432
##               location_conditionControl 9.267e-03 0.096263 -1.00
## Residual                        6.279e-01 0.792398
## Number of obs: 414, groups:  participant, 30; item_id, 4
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)          0.2112    0.1385   1.524
## location_conditionControl -0.4035    0.0921  -4.381
##
## Correlation of Fixed Effects:
##              (Intr)
## lctn_cndtnC -0.234
## 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: 989.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -5.0353 -0.3231 -0.0190  0.2574 10.6239
##
## Random effects:
##   Groups      Name                Variance Std.Dev. Corr
##   participant (Intercept)          0.60011  0.7747
##               location_conditionPatient 0.01830  0.1353  -1.00
##   item_id     (Intercept)          0.04665  0.2160

```

```

##           location_conditionPatient 0.03036  0.1742  1.00
## Residual                        0.51965  0.7209
## Number of obs: 412, groups:  participant, 30; item_id, 4
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)      -0.07725   0.18607  -0.415
## location_conditionPatient  0.16463   0.11555   1.425
##
## Correlation of Fixed Effects:
##           (Intr)
## lctn_cndtnP 0.155
## convergence code: 0
## boundary (singular) fit: see ?isSingular
##
##
## ### duration of Verb
## 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: 1001.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.0290 -0.4189 -0.0545  0.2680  9.7236
##
## Random effects:
##   Groups      Name                Variance Std.Dev. Corr
##   participant (Intercept)          0.30232  0.5498
##               location_conditionVerb 0.04098  0.2024  1.00
##   item_id     (Intercept)          0.02007  0.1417
##               location_conditionVerb 0.03941  0.1985  1.00
## Residual                        0.53909  0.7342
## Number of obs: 414, groups:  participant, 29; item_id, 4
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)      -0.2743   0.1350  -2.032
## location_conditionVerb  0.4465   0.1290   3.460
##
## Correlation of Fixed Effects:
##           (Intr)
## lctn_cndtnV 0.485
## convergence code: 0
## boundary (singular) fit: see ?isSingular
##
##
##

```



```

## ### 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: 925
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -5.7710 -0.3428  0.0734  0.5581  2.1033
##
## Random effects:
##   Groups             Name                Variance Std.Dev. Corr
##   participant (Intercept)                0.612466 0.78260
##               location_conditionControl 0.009811 0.09905  -0.16
##   item_id      (Intercept)                0.124956 0.35349
##               location_conditionControl 0.005040 0.07099   1.00
## Residual                                0.424299 0.65138
## Number of obs: 414, groups:  participant, 30; item_id, 4
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)          0.02961   0.23274   0.127
## location_conditionControl -0.13897   0.07609  -1.826
##
## Correlation of Fixed Effects:
##              (Intr)
## lctn_cndtnC 0.209
## 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: 927.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.3617 -0.4795 -0.0132  0.5146  3.6817
##
## Random effects:
##   Groups             Name                Variance Std.Dev. Corr

```

```

## participant (Intercept)          0.6433031 0.80206
##           location_conditionPatient 0.0007132 0.02671 1.00
## item_id (Intercept)              0.0000000 0.00000
##           location_conditionPatient 0.0174660 0.13216 NaN
## Residual                        0.4400266 0.66334
## Number of obs: 412, groups: participant, 30; item_id, 4
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)      -0.1397    0.1547  -0.903
## location_conditionPatient  0.2499    0.0936   2.670
##
## Correlation of Fixed Effects:
##           (Intr)
## lctn_cndtnP -0.095
## convergence code: 0
## boundary (singular) fit: see ?isSingular
##
##
## ### meanIntensity of Verb
## 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: 847.7
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.6432 -0.6308 -0.0516  0.6495  3.2893
##
## Random effects:
##   Groups      Name                Variance Std.Dev. Corr
## participant (Intercept)          0.866730 0.93098
##           location_conditionVerb 0.003878 0.06228 1.00
## item_id (Intercept)             0.116009 0.34060
##           location_conditionVerb 0.018870 0.13737 1.00
## Residual                        0.339431 0.58261
## Number of obs: 414, groups: participant, 29; item_id, 4
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)      -0.19836    0.24648  -0.805
## location_conditionVerb  0.30175    0.09051   3.334
##
## Correlation of Fixed Effects:
##           (Intr)
## lctn_cndtnV 0.545
## convergence code: 0
## boundary (singular) fit: see ?isSingular

```

```

##
##
##
## ### 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: 988.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.5917 -0.2774  0.0049  0.2450  3.3038
##
## Random effects:
##   Groups             Name                Variance Std.Dev. Corr
##   participant (Intercept)                0.513890 0.71686
##               location_conditionControl 0.121890 0.34913 -0.25
##   item_id      (Intercept)                0.010452 0.10224
##               location_conditionControl 0.006902 0.08308 -1.00
## Residual                        0.500149 0.70721
## Number of obs: 414, groups:  participant, 30; item_id, 4
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)          0.03464    0.15088   0.230
## location_conditionControl -0.10487    0.10513  -0.997
##
## Correlation of Fixed Effects:
##              (Intr)
## lctn_cndtnC -0.434
## convergence code: 0
## boundary (singular) fit: see ?isSingular
##
##
##
## ### meanpit of Patient

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 0.00252213
## (tol = 0.002, component 1)

## 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: 939.2
##
## Scaled residuals:

```

```

##      Min      1Q  Median      3Q      Max
## -3.7880 -0.2432  0.0182  0.3437  3.3993
##
## Random effects:
##      Groups      Name              Variance Std.Dev. Corr
## participant (Intercept)          5.829e-01 0.763473
##      location_conditionPatient 2.996e-01 0.547392 -0.34
## item_id      (Intercept)          4.676e-03 0.068382
##      location_conditionPatient 1.039e-05 0.003223 1.00
## Residual                          4.238e-01 0.651021
## Number of obs: 412, groups: participant, 30; item_id, 4
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)      -0.040341   0.152433  -0.265
## location_conditionPatient -0.003267   0.123012  -0.027
##
## Correlation of Fixed Effects:
##      (Intr)
## lctn_cndtnP -0.375
## convergence code: 0
## Model failed to converge with max|grad| = 0.00252213 (tol = 0.002, component 1)
##
##
##
## ### meanpit of Verb
## 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: 848.4
##
## Scaled residuals:
##      Min      1Q  Median      3Q      Max
## -4.6428 -0.2972  0.0447  0.3578  4.7974
##
## Random effects:
##      Groups      Name              Variance Std.Dev. Corr
## participant (Intercept)          7.198e-01 0.8484200
##      location_conditionVerb 4.132e-03 0.0642815 -1.00
## item_id      (Intercept)          0.000e+00 0.0000000
##      location_conditionVerb 1.171e-08 0.0001082  NaN
## Residual                          3.587e-01 0.5988795
## Number of obs: 414, groups: participant, 29; item_id, 4
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)      0.002704   0.163626   0.017
## location_conditionVerb -0.072550   0.060337  -1.202
##

```

```
## Correlation of Fixed Effects:
##          (Intr)
## lctn_cndtnV -0.360
## convergence code: 0
## boundary (singular) fit: see ?isSingular
##
##
# r = lmer(get(observation) ~ condition + (1 | participant) + (1 | verb), data=df)
```