# Preliminary analysis

January 29, 2017

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

```
library(readr)
results_raw <- read_csv("./results/results.csv", na = "--undefined--")

## Parsed with column specification:
## cols(
## .default = col_double(),
## speaker = col_character(),
## idx = col_integer(),
## word = col_character()
## )

## See spec(...) for full column specifications.

words <- read.csv("./task/task-src/words.csv")</pre>
```

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 == "mono")
bi <- subset(results, syl == "di")</pre>
```

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

The mean length of monosyllabic words is 434.01 msec [sd = 88.01], while that of bisyllabic words is 478.63 msec [sd = 63.21].

```
plot(
    c(), c(),
    xlim = c(100,900),
    ylim = c(0,150),
    xlab = "duration (msec)", ylab = "frequency"
)

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

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

legend("topright",
    c("mono", "di"),
    fill = c("darkgrey","lightgrey")
    )
```

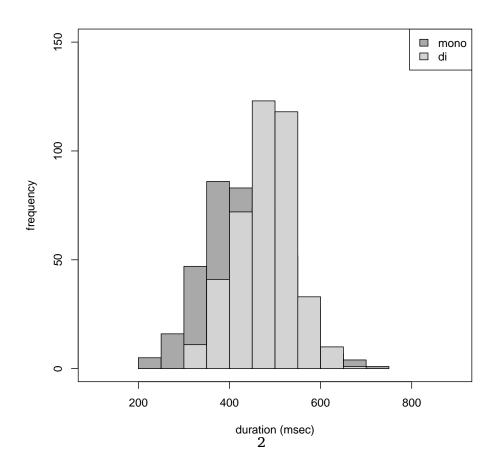
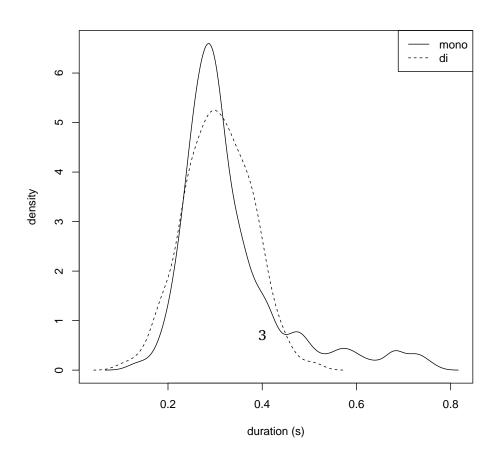


Figure 1: Duration of mono- vs. bisyllabic words.

```
norm_vowel_mono_density <- density(mono$norm_vowel, na.rm = TRUE)</pre>
norm_vowel_di_density <- density(bi$norm_vowel, na.rm = TRUE)</pre>
norm_vowel_density_x <- c(norm_vowel_mono_density$x,</pre>
                                    norm_vowel_di_density$x)
norm_vowel_density_y <- c(norm_vowel_mono_density$y,</pre>
                                    norm_vowel_di_density$y)
x.limits <- range(norm_vowel_density_x)</pre>
y.limits <- range(norm_vowel_density_y)</pre>
plot(
c(), c(),
xlim = x.limits,
ylim = y.limits,
xlab = "duration (s)", ylab = "density"
lines(norm_vowel_mono_density, lw = 1)
lines(norm_vowel_di_density, lw = 1, lty = 2)
legend("topright",
       c("mono", "di"),
       lty = c(1,2)
```



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.

```
shapiro.test(mono$norm_vowel)

##

## Shapiro-Wilk normality test

##

## data: mono$norm_vowel

## W = 0.82764, p-value < 2.2e-16

shapiro.test(bi$norm_vowel)

##

## Shapiro-Wilk normality test

##

## data: bi$norm_vowel

##

## data: bi$norm_vowel

##

## data: bi$norm_vowel</pre>
```

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

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

##

## Wilcoxon rank sum test with continuity correction

##

## data: mono$norm_vowel and bi$norm_vowel

## W = 82451, p-value = 0.309

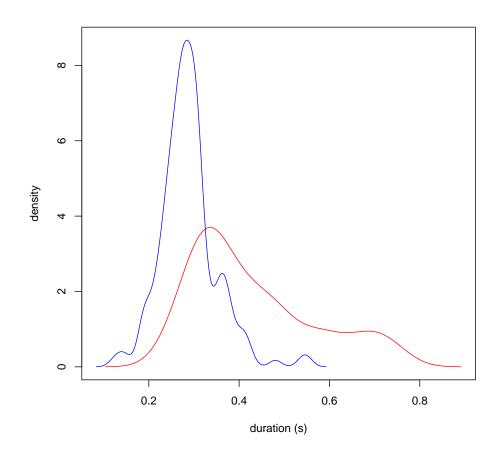
## alternative hypothesis: true location shift is not equal to 0
```

## 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")</pre>
di_stop <- subset(bi, manner == "stop")</pre>
mono_stop_asp <- mono_stop$norm_vowel[mono_stop$asp == "yes"]</pre>
mono_stop_nasp <- mono_stop$norm_vowel[mono_stop$asp == "no"]</pre>
norm_vowel_asp_density <- density(mono_stop_asp)</pre>
norm_vowel_nasp_density <- density(mono_stop_nasp)</pre>
norm_vowel_density_x <- c(norm_vowel_asp_density$x,</pre>
                                     norm_vowel_nasp_density$x)
norm_vowel_density_y <- c(norm_vowel_asp_density$y,</pre>
                                     norm_vowel_nasp_density$y)
x.limits <- range(norm_vowel_density_x)</pre>
y.limits <- range(norm_vowel_density_y)</pre>
plot(
c(), c(),
xlim = x.limits,
ylim = y.limits,
xlab = "duration (s)", ylab = "density"
)
lines(norm_vowel_asp_density, lw = 1, col = "blue")
lines(norm_vowel_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

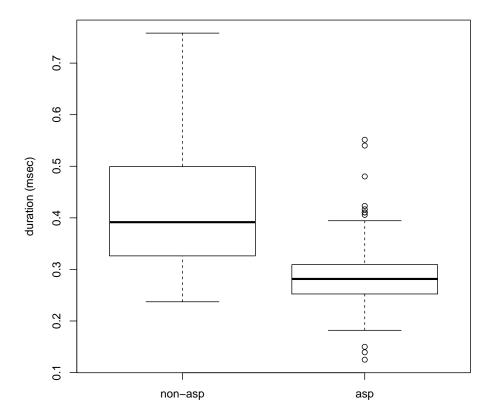
## data: mono_stop_nasp

## 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).

### Duration (in seconds) of modal voicing in CVCC words



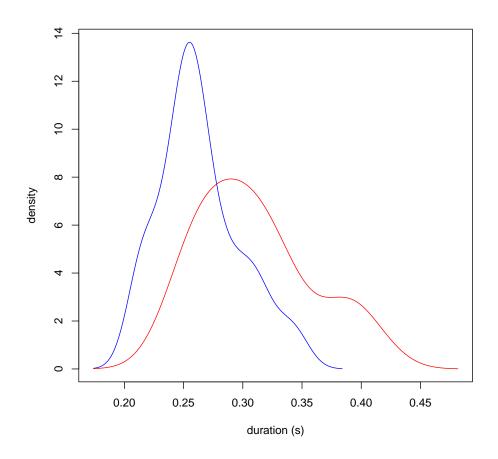
```
wilcox.test(norm_vowel ~ asp, data = mono_stop)
##
```

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

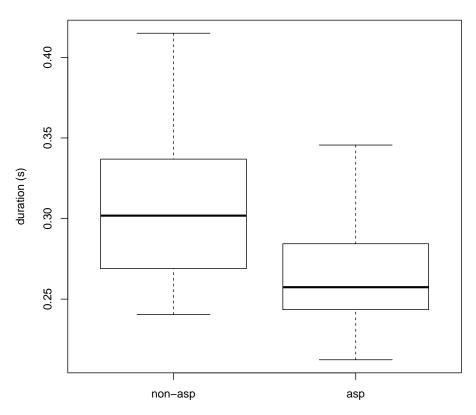
#### 1.2 Nasals

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

```
mono_nas_asp <- mono_nas$norm_vowel[mono_nas$asp == "yes"]</pre>
mono_nas_nasp <- mono_nas$norm_vowel[mono_nas$asp == "no"]</pre>
norm_vowel_asp_density <- density(mono_nas_asp)</pre>
norm_vowel_nasp_density <- density(mono_nas_nasp)</pre>
norm_vowel_density_x <- c(norm_vowel_asp_density$x,</pre>
                                     norm_vowel_nasp_density$x)
norm_vowel_density_y <- c(norm_vowel_asp_density$y,</pre>
                                     norm_vowel_nasp_density$y)
x.limits <- range(norm_vowel_density_x)</pre>
y.limits <- range(norm_vowel_density_y)</pre>
plot(
c(), c(),
xlim = x.limits,
ylim = y.limits,
xlab = "duration (s)", ylab = "density"
lines(norm_vowel_asp_density, lw = 1, col = "blue")
lines(norm_vowel_nasp_density, lw = 1, col = "red")
```



## Duration (in seconds) of modal voicing in CVNC words



```
shapiro.test(mono_nas_asp)

##

## Shapiro-Wilk normality test

##

## data: mono_nas_asp

## W = 0.94983, p-value = 0.0879

shapiro.test(mono_nas_nasp)

##

## Shapiro-Wilk normality test

##

## data: mono_nas_nasp

## w = 0.94149, p-value = 0.1097

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

```
##
## Welch Two Sample t-test
##
## data: norm_vowel by asp
## t = 4.3754, df = 48.072, p-value = 6.497e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 0.02495378 0.06738528
## sample estimates:
## mean in group no mean in group yes
## 0.3098634 0.2636939
```

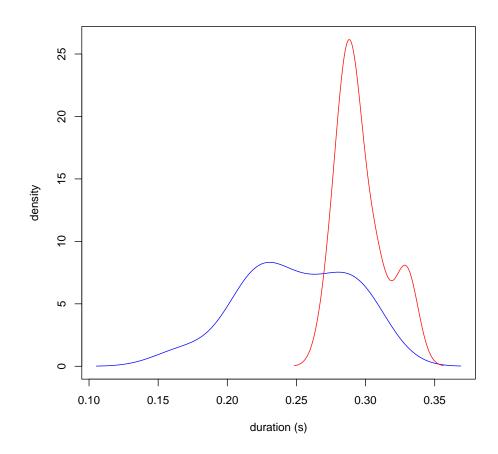
According to a two-sample t-test, the duration of modal voicing in words ending with an  $\N$ C cluster (0.26 seconds) is shorter than in words ending with an NC cluster (0.31 seconds).

#### 1.3 Laterals

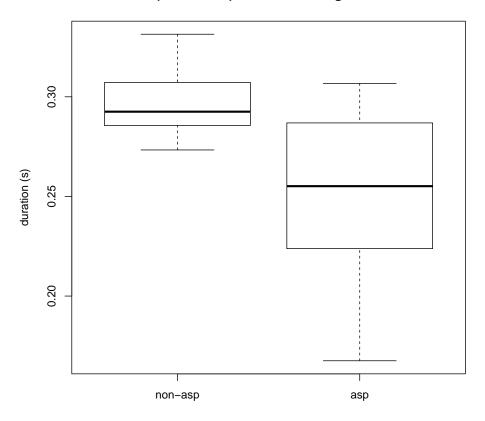
```
mono_lat <- subset(mono, manner == "lateral")
```

```
mono_lat_asp <- mono_lat$norm_vowel[mono_lat$asp == "yes"]</pre>
mono_lat_nasp <- mono_lat$norm_vowel[mono_lat$asp == "no"]</pre>
norm_vowel_asp_density <- density(mono_lat_asp)</pre>
norm_vowel_nasp_density <- density(mono_lat_nasp)</pre>
norm_vowel_density_x <- c(norm_vowel_asp_density$x,</pre>
                                     norm_vowel_nasp_density$x)
norm_vowel_density_y <- c(norm_vowel_asp_density$y,</pre>
                                     norm_vowel_nasp_density$y)
x.limits <- range(norm_vowel_density_x)</pre>
y.limits <- range(norm_vowel_density_y)</pre>
plot(
c(), c(),
xlim = x.limits,
ylim = y.limits,
xlab = "duration (s)", ylab = "density"
lines(norm_vowel_asp_density, lw = 1, col = "blue")
```

```
lines(norm_vowel_nasp_density, lw = 1, col = "red")
```



## 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

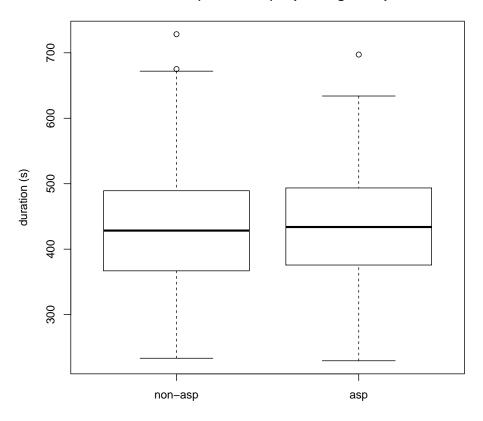
## w = 0.89635, p-value = 0.1423

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

```
##
## Welch Two Sample t-test
##
## data: norm_vowel 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.

### 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 = 0.9912, p-value = 0.2007

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 = -0.13529, df = 351.41, p-value = 0.8925

## alternative hypothesis: true difference in means is not equal to 0

## 95 percent confidence interval:

## -18.88264 16.45203

## sample estimates:

## mean in group no mean in group yes

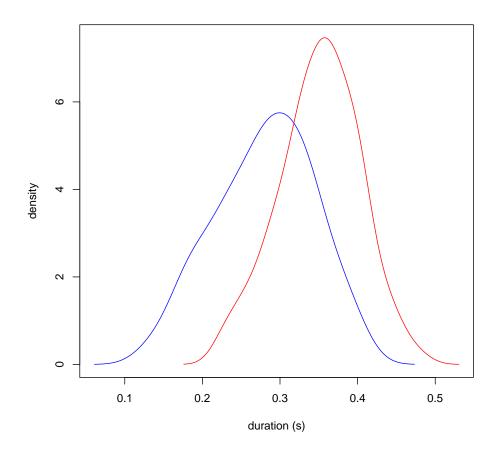
## 433.3422 434.5575
```

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

```
lines(norm_vowel_asp_density, lw = 1, col = "blue")
lines(norm_vowel_nasp_density, lw = 1, col = "red")
```



```
shapiro.test(di_stop_asp)

##

## Shapiro-Wilk normality test

##

## data: di_stop_asp

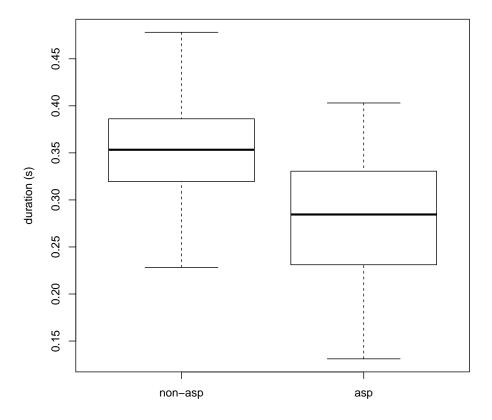
## W = 0.98478, p-value = 0.41

shapiro.test(di_stop_nasp)

##
```

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

### Duration (in seconds) of modal voicing in CVCCV words



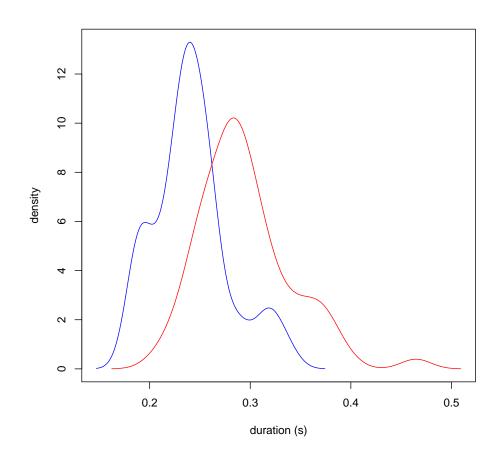
```
wilcox.test(norm_vowel ~ asp, data = mono_stop)
##
```

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

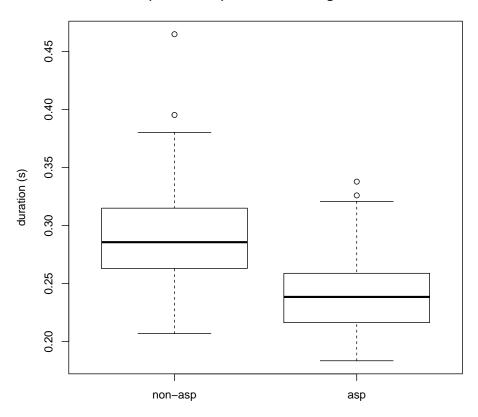
#### 2.2 Nasals

```
di_nas <- subset(bi, manner == "nasal")</pre>
```

```
di_nas_asp <- di_nas$norm_vowel[di_nas$asp == "yes"]</pre>
di_nas_nasp <- di_nas$norm_vowel[di_nas$asp == "no"]</pre>
norm_vowel_asp_density <- density(di_nas_asp)</pre>
norm_vowel_nasp_density <- density(di_nas_nasp)</pre>
norm_vowel_density_x <- c(norm_vowel_asp_density$x,</pre>
                                     norm_vowel_nasp_density$x)
norm_vowel_density_y <- c(norm_vowel_asp_density$y,</pre>
                                     norm_vowel_nasp_density$y)
x.limits <- range(norm_vowel_density_x)</pre>
y.limits <- range(norm_vowel_density_y)</pre>
plot(
c(), c(),
xlim = x.limits,
ylim = y.limits,
xlab = "duration (s)", ylab = "density"
lines(norm_vowel_asp_density, lw = 1, col = "blue")
lines(norm_vowel_nasp_density, lw = 1, col = "red")
```



## Duration (in seconds) of modal voicing in CVNCV words



```
shapiro.test(di_nas_asp)

##

## Shapiro-Wilk normality test

##

## data: di_nas_asp

## W = 0.9445, p-value = 0.01442

shapiro.test(di_nas_nasp)

##

## Shapiro-Wilk normality test

##

## data: di_nas_nasp

## W = 0.94685, p-value = 0.005831

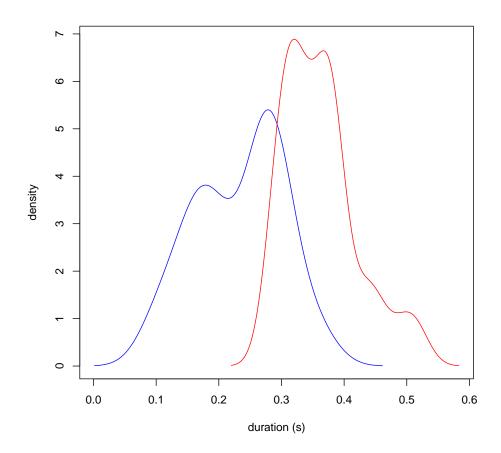
wilcox.test(norm_vowel ~ asp, data = di_nas)
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: norm_vowel by asp
## W = 3052, p-value = 3.721e-10
## alternative hypothesis: true location shift is not equal to 0
```

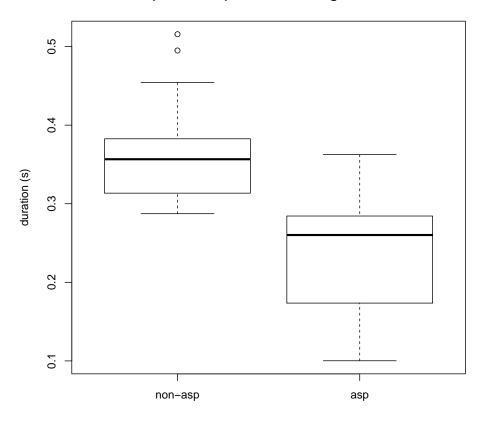
#### 2.3 Laterals

```
di_lat <- subset(bi, manner == "lateral")</pre>
```

```
di_lat_asp <- di_lat$norm_vowel[di_lat$asp == "yes"]</pre>
di_lat_nasp <- di_lat$norm_vowel[di_lat$asp == "no"]</pre>
norm_vowel_asp_density <- density(di_lat_asp)</pre>
norm_vowel_nasp_density <- density(di_lat_nasp)</pre>
norm_vowel_density_x <- c(norm_vowel_asp_density$x,</pre>
                                     norm_vowel_nasp_density$x)
norm_vowel_density_y <- c(norm_vowel_asp_density$y,</pre>
                                     norm_vowel_nasp_density$y)
x.limits <- range(norm_vowel_density_x)</pre>
y.limits <- range(norm_vowel_density_y)</pre>
plot(
c(), c(),
xlim = x.limits,
ylim = y.limits,
xlab = "duration (s)", ylab = "density"
)
lines(norm_vowel_asp_density, lw = 1, col = "blue")
lines(norm_vowel_nasp_density, lw = 1, col = "red")
```



## Duration (in seconds) of modal voicing in CVLCV words



```
shapiro.test(di_lat_asp)

##

## Shapiro-Wilk normality test

##

## data: di_lat_asp

## W = 0.95378, p-value = 0.2644

shapiro.test(di_lat_nasp)

##

## Shapiro-Wilk normality test

##

## data: di_lat_nasp

## W = 0.91062, p-value = 0.01543

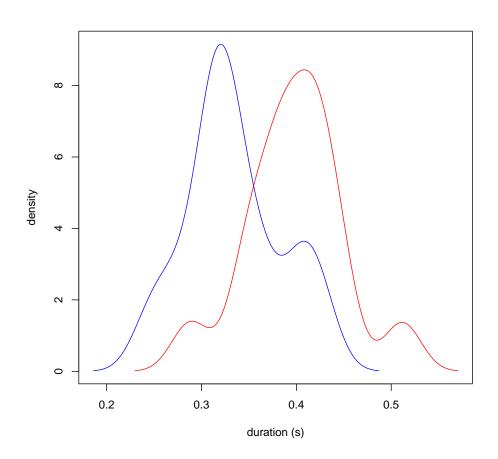
wilcox.test(norm_vowel ~ asp, data = di_lat)
```

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

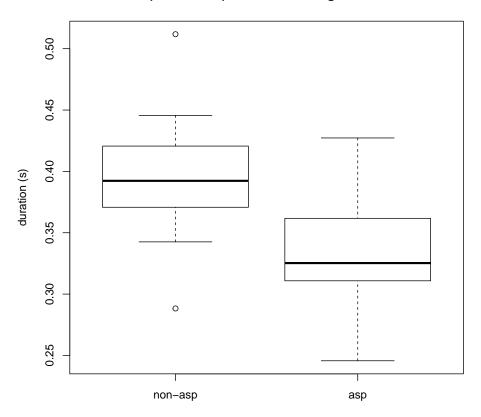
#### 2.4 Rhotics

```
di_rho <- subset(bi, manner == "rhotic")</pre>
```

```
di_rho_asp <- di_rho$norm_vowel[di_rho$asp == "yes"]</pre>
di_rho_nasp <- di_rho$norm_vowel[di_rho$asp == "no"]</pre>
norm_vowel_asp_density <- density(di_rho_asp)</pre>
norm_vowel_nasp_density <- density(di_rho_nasp)</pre>
norm_vowel_density_x <- c(norm_vowel_asp_density$x,</pre>
                                     norm_vowel_nasp_density$x)
norm_vowel_density_y <- c(norm_vowel_asp_density$y,</pre>
                                     norm_vowel_nasp_density$y)
x.limits <- range(norm_vowel_density_x)</pre>
y.limits <- range(norm_vowel_density_y)</pre>
plot(
c(), c(),
xlim = x.limits,
ylim = y.limits,
xlab = "duration (s)", ylab = "density"
)
lines(norm_vowel_asp_density, lw = 1, col = "blue")
lines(norm_vowel_nasp_density, lw = 1, col = "red")
```



## Duration (in seconds) of modal voicing in CVRCV words



```
shapiro.test(di_rho_asp)

##

## Shapiro-Wilk normality test

##

## data: di_rho_asp

## W = 0.96487, p-value = 0.7762

shapiro.test(di_rho_nasp)

##

## Shapiro-Wilk normality test

##

## data: di_rho_nasp

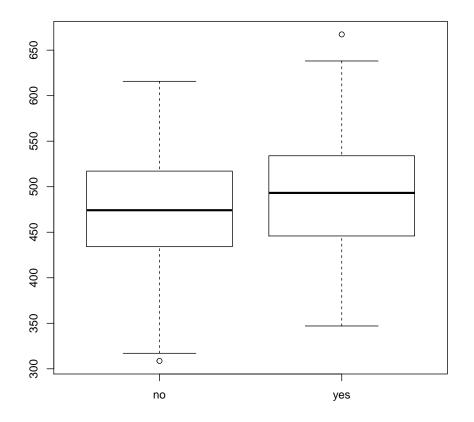
## W = 0.97208, p-value = 0.8876

t.test(norm_vowel ~ asp, data = di_rho)
```

```
##
## Welch Two Sample t-test
##
## data: norm_vowel 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 bisyllabic 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 = 0.9923, p-value = 0.4437

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

##

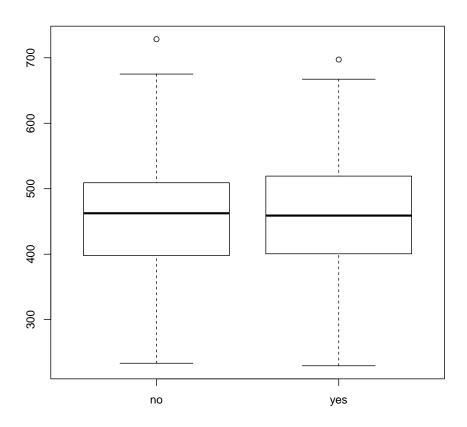
## data: dur_word by asp

## W = 17319, p-value = 0.004711

## 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 = 0.99506, p-value = 0.2236

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
```

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

##

## Wilcoxon rank sum test with continuity correction

##

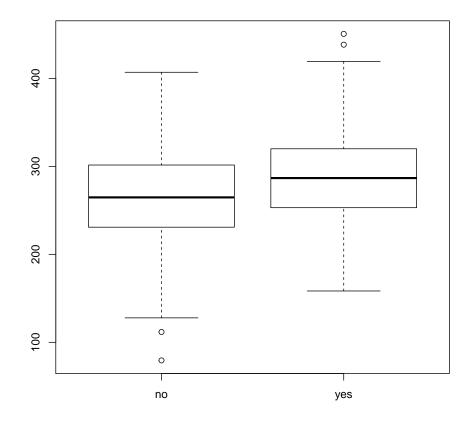
## data: dur_word by asp

## W = 79555, p-value = 0.3918

## alternative hypothesis: true location shift is not equal to 0
```

# 3 VOR

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



```
shapiro.test(results$vor[results$asp == "yes"])
##
##
   Shapiro-Wilk normality test
## data: results$vor[results$asp == "yes"]
## W = 0.9958, p-value = 0.3577
shapiro.test(results$vor[results$asp == "no"])
##
   Shapiro-Wilk normality test
## data: results$vor[results$asp == "no"]
## W = 0.99719, p-value = 0.7422
t.test(vor ~ asp, data = results)
##
## Welch Two Sample t-test
##
## data: vor by asp
## t = -6.1393, df = 790.14, p-value = 1.311e-09
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -28.86452 -14.87835
## sample estimates:
## mean in group no mean in group yes
          265.3115
```

### 4 Closure

```
wilcox.test(norm_clos ~ asp, mono_nas)

##

## Wilcoxon rank sum test

##

## data: norm_clos by asp

## W = 643, p-value = 0.2487

## alternative hypothesis: true location shift is not equal to 0

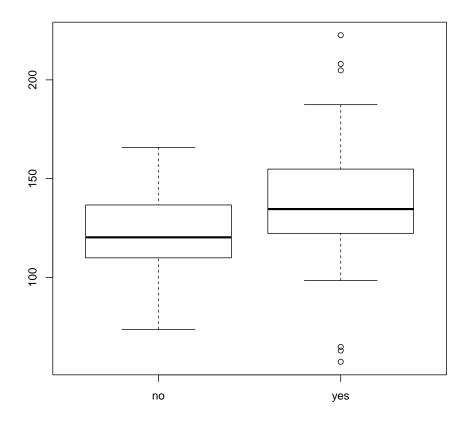
wilcox.test(norm_clos ~ asp, mono_lat)
```

```
## Wilcoxon rank sum test
##
## data: norm_clos by asp
## W = 122, p-value = 0.1257
## alternative hypothesis: true location shift is not equal to 0
wilcox.test(norm_clos ~ asp, di_nas)
##
## Wilcoxon rank sum test with continuity correction
## data: norm_clos by asp
## W = 1829, p-value = 0.9733
## alternative hypothesis: true location shift is not equal to 0
wilcox.test(norm_clos ~ asp, di_lat)
##
## Wilcoxon rank sum test
##
## data: norm_clos by asp
## W = 546, p-value = 0.02392
## alternative hypothesis: true location shift is not equal to 0
wilcox.test(norm_clos ~ asp, di_rho)
##
## Wilcoxon rank sum test
##
## data: norm_clos by asp
## W = 206, p-value = 2.655e-05
\#\# alternative hypothesis: true location shift is not equal to 0
```

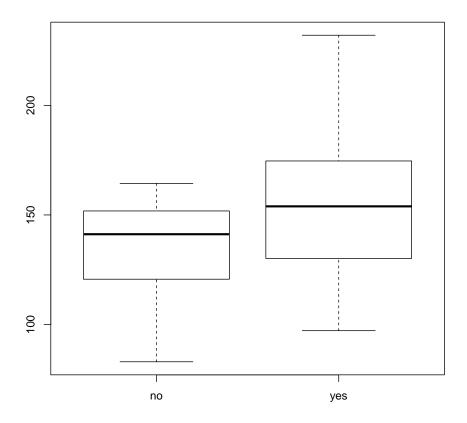
### 5 Sonorant duration

```
wilcox.test(norm_mann ~ asp, mono_nas)
##
## Wilcoxon rank sum test
##
## data: norm_mann by asp
```

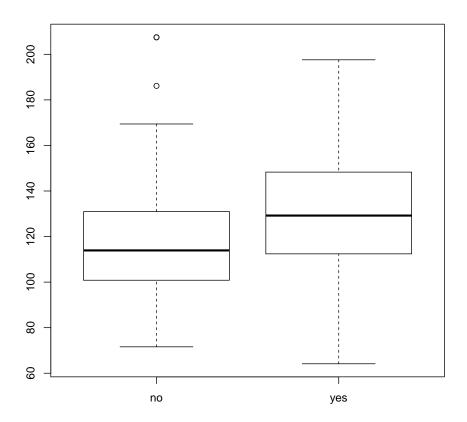
```
## W = 338, p-value = 0.006585
## alternative hypothesis: true location shift is not equal to 0
wilcox.test(norm_mann ~ asp, mono_lat)
##
## Wilcoxon rank sum test
##
## data: norm_mann by asp
## W = 25, p-value = 0.0009344
## alternative hypothesis: true location shift is not equal to 0
wilcox.test(norm_mann ~ asp, di_nas)
##
## Wilcoxon rank sum test with continuity correction
## data: norm_mann by asp
## W = 1190, p-value = 0.0008771
## alternative hypothesis: true location shift is not equal to 0
wilcox.test(norm_mann ~ asp, di_lat)
##
## Wilcoxon rank sum test
##
## data: norm_mann by asp
## W = 58, p-value = 6.607e-10
## alternative hypothesis: true location shift is not equal to 0
wilcox.test(norm_mann ~ asp, di_rho)
## Wilcoxon rank sum test
##
## data: norm_mann by asp
## W = 6, p-value = 3.868e-07
\#\# alternative hypothesis: true location shift is not equal to 0
boxplot(dur_mann ~ asp, mono_nas)
```



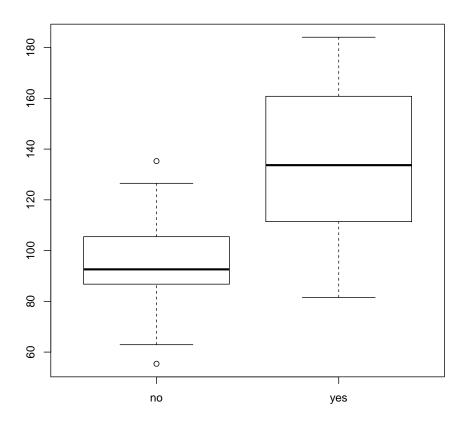
boxplot(dur\_mann ~ asp, mono\_lat)



boxplot(dur\_mann ~ asp, di\_nas)



boxplot(dur\_mann ~ asp, di\_lat)



boxplot(dur\_mann ~ asp, di\_rho)

