Vowel duration

Stefano Coretta

February 27, 2017

1 Import data

Let's read the files.

```
vowels <- list.files(path = "pilot/results",
                     pattern = "*-vowel-durations.csv",
                     full.names = TRUE) %>%
    map_df(~read_csv(.))
```

```
## Parsed with column specification:
## cols(
##   index = col_integer(),
##   speaker = col_character(),
##   word = col_character(),
##   duration = col_double()
## )
## Parsed with column specification:
## cols(
##   index = col_integer(),
##   speaker = col_character(),
##   word = col_character(),
##   duration = col_double()
## )
## Parsed with column specification:
## cols(
##   index = col_integer(),
##   speaker = col_character(),
##   word = col_character(),
##   duration = col_double()
## )
```

```
languages <- read_csv("./pilot/stimuli/languages.csv")
```

```
## Parsed with column specification:
## cols(
##   speaker = col_character(),
##   language = col_character()
## )
```

```
words <- read_csv("./pilot/stimuli/nonce.csv")
```

```
## Parsed with column specification:
## cols(
##   item = col_integer(),
##   word = col_character(),
##   ipa = col_character(),
##   c1 = col_character(),
##   c1phonation = col_character(),
```

```
##   vowel = col_character(),
##   anteropost = col_character(),
##   height = col_character(),
##   c2 = col_character(),
##   c2phonation = col_character(),
##   c2place = col_character(),
##   language = col_character()
## )
```

```
vowels <- left_join(vowels, languages) %>%
    left_join(y = words) %>%
    mutate_if(is.character, as.factor) %>%
    mutate(duration.norm = (duration - mean(duration)) / sd(duration))
```

```
## Joining, by = "speaker"
```

```
## Joining, by = c("word", "language")
```

2 Italian

Fit a mixed effect linear model for Italian.

```
vowels.it <- filter(vowels, language == "italian")

model.it <- lmer(duration.norm ~ c2phonation * vowel + c1phonation + c2place +
                   (1 | word) + (1|speaker),
                 data = vowels.it
                 )

summary(model.it)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: duration.norm ~ c2phonation * vowel + c1phonation + c2place +
##     (1 | word) + (1 | speaker)
##    Data: vowels.it
##
## REML criterion at convergence: 264.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4316 -0.5556  0.0513  0.5913  3.0894
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  word     (Intercept) 0.02202  0.1484
##  speaker  (Intercept) 0.41338  0.6429
##  Residual             0.13394  0.3660
## Number of obs: 263, groups:  word, 24; speaker, 2
##
## Fixed effects:
##                             Estimate Std. Error t value
## (Intercept)                  1.26767    0.46701   2.714
## c2phonationvoiceless        -0.58864    0.13105  -4.492
## vowelo                      -0.16970    0.13077  -1.298
```

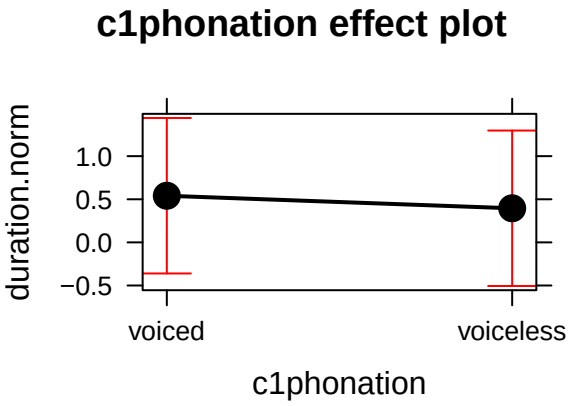
```
## vowelu                       -1.08921    0.13077  -8.329
## c1phonationvoiceless         -0.14535    0.07555  -1.924
## c2placevelar                 -0.16564    0.07555  -2.192
## c2phonationvoiceless:vowelo   0.02321    0.18513   0.125
## c2phonationvoiceless:vowelu   0.39704    0.18513   2.145
##
## Correlation of Fixed Effects:
##                              (Intr) c2phnt vowelo vowelu c1phnt c2plcv
## c2phntnvcls                  -0.140
## vowelo                       -0.140  0.499
## vowelu                       -0.140  0.499  0.500
## c1phntnvcls                  -0.081 -0.003  0.000  0.000
## c2placevelr                  -0.081  0.003  0.000  0.000 -0.001
## c2phonationvoiceless:vowelo   0.099 -0.708 -0.706 -0.353  0.002 -0.002
## c2phonationvoiceless:vowelu   0.099 -0.708 -0.353 -0.706  0.002 -0.002
##                              c2phonationvoiceless:vowelo
## c2phntnvcls
## vowelo
## vowelu
## c1phntnvcls
## c2placevelr
## c2phonationvoiceless:vowelo
## c2phonationvoiceless:vowelu  0.501
```

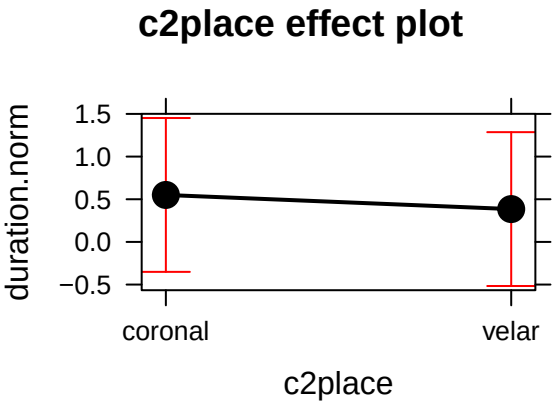
```
mixed(duration.norm ~ c2phonation * vowel + c1phonation + c2place +
                 (1 | word) + (1|speaker),
             data = vowels.it)
```

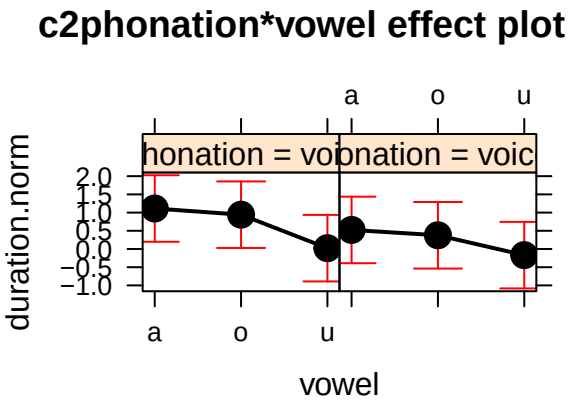
```
## Fitting 6 (g)lmer() models:
## [......]
## Obtaining 5 p-values:
## [.....]
```

```
##              Effect       df F.scaling         F p.value
## 1       c2phonation 1, 16.09      1.00 20.17 ***   .0004
## 2             vowel 2, 15.95      1.00 40.17 ***  <.0001
## 3       c1phonation 1, 16.00      1.00    3.70 +     .07
## 4           c2place 1, 16.00      1.00    4.81 *     .04
## 5 c2phonation:vowel 2, 16.00      1.00    2.90 +     .08
```

```
plot(allEffects(model.it))
```







3 Polish

Fit a mixed effect linear model for Polish

```
vowels.pl <- filter(vowels, language == "polish")

model.pl <- lmer(duration ~ c2phonation * vowel + c2place +
                   (1 | word),
                 data = vowels.pl
                 )

summary(model.pl)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: duration ~ c2phonation * vowel + c2place + (1 | word)
##    Data: vowels.pl
##
## REML criterion at convergence: 789.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.6720 -0.5792 -0.1033  0.3587  6.1371
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  word     (Intercept)  51.93    7.206
```

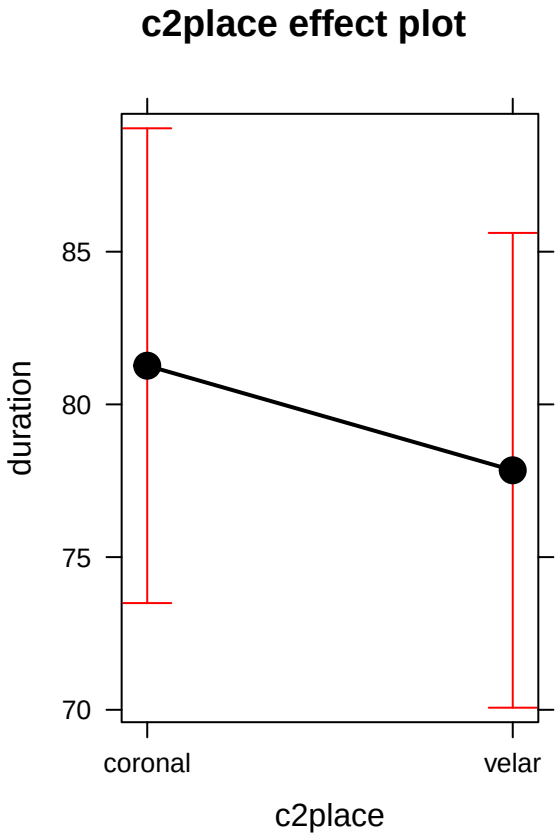
```
##  Residual             319.10   17.863
## Number of obs: 96, groups:  word, 12
##
## Fixed effects:
##                             Estimate Std. Error t value
## (Intercept)                   95.489      7.319  13.047
## c2phonationvoiceless          -7.286      9.582  -0.760
## vowelo                        -9.843      9.582  -1.027
## vowelu                       -28.654      9.582  -2.990
## c2placevelar                  -3.425      5.532  -0.619
## c2phonationvoiceless:vowelo    7.162     13.551   0.528
## c2phonationvoiceless:vowelu    6.360     13.551   0.469
##
## Correlation of Fixed Effects:
##                             (Intr) c2phnt vowelo vowelu c2plcv
## c2phntnvcls                 -0.655
## vowelo                      -0.655  0.500
## vowelu                      -0.655  0.500  0.500
## c2placevelr                 -0.378  0.000  0.000  0.000
## c2phonationvoiceless:vowelo  0.463 -0.707 -0.707 -0.354  0.000
## c2phonationvoiceless:vowelu  0.463 -0.707 -0.354 -0.707  0.000
##                             c2phonationvoiceless:vowelo
## c2phntnvcls
## vowelo
## vowelu
## c2placevelr
## c2phonationvoiceless:vowelo
## c2phonationvoiceless:vowelu  0.500
```

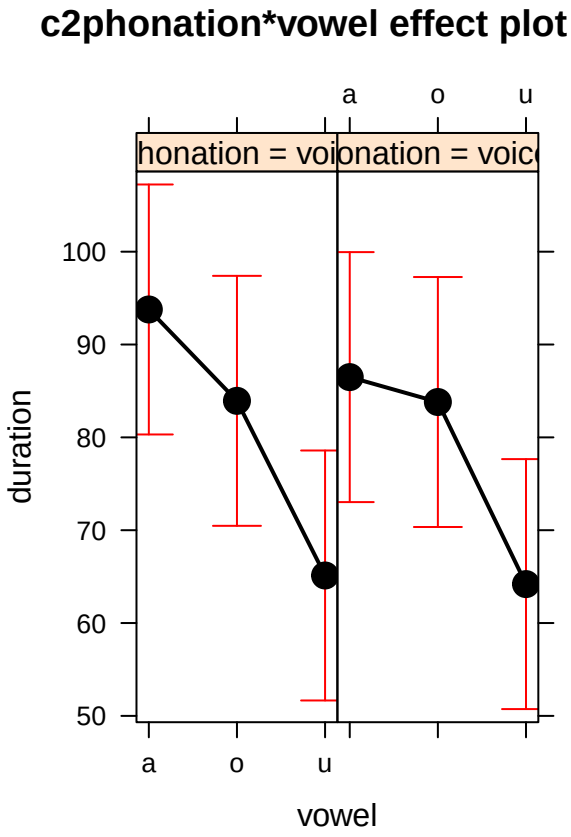
```
mixed(duration ~ c2phonation * vowel + c2place +
                  (1 | word),
              data = vowels.pl)
```

```
## Fitting 5 (g)lmer() models:
## [.....]
## Obtaining 4 p-values:
## [....]
```

```
##              Effect   df F.scaling      F p.value
## 1       c2phonation 1, 5      1.00   0.58     .48
## 2             vowel 2, 5      1.00 4.62 +     .07
## 3           c2place 1, 5      1.00   0.38     .56
## 4 c2phonation:vowel 2, 5      1.00   0.17     .85
```

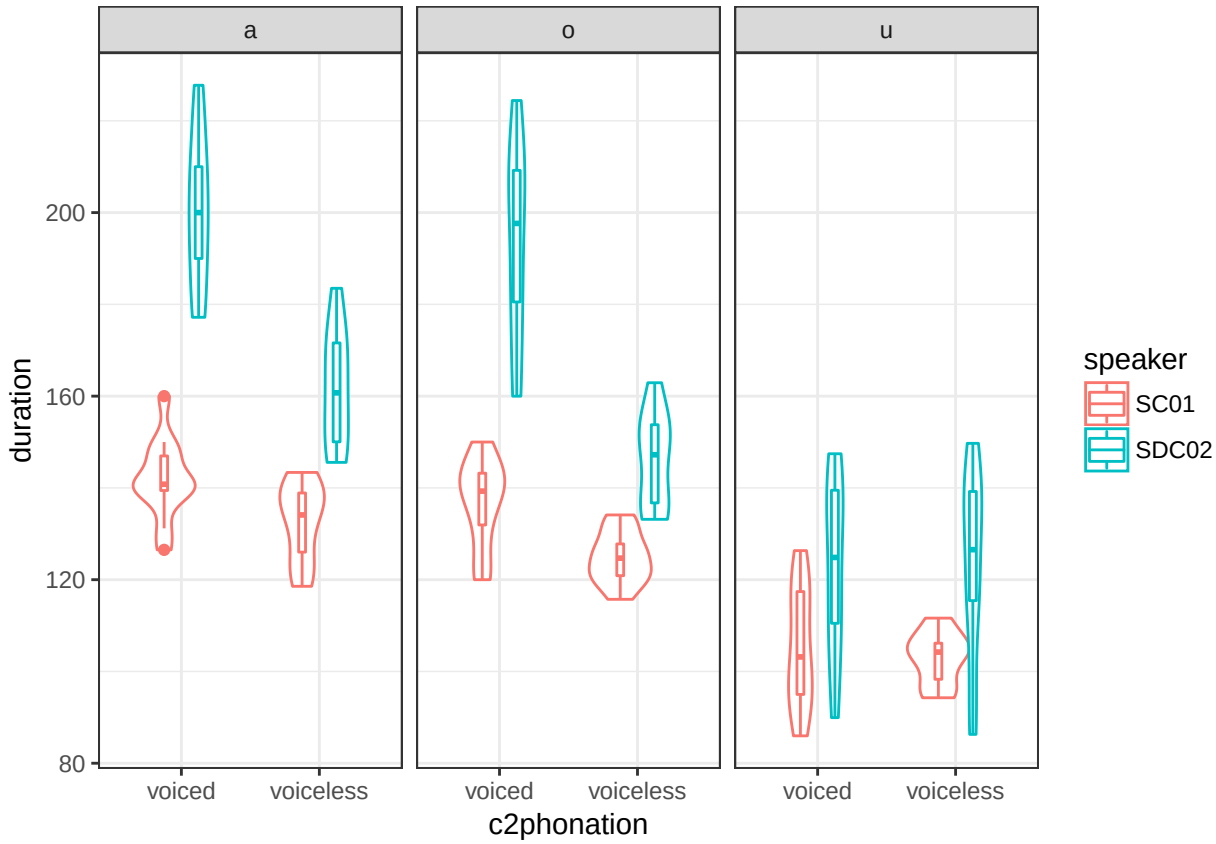
```
plot(allEffects(model.pl))
```



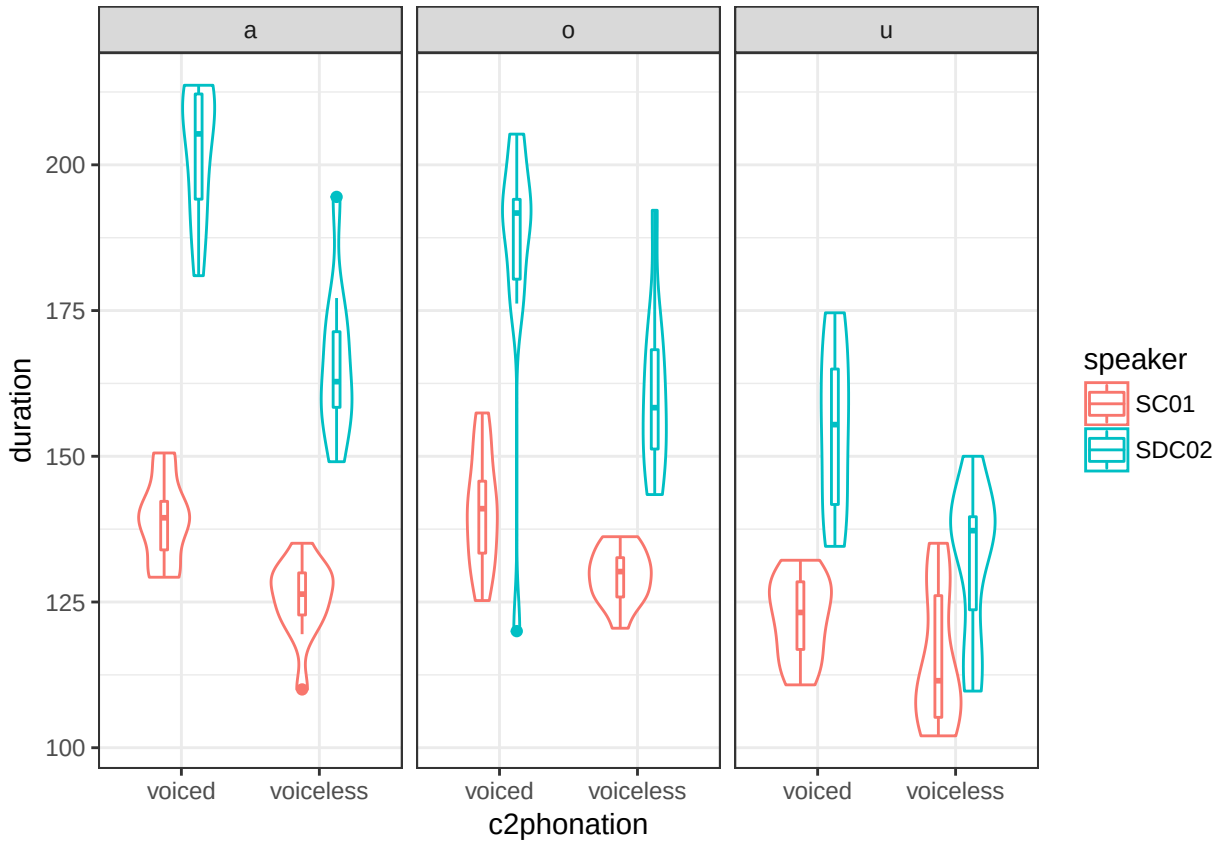


4 Inter-speaker variation

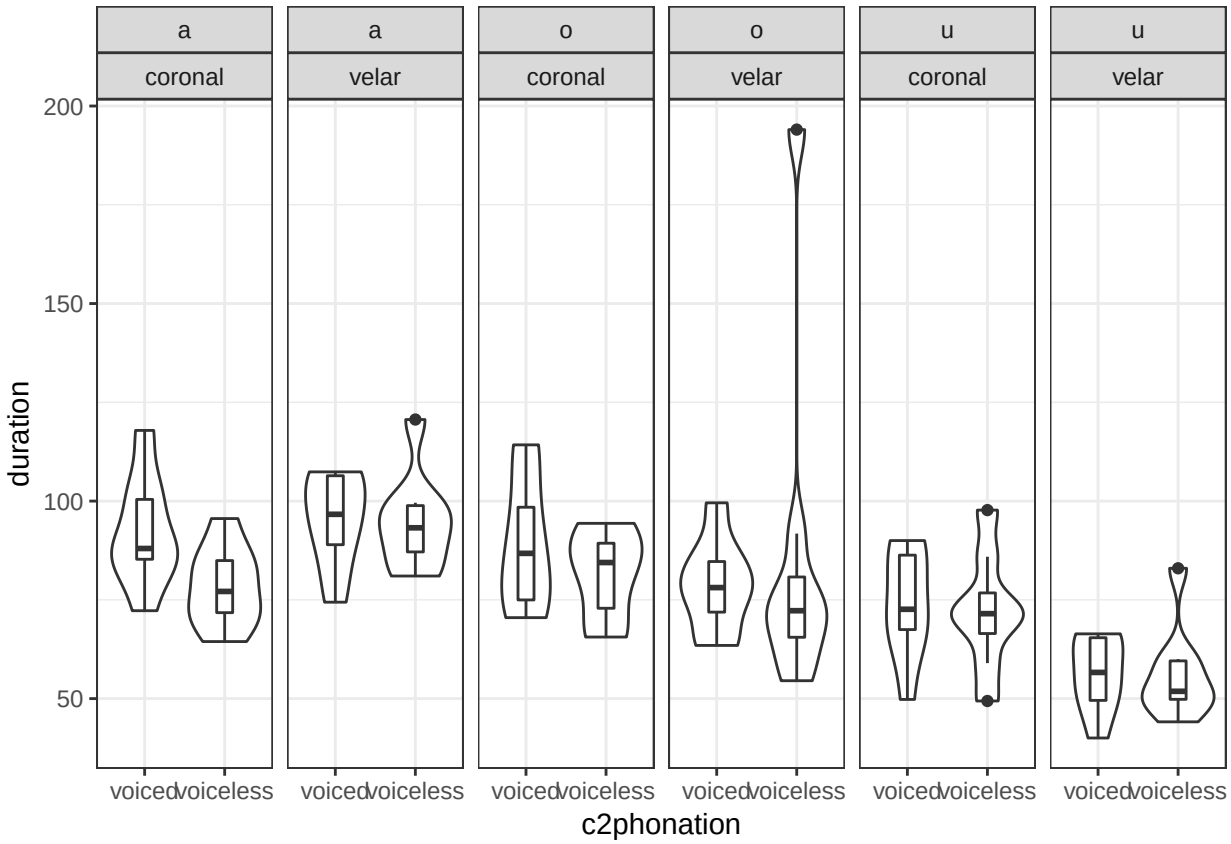
```
dodge <- position_dodge(width = 0.5)

ggplot(filter(vowels.it, c2place == "velar"), aes(c2phonation, duration, colour = speaker)) +
    facet_grid(. ~ vowel) +
    geom_violin(position = dodge) +
    geom_boxplot(width=0.1, position = dodge)
```




```
ggplot(filter(vowels.it, c2place == "coronal"), aes(c2phonation, duration, colour = speaker)) +
    facet_grid(. ~ vowel) +
    geom_violin(position = dodge) +
    geom_boxplot(width=0.1, position = dodge)
```




```
ggplot(vowels.pl, aes(c2phonation, duration)) +
    facet_grid(. ~ vowel + c2place) +
    geom_violin() +
    geom_boxplot(width=0.2)
```


a
coronal
a
velar
o
coronal
o
velar
u
coronal
u
velar
200
150
100
50
duration
voiced
voiceless
c2phonation