Preliminary analysis

August 10, 2016

The following code loads the csv file with the word list used in the task and the result file.

```
results_raw <- read.csv("analysis/results.csv")
words <- read.csv("task/task-src/words.csv")
```

Now we can merge the word list and result files into one single data frame.

```
results <- merge(results_raw, words, by.x = "word", by.y = "lex")
```

To check the duration of each word token as spoken by the participants, we first split the data into two subsets: one with monosyllabic words and one with bisyllabic words.
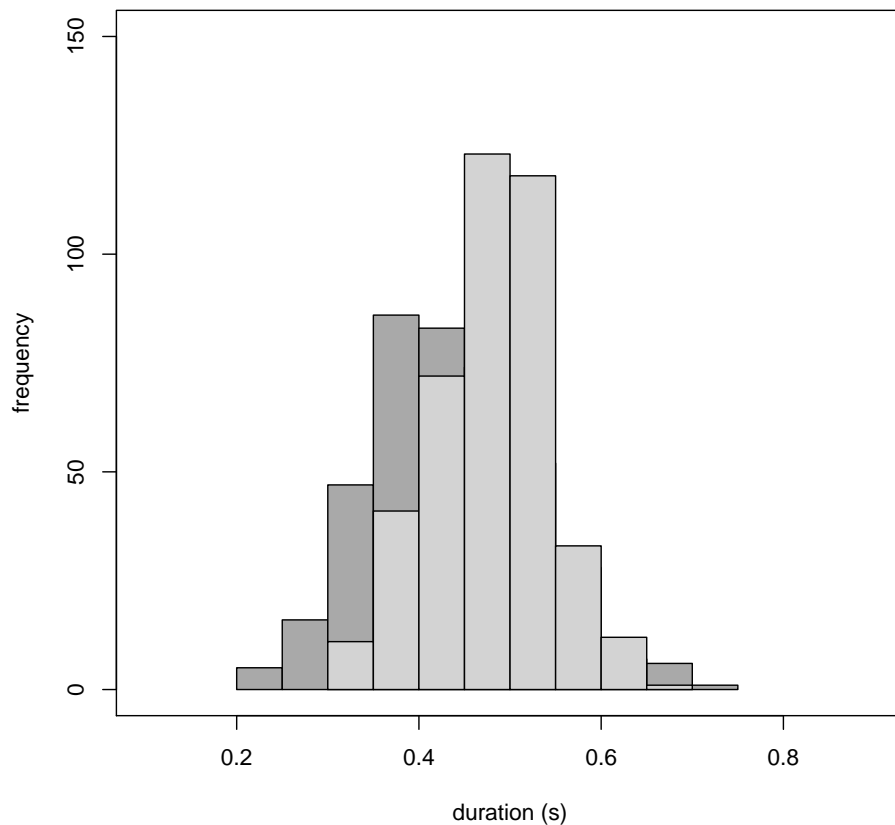
```
mono <- subset(results, syl_no == "mono")
bi <- subset(results, syl_no == "bi")
```

We can now separately check the duration of monosyllabic and bisyllabic words.

```
plot(
    c(), c(),
    xlim = c(0.1,0.9),
    ylim = c(0,150),
    xlab = "duration (s)", ylab = "frequency"
)

hist(
    mono$dur_word, add = TRUE, col = "darkgrey"
)

hist(
    bi$dur_word, add = TRUE, col = "lightgrey"
)
```

The mean length of monosyllabic words is 0.44 seconds, while that of bisyllabic words is 0.48 seconds.

Since I am looking at the duration of the normalised modal voicing of the vocalic gesture, we should first check if there is a significant difference in length of modal voicing between mono and bisyllabic words.

```r
norm_abs_voic_mono_density <- density(mono$norm_abs_voic)
norm_abs_voic_bi_density <- density(bi$norm_abs_voic)

norm_abs_voic_density_x <- c(norm_abs_voic_mono_density$x,
                             norm_abs_voic_bi_density$x)
norm_abs_voic_density_y <- c(norm_abs_voic_mono_density$y,
                             norm_abs_voic_bi_density$y)

x.limits <- range(norm_abs_voic_density_x)
y.limits <- range(norm_abs_voic_density_y)
```
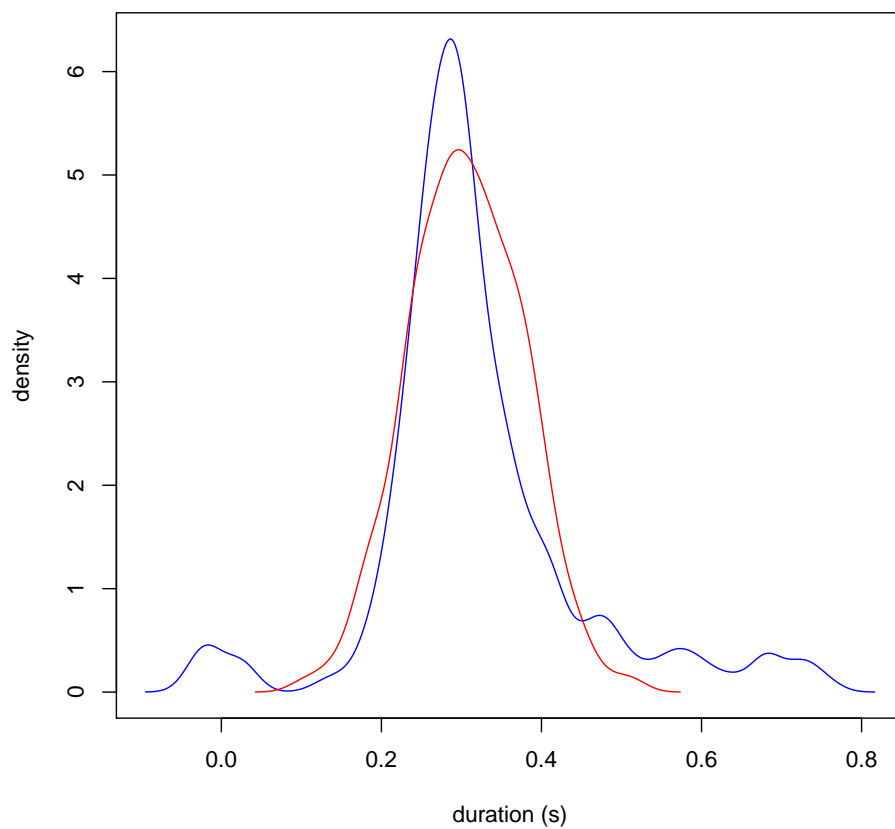
```
plot(
c(), c(),
xlim = x.limits,
ylim = y.limits,
xlab = "duration (s)", ylab = "density"
)

lines(norm_abs_voic_mono_density, lw = 1, col = "blue")
lines(norm_abs_voic_bi_density, lw = 1, col = "red")
```



```
shapiro.test(mono$norm_abs_voic)

##
##  Shapiro-Wilk normality test
##
## data:  mono$norm_abs_voic
```

```
## W = 0.88286, p-value < 2.2e-16

shapiro.test(bi$norm_abs_voic)

##
##  Shapiro-Wilk normality test
##
## data:  bi$norm_abs_voic
## W = 0.99834, p-value = 0.9647
```

A Wilcox test shows that the mean length of modal voicing is significantly longer in monosyllabic words (0.32) than in bisyllabic words (0.31). Subsequent tests will be performed for monosyllabic and bisyllabic words separately.

```
wilcox.test(mono$norm_abs_voic, bi$norm_abs_voic)

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  mono$norm_abs_voic and bi$norm_abs_voic
## W = 83416, p-value = 0.9555
## alternative hypothesis: true location shift is not equal to 0
```

1 Monosyllabic words

1.1 Stops

We can now have a look at monosyllabic words, starting from CVCC words (ending in a geminate stop). These words start with either an aspirated, a non-aspirated or a sonorant consonant.

```
mono_stop <- subset(mono, manner == "stop")

bi_stop <- subset(bi, manner == "stop")

mono_stop_asp <- mono_stop$norm_abs_voic[mono_stop$asp == "yes"]
mono_stop_nasp <- mono_stop$norm_abs_voic[mono_stop$asp == "no"]

norm_abs_voic_asp_density <- density(mono_stop_asp)
norm_abs_voic_nasp_density <- density(mono_stop_nasp)
```

```r
norm_abs_voic_density_x <- c(norm_abs_voic_asp_density$x,
                             norm_abs_voic_nasp_density$x)
norm_abs_voic_density_y <- c(norm_abs_voic_asp_density$y,
                             norm_abs_voic_nasp_density$y)

x.limits <- range(norm_abs_voic_density_x)
y.limits <- range(norm_abs_voic_density_y)

plot(
c(), c(),
xlim = x.limits,
ylim = y.limits,
xlab = "duration (s)", ylab = "density"
)

lines(norm_abs_voic_asp_density, lw = 1, col = "blue")
lines(norm_abs_voic_nasp_density, lw = 1, col = "red")
```
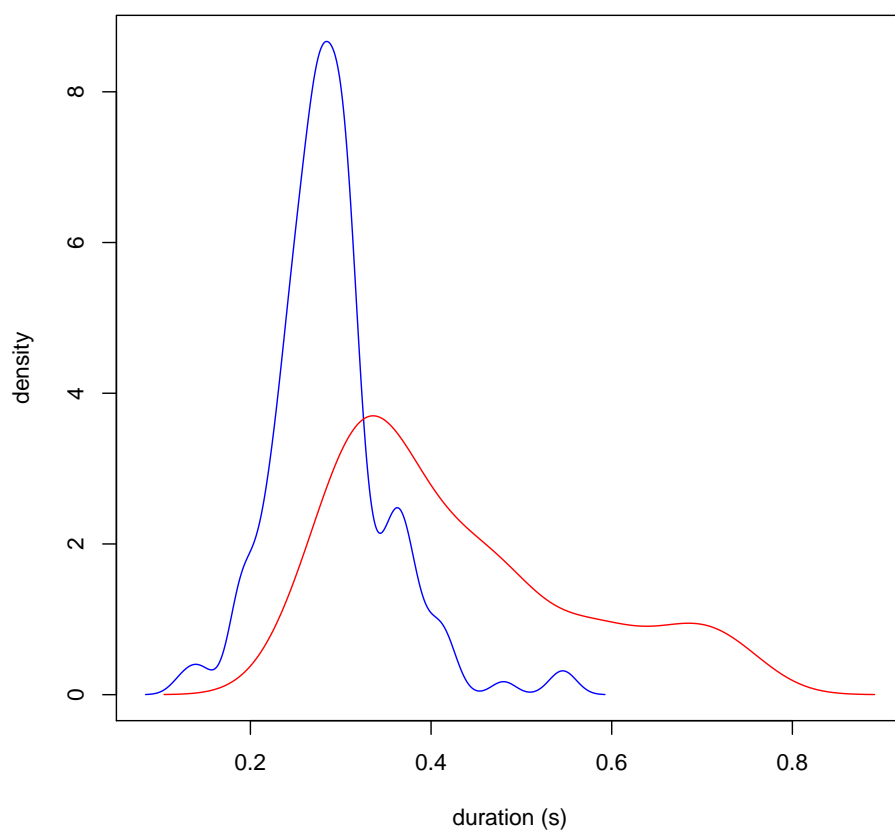
```
shapiro.test(mono_stop_asp)

##
##  Shapiro-Wilk normality test
##
## data:  mono_stop_asp
## W = 0.94276, p-value = 2.565e-06

shapiro.test(mono_stop_nasp)

##
##  Shapiro-Wilk normality test
##
## data:  mono_stop_nasp
## W = 0.90257, p-value = 1.673e-07
```

Since the distributions of the duration of modal voicing in the two con-

ditions (pre-aspirated and non-aspirated) are significantly different from the normal distribution, we should perform a Wilcoxon test. According to the test, the mean duration of modal voicing in words ending with a pre-aspirated stop is significantly shorter (0.29 seconds) than in words ending with a non-aspirated stop (0.43 seconds).

```r
boxplot(norm_abs_voic ~ asp,
        data = mono_stop,
        names = c("non-asp", "asp"),
        ylab = "duration (s)",
     main = "Duration (in seconds) of modal voicing in CVCC words"
        )
wilcox.test(norm_abs_voic ~ asp, data = mono_stop)

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  norm_abs_voic by asp
## W = 18010, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```

Duration (in seconds) of modal voicing in CVCC words
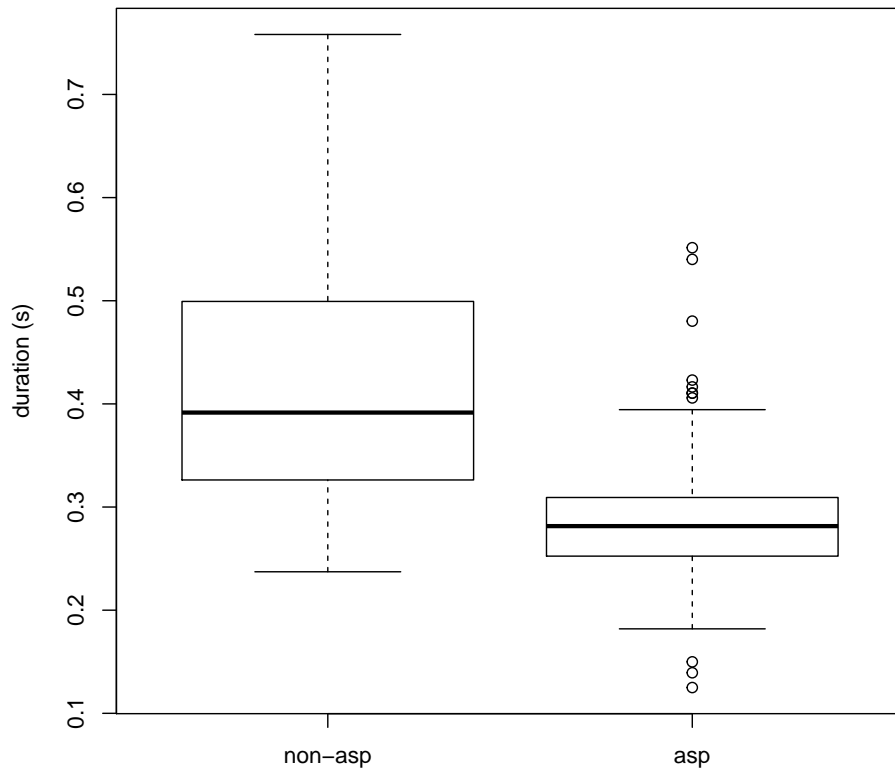
1.2 Nasals

```
mono_nas <- subset(mono, manner == "nasal")
```

```
mono_nas_asp <- mono_nas$norm_abs_voic[mono_nas$asp == "yes"]
mono_nas_nasp <- mono_nas$norm_abs_voic[mono_nas$asp == "no"]

norm_abs_voic_asp_density <- density(mono_nas_asp)
norm_abs_voic_nasp_density <- density(mono_nas_nasp)

norm_abs_voic_density_x <- c(norm_abs_voic_asp_density$x,
                             norm_abs_voic_nasp_density$x)
norm_abs_voic_density_y <- c(norm_abs_voic_asp_density$y,
                             norm_abs_voic_nasp_density$y)
```
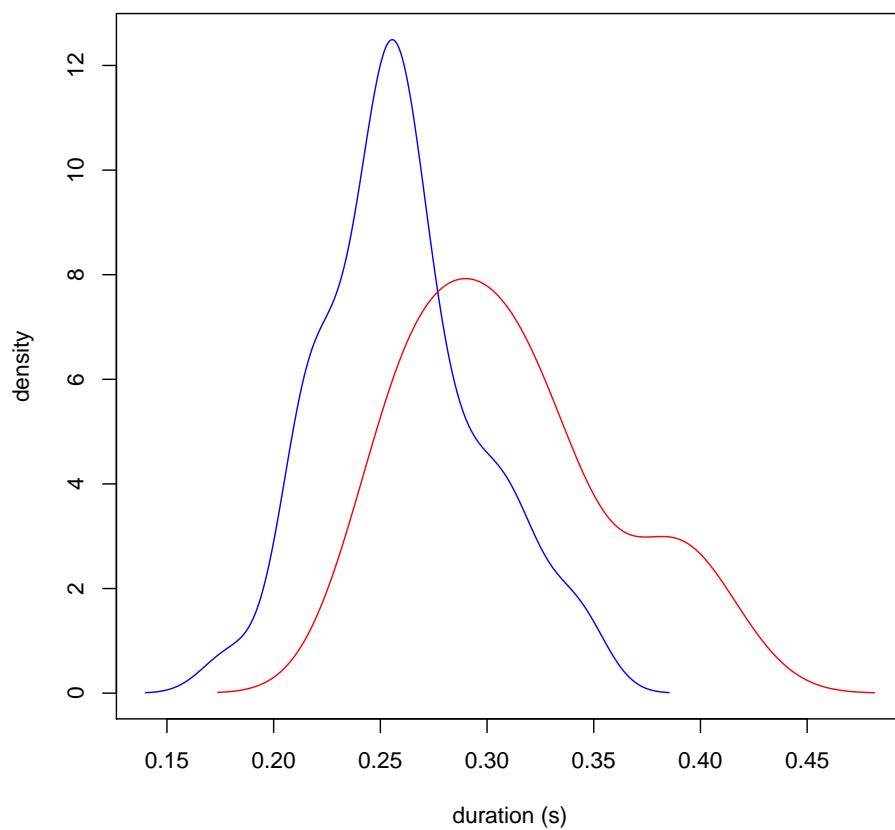
```r
x.limits <- range(norm_abs_voic_density_x)
y.limits <- range(norm_abs_voic_density_y)

plot(
c(), c(),
xlim = x.limits,
ylim = y.limits,
xlab = "duration (s)", ylab = "density"
)

lines(norm_abs_voic_asp_density, lw = 1, col = "blue")
lines(norm_abs_voic_nasp_density, lw = 1, col = "red")
```
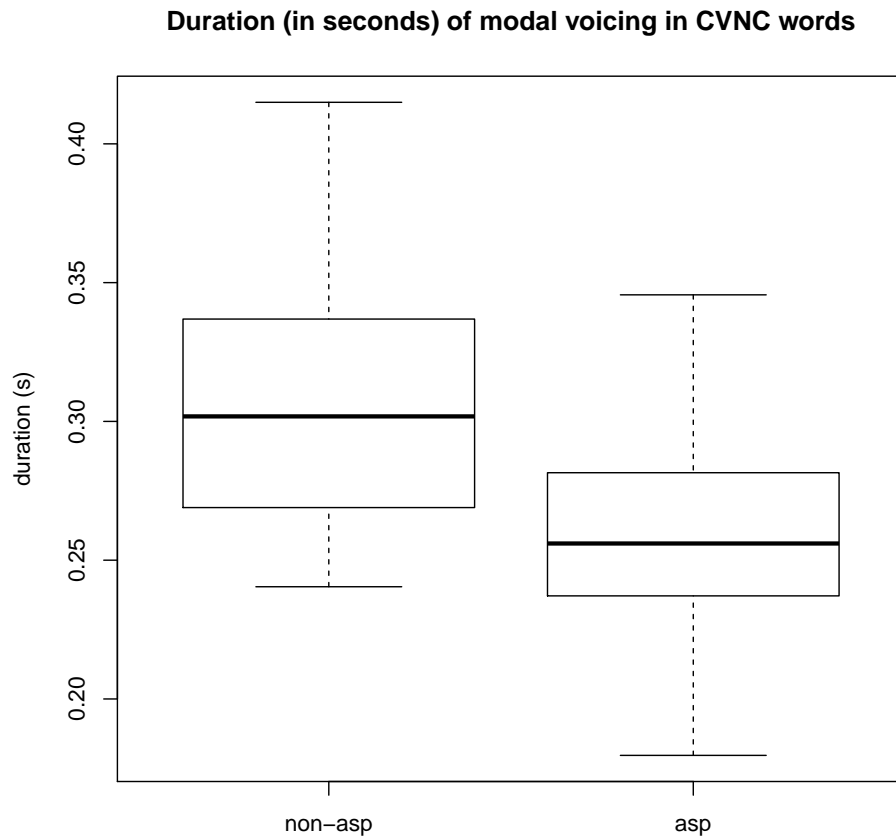


```r
boxplot(norm_abs_voic ~ asp,
        data = mono_nas,
```

```
        names = c("non-asp", "asp"),
        ylab = "duration (s)",
    main = "Duration (in seconds) of modal voicing in CVNC words"
        )
```

Duration (in seconds) of modal voicing in CVNC words


```
shapiro.test(mono_nas_asp)

##
##  Shapiro-Wilk normality test
##
## data:  mono_nas_asp
## W = 0.97229, p-value = 0.4241

shapiro.test(mono_nas_nasp)

##
```

```
##  Shapiro-Wilk normality test
##
## data:  mono_nas_nasp
## W = 0.94149, p-value = 0.1097
```

```
t.test(norm_abs_voic ~ asp, data = mono_nas)
```

```
##
##  Welch Two Sample t-test
##
## data:  norm_abs_voic by asp
## t = 4.6522, df = 50.092, p-value = 2.433e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   0.02829407 0.07128385
## sample estimates:
##   mean in group no mean in group yes
##          0.3098634           0.2600745
```

According to a two-sample t-test, the duration of modal voicing in words ending with an ṆC cluster (0.26 seconds) is shorter than in words ending with an NC cluster (0.31 seconds).

1.3   Laterals

```
mono_lat <- subset(mono, manner == "lateral")
```

```
mono_lat_asp <- mono_lat$norm_abs_voic[mono_lat$asp == "yes"]
mono_lat_nasp <- mono_lat$norm_abs_voic[mono_lat$asp == "no"]

norm_abs_voic_asp_density <- density(mono_lat_asp)
norm_abs_voic_nasp_density <- density(mono_lat_nasp)

norm_abs_voic_density_x <- c(norm_abs_voic_asp_density$x,
                             norm_abs_voic_nasp_density$x)
norm_abs_voic_density_y <- c(norm_abs_voic_asp_density$y,
                             norm_abs_voic_nasp_density$y)

x.limits <- range(norm_abs_voic_density_x)
y.limits <- range(norm_abs_voic_density_y)

plot(
```
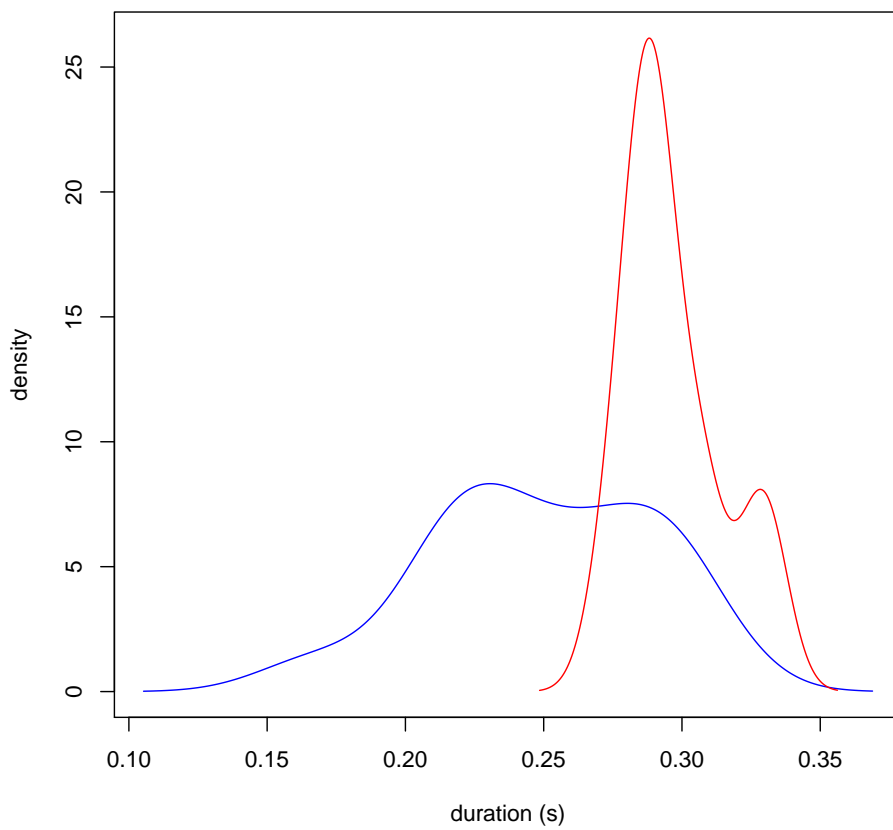
```
c(), c(),
xlim = x.limits,
ylim = y.limits,
xlab = "duration (s)", ylab = "density"
)

lines(norm_abs_voic_asp_density, lw = 1, col = "blue")
lines(norm_abs_voic_nasp_density, lw = 1, col = "red")
```
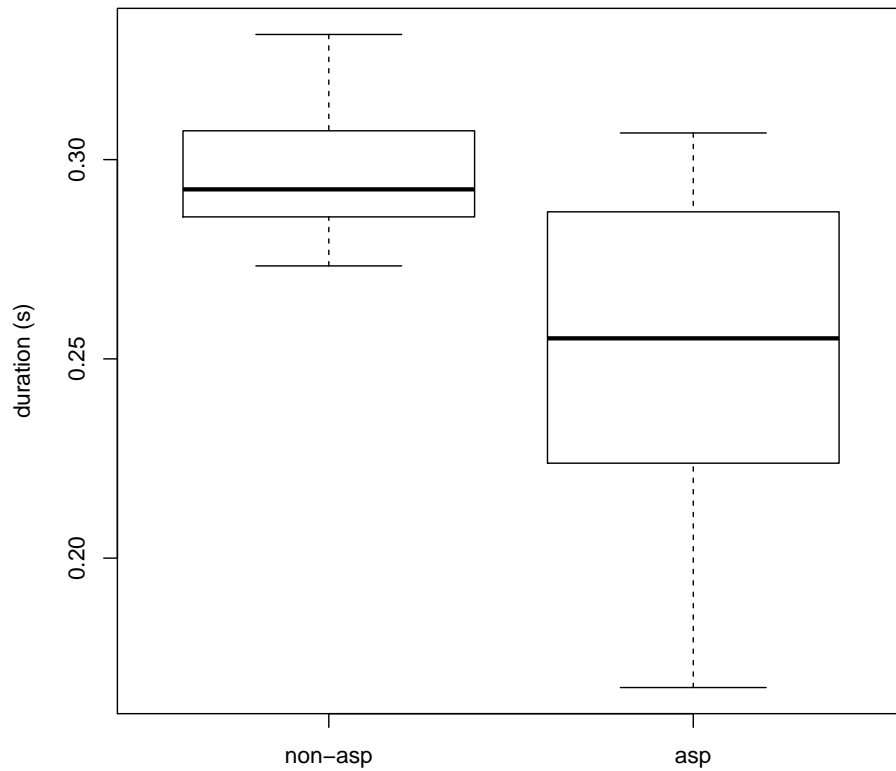


```
boxplot(norm_abs_voic ~ asp,
        data = mono_lat,
        names = c("non-asp", "asp"),
        ylab = "duration (s)",
    main = "Duration (in seconds) of modal voicing in CVLC words"
        )
```

Duration (in seconds) of modal voicing in CVLC words


```
shapiro.test(mono_lat_asp)

##
##  Shapiro-Wilk normality test
##
## data:  mono_lat_asp
## W = 0.94973, p-value = 0.5202

shapiro.test(mono_lat_nasp)

##
##  Shapiro-Wilk normality test
##
## data:  mono_lat_nasp
## W = 0.89635, p-value = 0.1423

t.test(norm_abs_voic ~ asp, data = mono_lat)
```
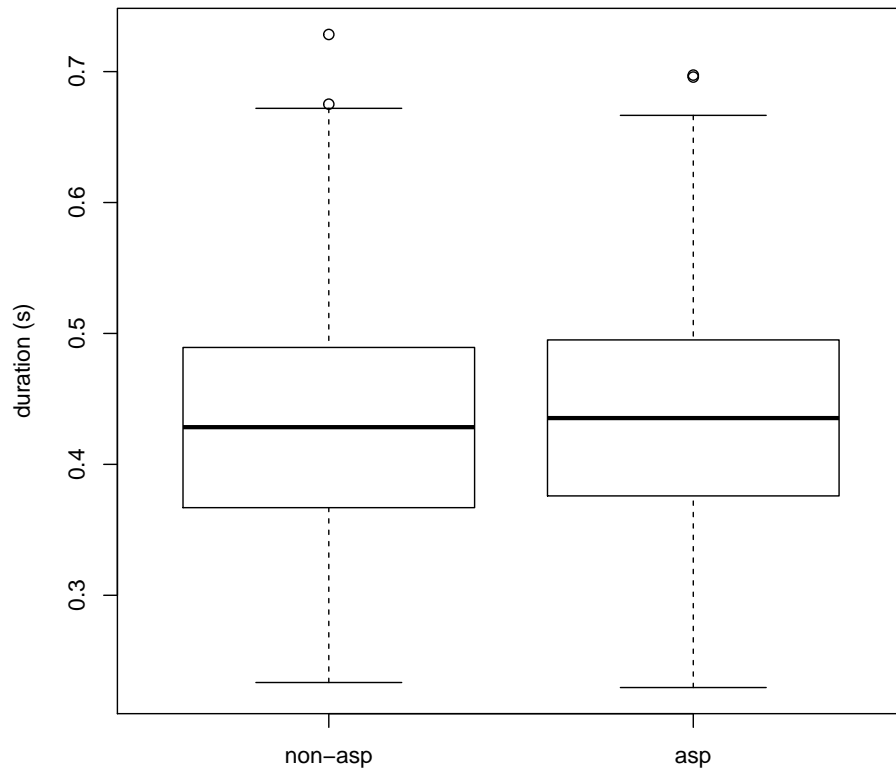
13

```
##
##  Welch Two Sample t-test
##
## data:  norm_abs_voic by asp
## t = 4.1099, df = 20.43, p-value = 0.0005241
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   0.02324930 0.07104155
## sample estimates:
##  mean in group no mean in group yes
##         0.2969002         0.2497548
```

It is good to check whether the duration of monosyllabic words is affected by the presence vs. absence of preaspiration.

```
boxplot(dur_word ~ asp,
        data = mono,
        names = c("non-asp", "asp"),
        ylab = "duration (s)",
     main = "Word duration (in seconds) depending on aspiration"
        )
```

Word duration (in seconds) depending on aspiration


```
shapiro.test(mono$dur_word[mono$asp == "yes"])

##
##  Shapiro-Wilk normality test
##
## data:  mono$dur_word[mono$asp == "yes"]
## W = 0.98811, p-value = 0.05986

shapiro.test(mono$dur_word[mono$asp == "no"])

##
##  Shapiro-Wilk normality test
##
## data:  mono$dur_word[mono$asp == "no"]
## W = 0.98869, p-value = 0.159
```

```
t.test(dur_word ~ asp, data = mono)

##
##  Welch Two Sample t-test
##
## data:  dur_word by asp
## t = -0.36443, df = 358.23, p-value = 0.7158
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.02116269  0.01454560
## sample estimates:
##  mean in group no mean in group yes
##           0.4333422         0.4366508
```

According to a *t*-test, the duration of monosyllabic words is not affected by the presence vs. absence of preaspiration. Thus, the difference in the mean duration of modal voicing can't be attributed to the differences in word duration.

2   Bisyllabic words

2.1   Stops

```
bi_stop_asp <- bi_stop$norm_abs_voic[bi_stop$asp == "yes"]
bi_stop_nasp <- bi_stop$norm_abs_voic[bi_stop$asp == "no"]

norm_abs_voic_asp_density <- density(bi_stop_asp)
norm_abs_voic_nasp_density <- density(bi_stop_nasp)

norm_abs_voic_density_x <- c(norm_abs_voic_asp_density$x,
                             norm_abs_voic_nasp_density$x)
norm_abs_voic_density_y <- c(norm_abs_voic_asp_density$y,
                             norm_abs_voic_nasp_density$y)

x.limits <- range(norm_abs_voic_density_x)
y.limits <- range(norm_abs_voic_density_y)

plot(
c(), c(),
xlim = x.limits,
ylim = y.limits,
xlab = "duration (s)", ylab = "density"
```
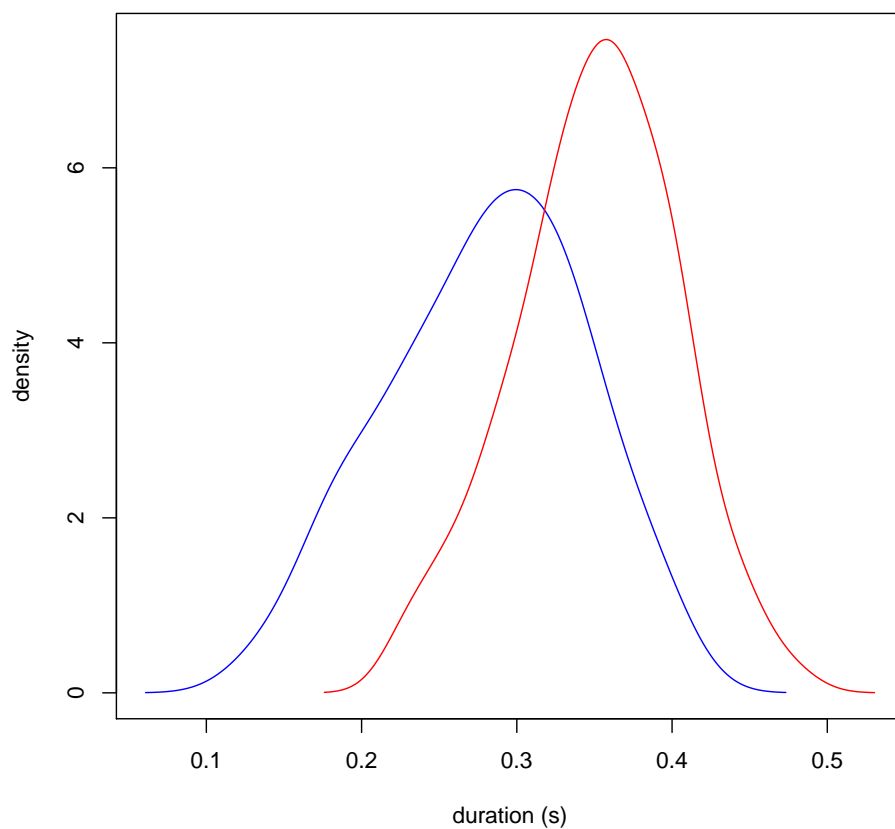
```
)

lines(norm_abs_voic_asp_density, lw = 1, col = "blue")
lines(norm_abs_voic_nasp_density, lw = 1, col = "red")
```



```
shapiro.test(bi_stop_asp)

##
##  Shapiro-Wilk normality test
##
## data:  bi_stop_asp
## W = 0.98478, p-value = 0.41

shapiro.test(bi_stop_nasp)

##
```

```
##  Shapiro-Wilk normality test
##
## data:  bi_stop_nasp
## W = 0.99166, p-value = 0.727
```

```
boxplot(norm_abs_voic ~ asp,
        data = bi_stop,
        names = c("non-asp", "asp"),
        ylab = "duration (s)",
     main = "Duration (in seconds) of modal voicing in CVCCV words"
        )
```

Duration (in seconds) of modal voicing in CVCCV words


```
wilcox.test(norm_abs_voic ~ asp, data = mono_stop)
```

```
##
```
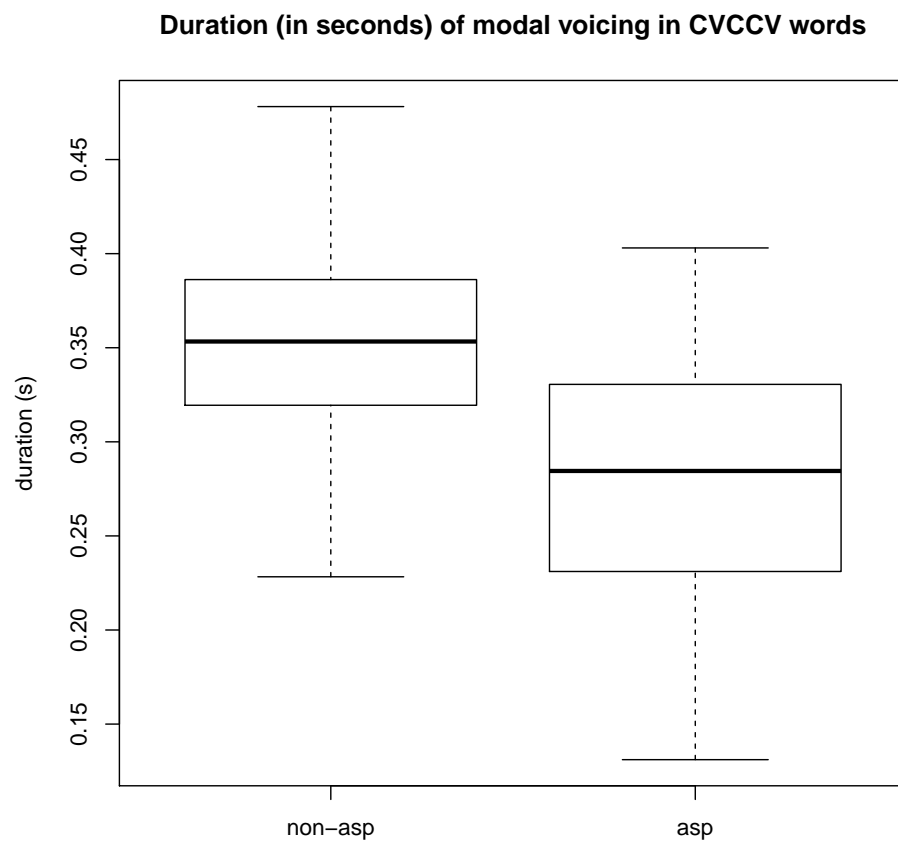
```
##  Wilcoxon rank sum test with continuity correction
##
## data:  norm_abs_voic by asp
## W = 18010, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```

2.2 Nasals

```r
bi_nas <- subset(bi, manner == "nasal")
```

```r
bi_nas_asp <- bi_nas$norm_abs_voic[bi_nas$asp == "yes"]
bi_nas_nasp <- bi_nas$norm_abs_voic[bi_nas$asp == "no"]

norm_abs_voic_asp_density <- density(bi_nas_asp)
norm_abs_voic_nasp_density <- density(bi_nas_nasp)

norm_abs_voic_density_x <- c(norm_abs_voic_asp_density$x,
                             norm_abs_voic_nasp_density$x)
norm_abs_voic_density_y <- c(norm_abs_voic_asp_density$y,
                             norm_abs_voic_nasp_density$y)

x.limits <- range(norm_abs_voic_density_x)
y.limits <- range(norm_abs_voic_density_y)

plot(
c(), c(),
xlim = x.limits,
ylim = y.limits,
xlab = "duration (s)", ylab = "density"
)

lines(norm_abs_voic_asp_density, lw = 1, col = "blue")
lines(norm_abs_voic_nasp_density, lw = 1, col = "red")
```
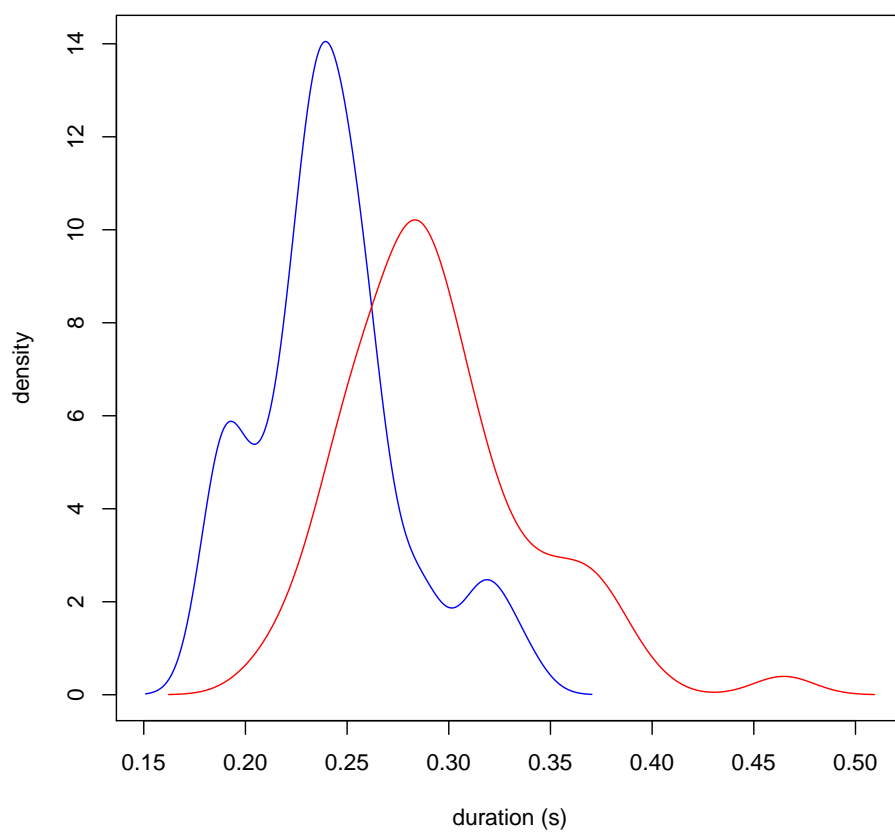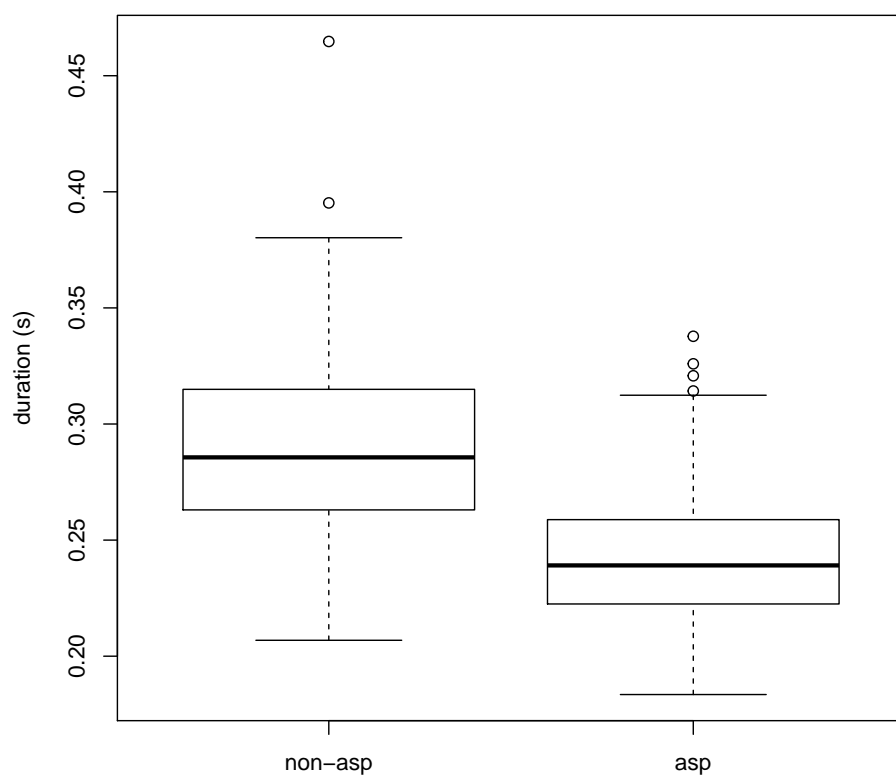
```r
boxplot(norm_abs_voic ~ asp,
        data = bi_nas,
        names = c("non-asp", "asp"),
        ylab = "duration (s)",
    main = "Duration (in seconds) of modal voicing in CVNCV words"
        )
```

Duration (in seconds) of modal voicing in CVNCV words


```
shapiro.test(bi_nas_asp)

##
##  Shapiro-Wilk normality test
##
## data:  bi_nas_asp
## W = 0.94884, p-value = 0.01743

shapiro.test(bi_nas_nasp)

##
##  Shapiro-Wilk normality test
##
## data:  bi_nas_nasp
## W = 0.94685, p-value = 0.005831

wilcox.test(norm_abs_voic ~ asp, data = bi_nas)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  norm_abs_voic by asp
## W = 3220, p-value = 2.121e-10
## alternative hypothesis: true location shift is not equal to 0
```

2.3   Laterals

```r
bi_lat <- subset(bi, manner == "lateral")
```

```r
bi_lat_asp <- bi_lat$norm_abs_voic[bi_lat$asp == "yes"]
bi_lat_nasp <- bi_lat$norm_abs_voic[bi_lat$asp == "no"]

norm_abs_voic_asp_density <- density(bi_lat_asp)
norm_abs_voic_nasp_density <- density(bi_lat_nasp)

norm_abs_voic_density_x <- c(norm_abs_voic_asp_density$x,
                             norm_abs_voic_nasp_density$x)
norm_abs_voic_density_y <- c(norm_abs_voic_asp_density$y,
                             norm_abs_voic_nasp_density$y)

x.limits <- range(norm_abs_voic_density_x)
y.limits <- range(norm_abs_voic_density_y)

plot(
c(), c(),
xlim = x.limits,
ylim = y.limits,
xlab = "duration (s)", ylab = "density"
)

lines(norm_abs_voic_asp_density, lw = 1, col = "blue")
lines(norm_abs_voic_nasp_density, lw = 1, col = "red")
```
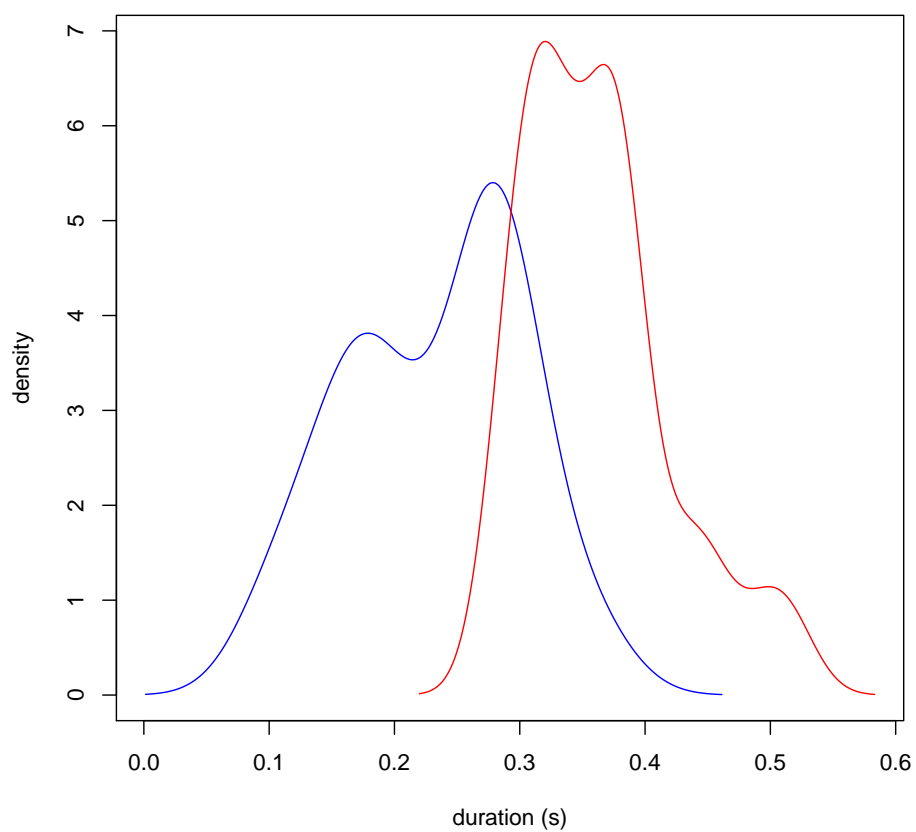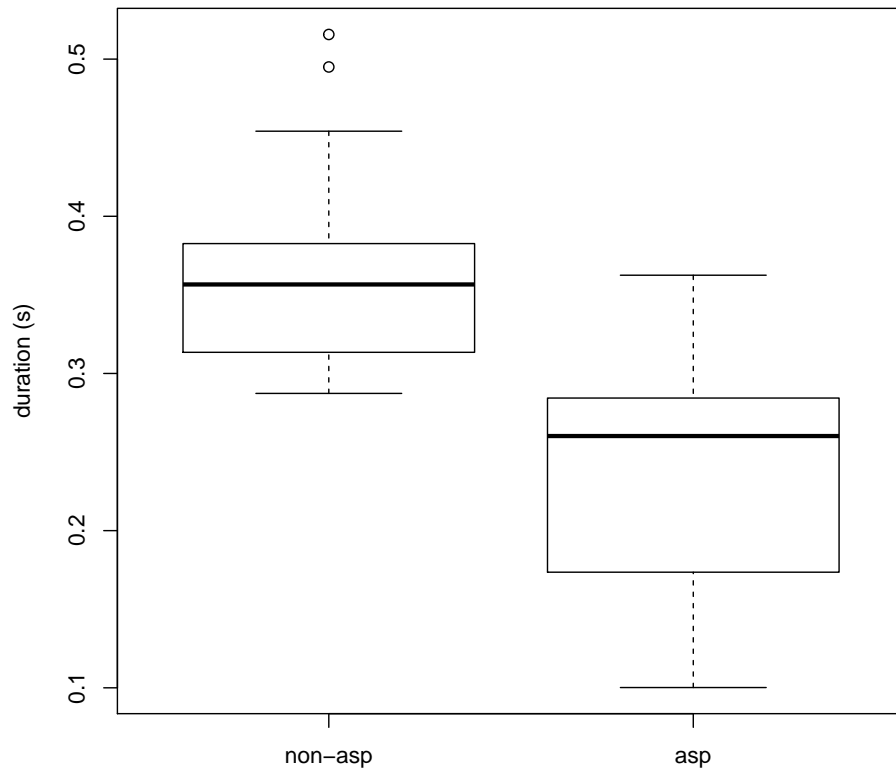
```
boxplot(norm_abs_voic ~ asp,
        data = bi_lat,
        names = c("non-asp", "asp"),
        ylab = "duration (s)",
    main = "Duration (in seconds) of modal voicing in CVLCV words"
        )
```

Duration (in seconds) of modal voicing in CVLCV words


```
shapiro.test(bi_lat_asp)

##
##  Shapiro-Wilk normality test
##
## data:  bi_lat_asp
## W = 0.95378, p-value = 0.2644

shapiro.test(bi_lat_nasp)

##
##  Shapiro-Wilk normality test
##
## data:  bi_lat_nasp
## W = 0.91062, p-value = 0.01543

wilcox.test(norm_abs_voic ~ asp, data = bi_lat)
```

```
##
##  Wilcoxon rank sum test
##
## data:  norm_abs_voic by asp
## W = 771, p-value = 2.525e-11
## alternative hypothesis: true location shift is not equal to 0
```

2.4 Rhotics

```r
bi_rho <- subset(bi, manner == "rhotic")
```

```r
bi_rho_asp <- bi_rho$norm_abs_voic[bi_rho$asp == "yes"]
bi_rho_nasp <- bi_rho$norm_abs_voic[bi_rho$asp == "no"]

norm_abs_voic_asp_density <- density(bi_rho_asp)
norm_abs_voic_nasp_density <- density(bi_rho_nasp)

norm_abs_voic_density_x <- c(norm_abs_voic_asp_density$x,
                             norm_abs_voic_nasp_density$x)
norm_abs_voic_density_y <- c(norm_abs_voic_asp_density$y,
                             norm_abs_voic_nasp_density$y)

x.limits <- range(norm_abs_voic_density_x)
y.limits <- range(norm_abs_voic_density_y)

plot(
c(), c(),
xlim = x.limits,
ylim = y.limits,
xlab = "duration (s)", ylab = "density"
)

lines(norm_abs_voic_asp_density, lw = 1, col = "blue")
lines(norm_abs_voic_nasp_density, lw = 1, col = "red")
```
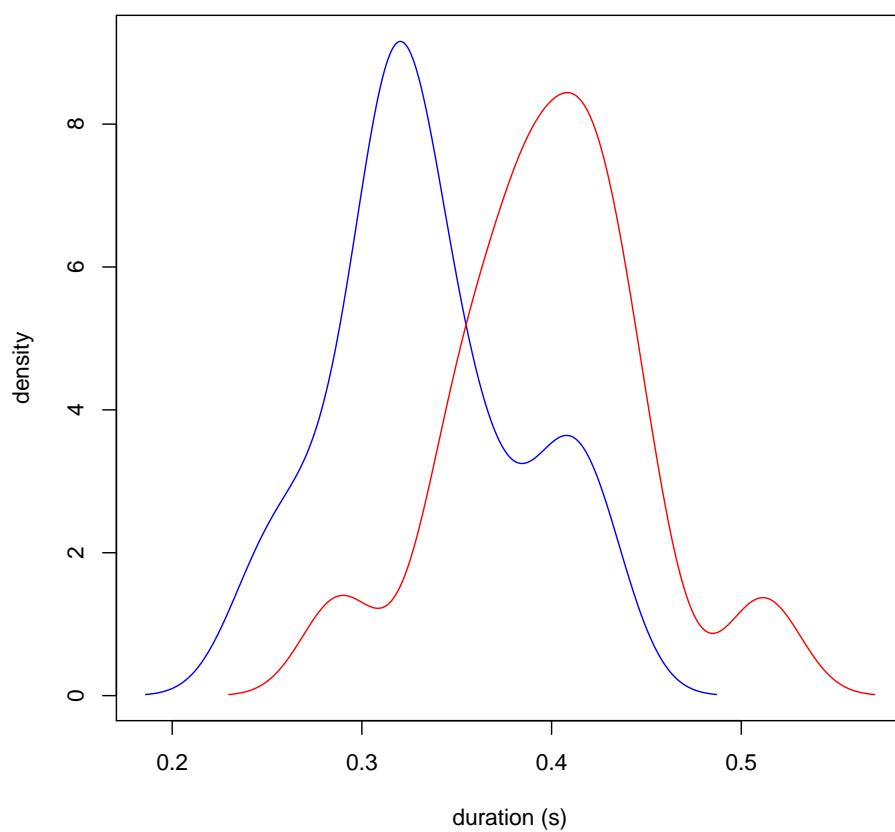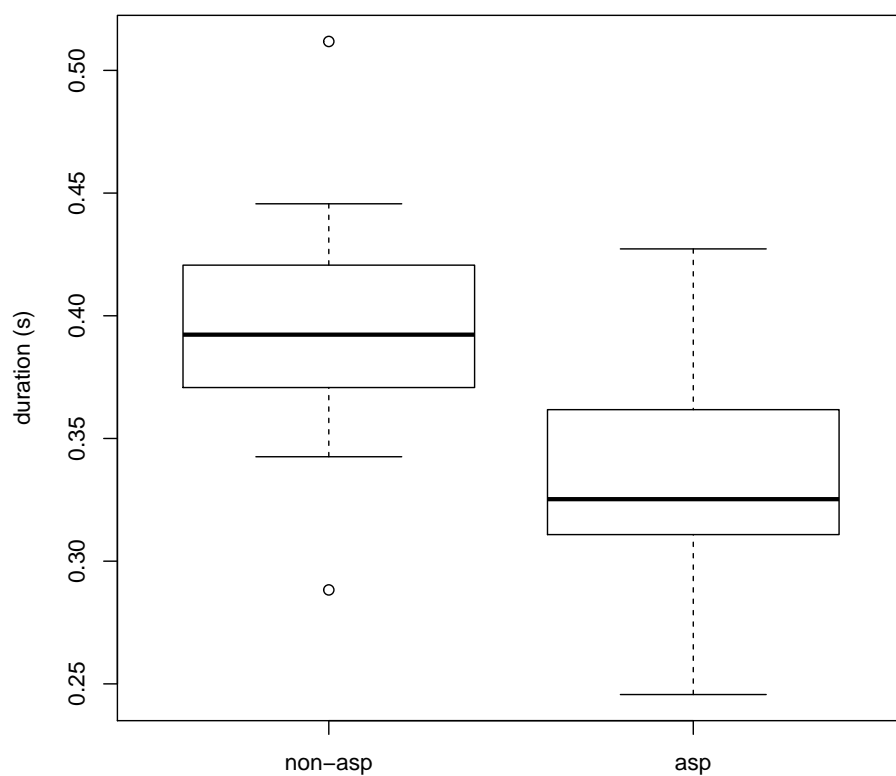
```
boxplot(norm_abs_voic ~ asp,
        data = bi_rho,
        names = c("non-asp", "asp"),
        ylab = "duration (s)",
    main = "Duration (in seconds) of modal voicing in CVRCV words"
        )
```

Duration (in seconds) of modal voicing in CVRCV words


```
shapiro.test(bi_rho_asp)

##
##  Shapiro-Wilk normality test
##
## data:  bi_rho_asp
## W = 0.96487, p-value = 0.7762

shapiro.test(bi_rho_nasp)

##
##  Shapiro-Wilk normality test
##
## data:  bi_rho_nasp
## W = 0.97208, p-value = 0.8876

t.test(norm_abs_voic ~ asp, data = bi_rho)
```

27

```
## 
##   Welch Two Sample t-test
## 
## data:  norm_abs_voic by asp
## t = 3.4013, df = 27.988, p-value = 0.002036
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   0.02544551 0.10250100
## sample estimates:
##   mean in group no mean in group yes
##          0.3982628          0.3342896
```

As with monosyllabic words, we should check if the duration of bisyl-
labic words is affected by the presence vs. absence of preaspiration.

```
boxplot(dur_word ~ asp, data = bi)
```
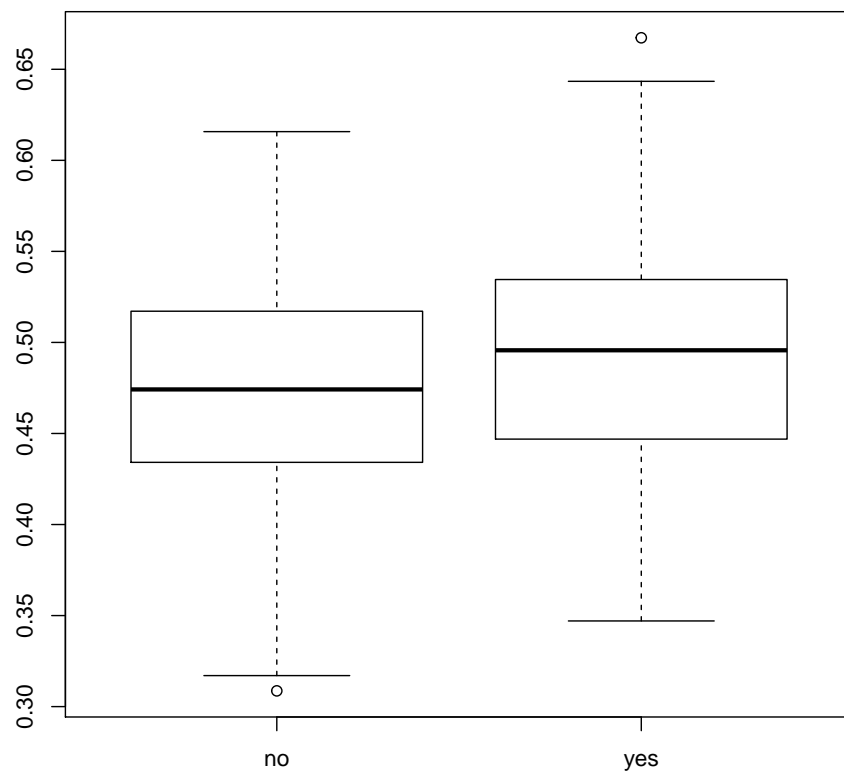
```
shapiro.test(bi$dur_word[bi$asp == "yes"])

##
##  Shapiro-Wilk normality test
##
## data:  bi$dur_word[bi$asp == "yes"]
## W = 0.99244, p-value = 0.4518

shapiro.test(bi$dur_word[bi$asp == "no"])

##
##  Shapiro-Wilk normality test
##
## data:  bi$dur_word[bi$asp == "no"]
## W = 0.98318, p-value = 0.008744
```
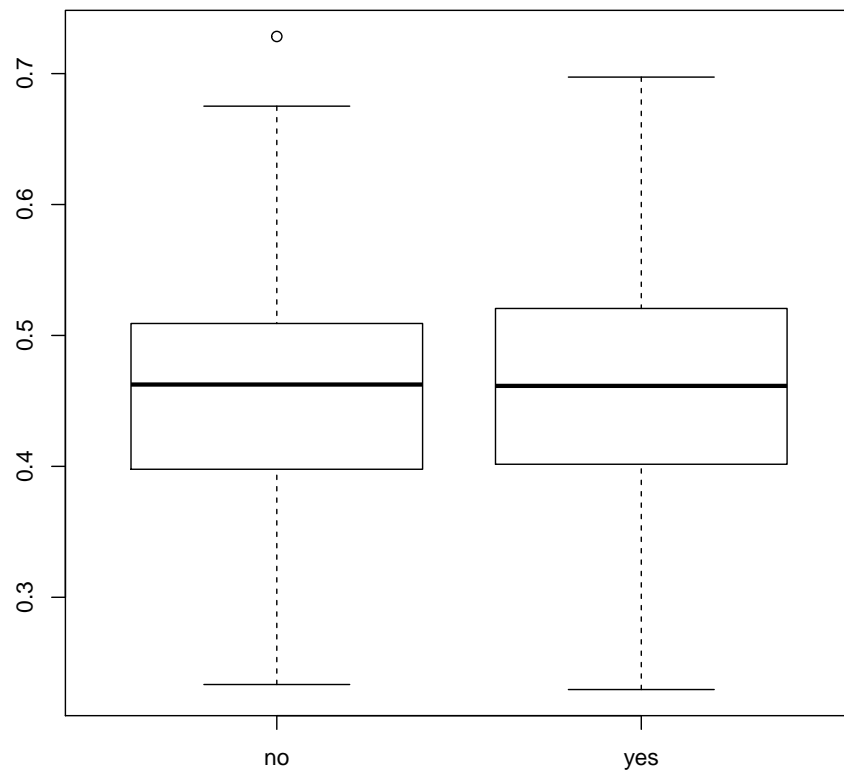
```
wilcox.test(dur_word ~ asp, data = bi)

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  dur_word by asp
## W = 17288, p-value = 0.00254
## alternative hypothesis: true location shift is not equal to 0
```

According to a Wilcoxon-test, the duration of bisyllabic words is affected by the presence vs. absence of preaspiration.

```
boxplot(dur_word ~ asp, data = results)
```

```
shapiro.test(results$dur_word[results$asp == "yes"])

##
##  Shapiro-Wilk normality test
##
## data:  results$dur_word[results$asp == "yes"]
## W = 0.99577, p-value = 0.3404

shapiro.test(results$dur_word[results$asp == "no"])

##
##  Shapiro-Wilk normality test
##
## data:  results$dur_word[results$asp == "no"]
## W = 0.99183, p-value = 0.02449
```

```r
wilcox.test(dur_word ~ asp, data = results)

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  dur_word by asp
## W = 79534, p-value = 0.2721
## alternative hypothesis: true location shift is not equal to 0
```