Description of system performance: using the pipeline v1 as example

AC

Contents

A first attempt

The goal here is to link system performance to characteristics of files and/or speakers on files. See this file for explanation of the fields.

Read in descriptor data

We read in background data (only exists for babytrain), data describing files and speakers. Typically, this does not depend on your system results, so you do not need to change it.

Read in system evaluation data

HUMAN LOOK HERE Typically you WILL need to change line 50 below, so that you read in your own system results. Please use pyannote.metrics to generate your results. They should be space separated.

IMPORTANT!!! THERE SHOULD BE NO WARNINGS IN IT. IF THERE ARE WARNINGS, DEAL WITH THEM AND REGENERATE A RESULTS FILE WITHOUT THEM

For this example, I manually removed the warnings in these 4 files…

To make full use of this description suite, you also need to generate a set of descriptors that cross the gold and the system output. To do so, navigate to computation/scripts/ and open the README. Follow the instructions there to set up the analysis environment. Next, assume you have a rttm file with all of the output from a single system for all the wav files.

This next chunk takes the output of the previous one and combines them into a table that can be used in further analyses

HUMAN END OF LOOK HERE

Combine descriptor and system data

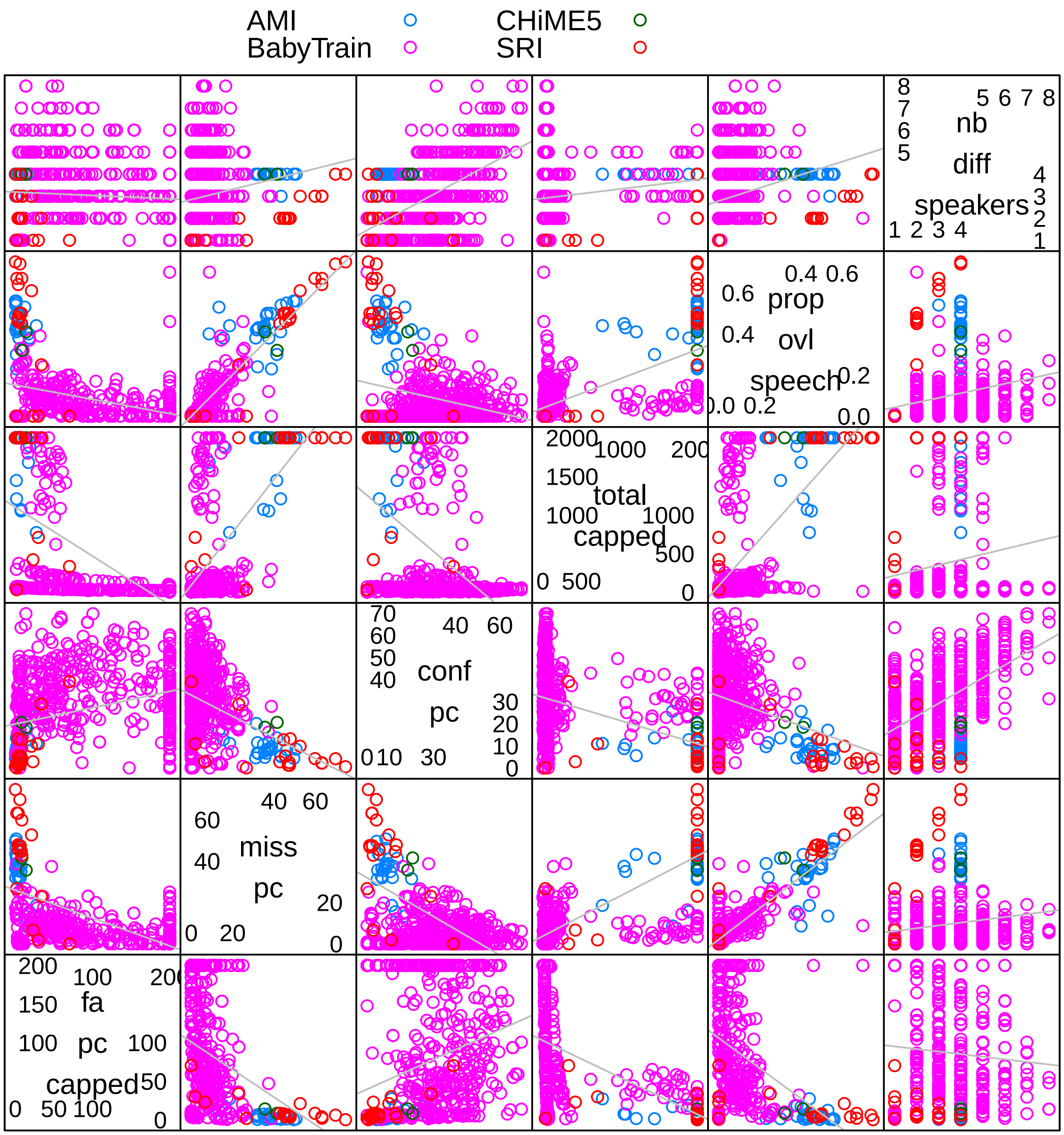
If all goes well, you won't need to change any of the following sections. After this code, the table file_eval has a combination of results and descriptors at the level of files.

Now you are ready to do some inspection. You can turn chunks off by adding ", eval=F" (e.g. {r spl,fig.height=10} below, it would become {r spl,fig.height=10, eval=F})

Example of analysis: Focusing on our pipeline V1

A scatter plot matrix shows many bivariate plots. In the one below, we focus exclusively on descriptors at the level of the file.

```
## Warning: package 'lattice' was built under R version 3.4.4
```



Scatter Plot Matrix

Focus on the last three rows, which show the correlations between percent false alarms (last row – capped at 200, meaning that any file that had more than 200 gets 200), percent misses (penultimate row), and percent confusion (antepen row), and the following selected characteristics (from left to right):

- proportion of speech that is overlapping (out of all speech in the file, which part was overlapping between two talkers)

- number of different speakers

For example, focusing on the last row, there is a lower percent of false alarms for corpora with more overlap (although this is mainly driven by corpus differences).

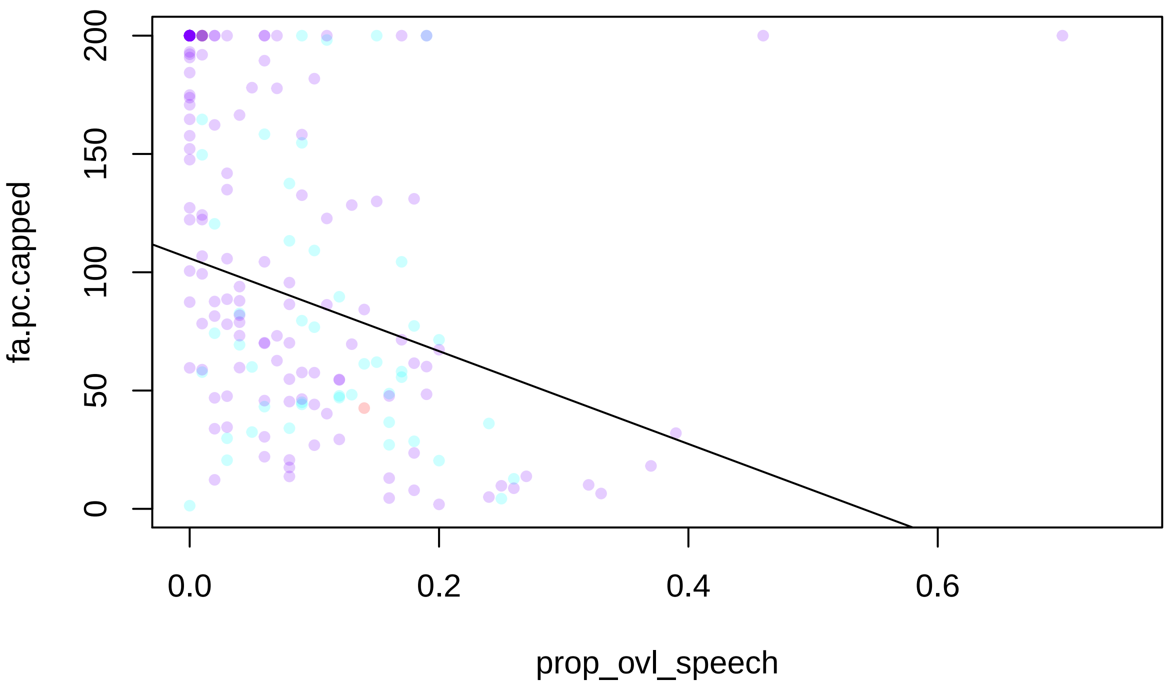
One row up, percent misses goes up with proportion of overlap (same caveat).

One row up, percent confusion goes down slightly with proportion overlap (but same caveat); and it goes up with number of talkers.

Example of analysis 2: Still using the pipeline v1

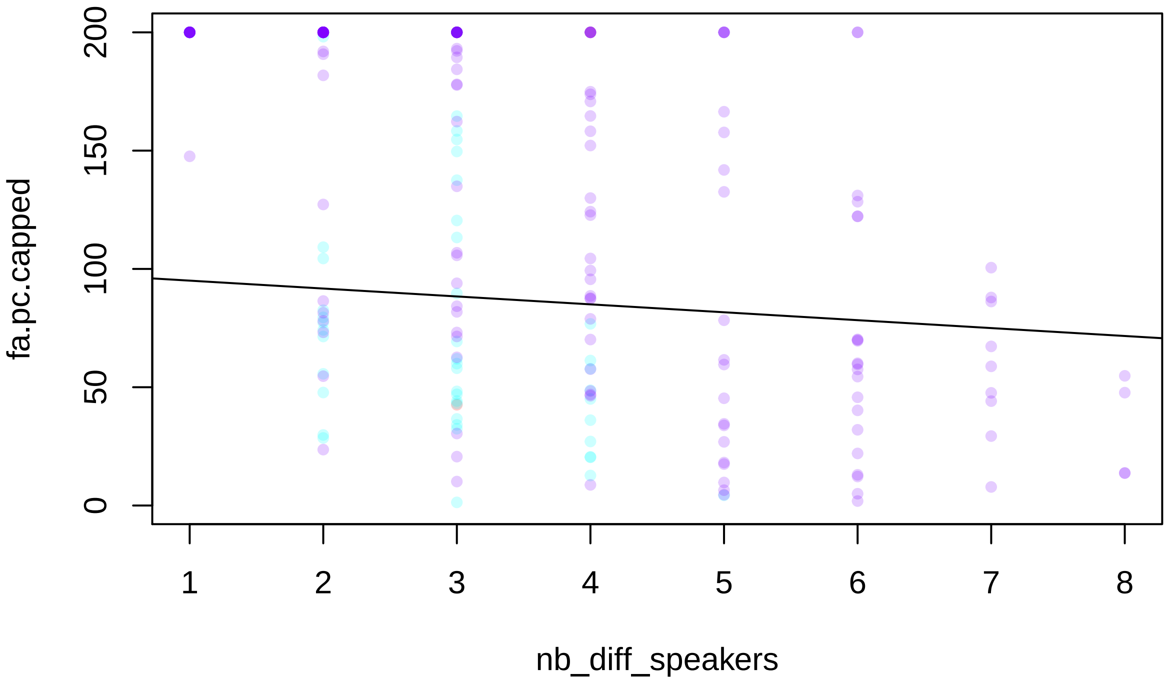
You can also focus on specific outcome and predictor variables and trim their distribution to see them more clearly.

```
## [1] "removing 0 outliers in fa.pc.capped specifically the following files:"
## character(0)
```



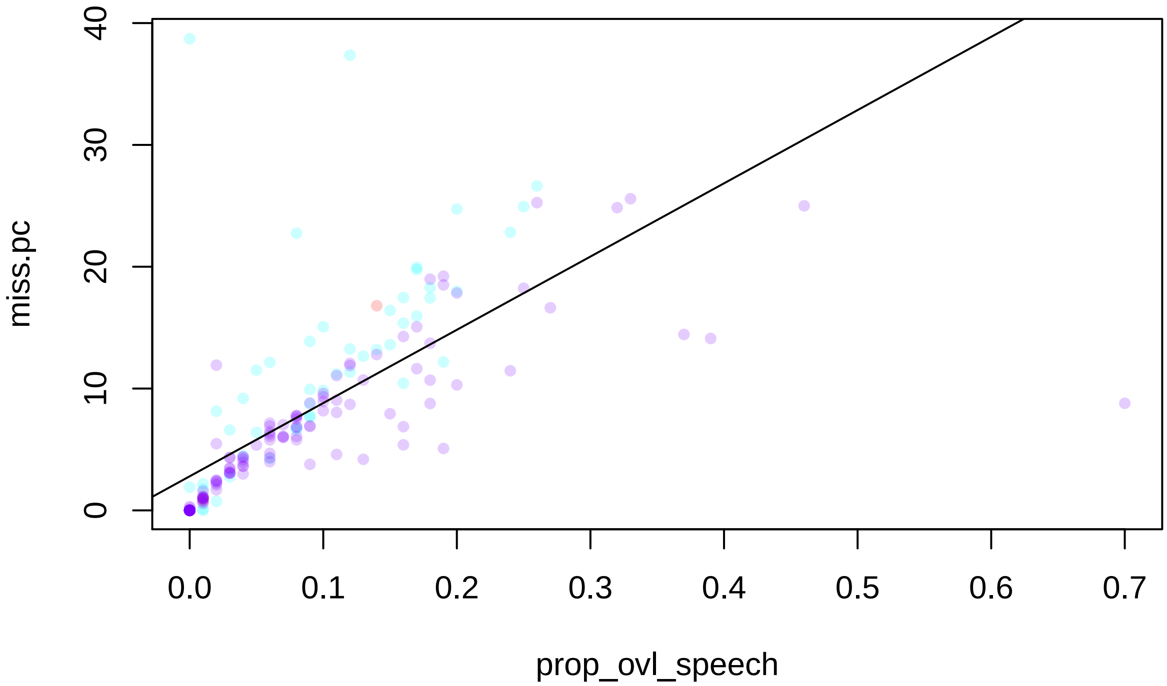

```
##
## Call:
## lm(formula = no_outliers[, thismet] ~ no_outliers[, thispred])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -105.48  -54.07  -11.62   81.08  231.43
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)              105.895      3.951  26.799  < 2e-16 ***
## no_outliers[, thispred] -196.175     23.131  -8.481 3.14e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 71.8 on 456 degrees of freedom
```

```
##   (139 observations deleted due to missingness)
## Multiple R-squared:  0.1362, Adjusted R-squared:  0.1343
## F-statistic: 71.93 on 1 and 456 DF,  p-value: 3.138e-16
```

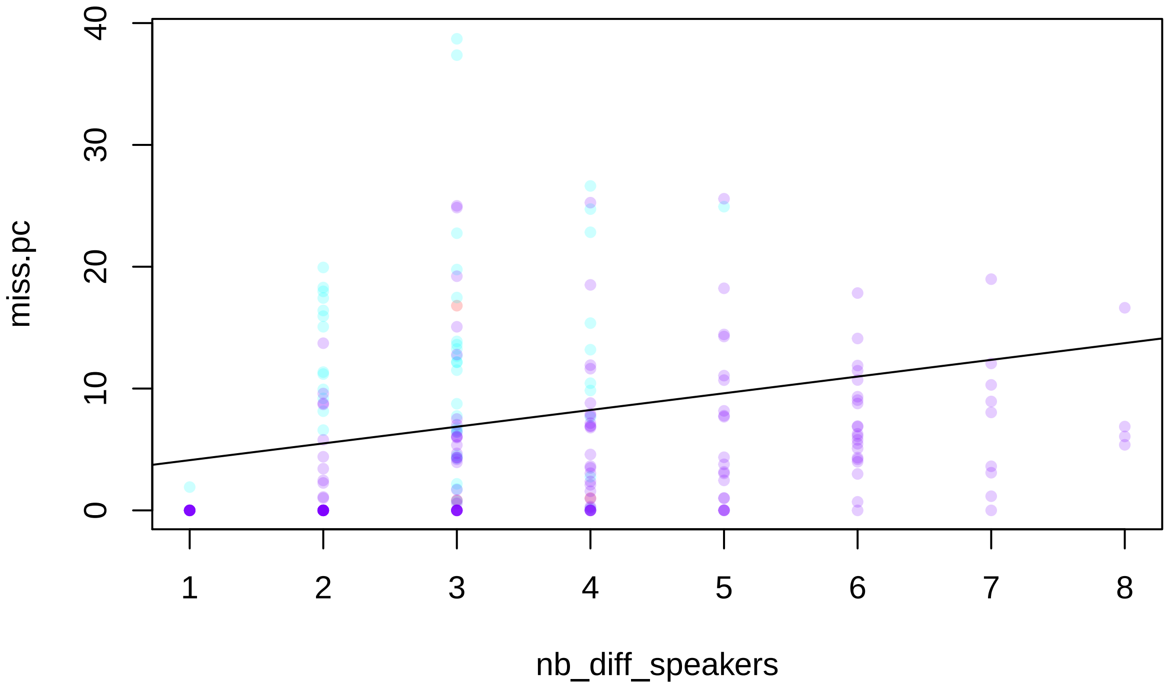



```
##
## Call:
## lm(formula = no_outliers[, thismet] ~ no_outliers[, thispred])
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -94.65 -73.20 -20.89  89.82 121.66
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)               98.399      7.560  13.016   <2e-16 ***
## no_outliers[, thispred]   -3.343      2.176  -1.536    0.125
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 77.05 on 456 degrees of freedom
##   (139 observations deleted due to missingness)
## Multiple R-squared:  0.005148,   Adjusted R-squared:  0.002967
## F-statistic:  2.36 on 1 and 456 DF,  p-value: 0.1252
##
## [1] "removing 121 outliers in miss.pc specifically the following files:"
##   [1] "EN2002a.Mix-Headset"       "EN2002b.Mix-Headset"
##   [3] "EN2002c.Mix-Headset"       "EN2002d.Mix-Headset"
##   [5] "jsalt-rm1-se06.2-mc05-stu" "jsalt-rm1-se06.2-mc09-lap"
##   [7] "jsalt-rm1-se06.2-mc10-lap" "jsalt-rm1-se06.2-mc16-ele"
##   [9] "jsalt-rm1-se07.1-mc05-stu" "jsalt-rm1-se07.1-mc09-lap"
##  [11] "jsalt-rm1-se07.1-mc10-lap" "jsalt-rm1-se07.1-mc16-ele"
##  [13] "jsalt-rm1-se07.2-mc05-stu" "jsalt-rm1-se07.2-mc09-lap"
##  [15] "jsalt-rm1-se07.2-mc10-lap" "jsalt-rm1-se07.2-mc16-ele"
```

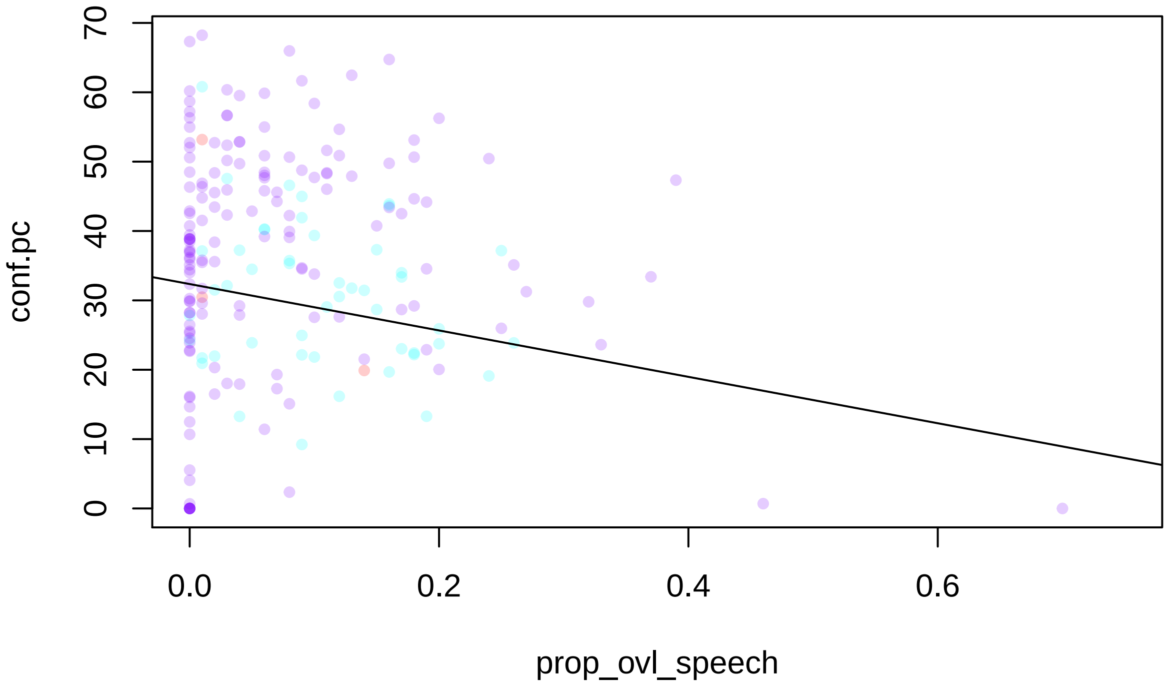
```
##  [17] "jsalt-rm1-se09.1-mc05-stu" "jsalt-rm1-se09.1-mc09-lap"
##  [19] "jsalt-rm1-se09.1-mc10-lap" "jsalt-rm1-se09.1-mc16-ele"
##  [21] "jsalt-rm1-se09.2-mc05-stu" "jsalt-rm1-se09.2-mc09-lap"
##  [23] "jsalt-rm1-se09.2-mc10-lap" "jsalt-rm1-se09.2-mc16-ele"
##  [25] "jsalt-rm2-se04.1-mc05-stu" "jsalt-rm2-se04.1-mc11-lap"
##  [27] "jsalt-rm2-se04.1-mc16-ele" "jsalt-rm2-se04.1-mc20-ele"
##  [29] "jsalt-rm2-se04.1-mc22-ele" "jsalt-rm2-se05.1-mc05-stu"
##  [31] "jsalt-rm2-se05.1-mc11-lap" "jsalt-rm2-se05.1-mc16-ele"
##  [33] "jsalt-rm2-se05.1-mc20-ele" "jsalt-rm2-se05.1-mc22-ele"
##  [35] "jsalt-rm2-se06.1-mc05-stu" "jsalt-rm2-se06.1-mc11-lap"
##  [37] "jsalt-rm2-se06.1-mc16-ele" "jsalt-rm2-se06.1-mc20-ele"
##  [39] "jsalt-rm2-se06.1-mc22-ele" "jsalt-rm2-se06.3-mc05-stu"
##  [41] "jsalt-rm2-se06.3-mc11-lap" "jsalt-rm2-se06.3-mc16-ele"
##  [43] "jsalt-rm2-se06.3-mc20-ele" "jsalt-rm2-se06.3-mc22-ele"
##  [45] "jsalt-rm2-se07.1-mc05-stu" "jsalt-rm2-se07.1-mc11-lap"
##  [47] "jsalt-rm2-se07.1-mc16-ele" "jsalt-rm2-se07.1-mc20-ele"
##  [49] "jsalt-rm2-se07.1-mc22-ele" "jsalt-rm2-se08.1-mc05-stu"
##  [51] "jsalt-rm2-se08.1-mc11-lap" "jsalt-rm2-se08.1-mc16-ele"
##  [53] "jsalt-rm2-se08.1-mc20-ele" "jsalt-rm2-se08.1-mc22-ele"
##  [55] "jsalt-rm2-se10.1-mc05-stu" "jsalt-rm2-se10.1-mc11-lap"
##  [57] "jsalt-rm2-se10.1-mc16-ele" "jsalt-rm2-se10.1-mc20-ele"
##  [59] "jsalt-rm2-se10.1-mc22-ele" "jsalt-rm2-se11.1-mc05-stu"
##  [61] "jsalt-rm2-se11.1-mc11-lap" "jsalt-rm2-se11.1-mc16-ele"
##  [63] "jsalt-rm2-se11.1-mc20-ele" "jsalt-rm2-se11.1-mc22-ele"
##  [65] "jsalt-rm3-se02.1-mc01-stu" "jsalt-rm3-se02.1-mc07-stu"
##  [67] "jsalt-rm3-se02.1-mc11-lap" "jsalt-rm3-se02.1-mc13-tab"
##  [69] "jsalt-rm3-se02.1-mc16-ele" "jsalt-rm3-se02.1-mc18-ele"
##  [71] "jsalt-rm3-se02.2-mc01-stu" "jsalt-rm3-se02.2-mc07-stu"
##  [73] "jsalt-rm3-se02.2-mc11-lap" "jsalt-rm3-se02.2-mc13-tab"
##  [75] "jsalt-rm3-se02.2-mc16-ele" "jsalt-rm3-se02.2-mc18-ele"
##  [77] "jsalt-rm3-se05.1-mc01-stu" "jsalt-rm3-se05.1-mc07-stu"
##  [79] "jsalt-rm3-se05.1-mc11-lap" "jsalt-rm3-se05.1-mc13-tab"
##  [81] "jsalt-rm3-se05.1-mc16-ele" "jsalt-rm3-se05.1-mc18-ele"
##  [83] "jsalt-rm3-se06.1-mc01-stu" "jsalt-rm3-se06.1-mc07-stu"
##  [85] "jsalt-rm3-se06.1-mc11-lap" "jsalt-rm3-se06.1-mc13-tab"
##  [87] "jsalt-rm3-se06.1-mc16-ele" "jsalt-rm3-se06.1-mc18-ele"
##  [89] "jsalt-rm3-se06.2-mc01-stu" "jsalt-rm3-se06.2-mc07-stu"
##  [91] "jsalt-rm3-se06.2-mc11-lap" "jsalt-rm3-se06.2-mc13-tab"
##  [93] "jsalt-rm3-se06.2-mc16-ele" "jsalt-rm3-se06.2-mc18-ele"
##  [95] "jsalt-rm3-se07.1-mc01-stu" "jsalt-rm3-se07.1-mc07-stu"
##  [97] "jsalt-rm3-se07.1-mc11-lap" "jsalt-rm3-se07.1-mc13-tab"
##  [99] "jsalt-rm3-se07.1-mc16-ele" "jsalt-rm3-se07.1-mc18-ele"
## [101] "jsalt-rm3-se07.2-mc01-stu" "jsalt-rm3-se07.2-mc07-stu"
## [103] "jsalt-rm3-se07.2-mc11-lap" "jsalt-rm3-se07.2-mc13-tab"
## [105] "jsalt-rm3-se07.2-mc16-ele" "jsalt-rm3-se07.2-mc18-ele"
## [107] "jsalt-rm3-se08.1-mc01-stu" "jsalt-rm3-se08.1-mc07-stu"
## [109] "jsalt-rm3-se08.1-mc11-lap" "jsalt-rm3-se08.1-mc13-tab"
## [111] "jsalt-rm3-se08.1-mc16-ele" "jsalt-rm3-se08.1-mc18-ele"
## [113] "jsalt-rm3-se08.2-mc01-stu" "jsalt-rm3-se08.2-mc07-stu"
## [115] "jsalt-rm3-se08.2-mc11-lap" "jsalt-rm3-se08.2-mc13-tab"
## [117] "jsalt-rm3-se08.2-mc16-ele" "jsalt-rm3-se08.2-mc18-ele"
## [119] "S21_U01"                   "TS3003a.Mix-Headset"
## [121] "TS3007a.Mix-Headset"
```



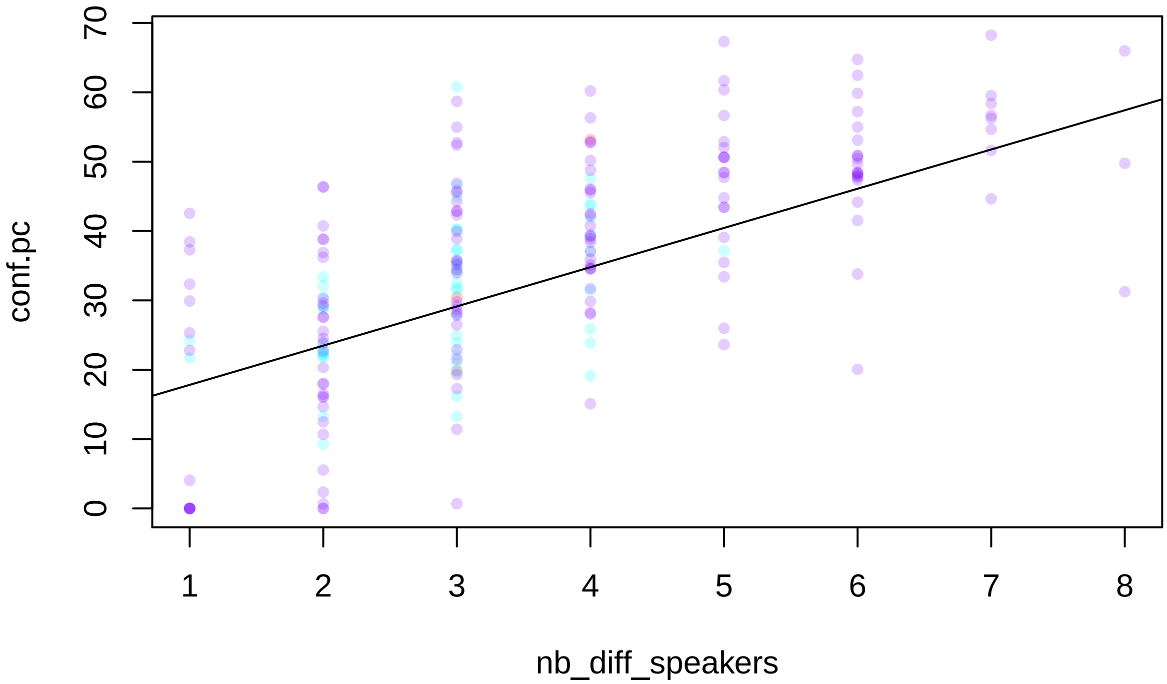

```
##
## Call:
## lm(formula = no_outliers[, thismet] ~ no_outliers[, thispred])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -36.094  -2.795  -1.570   1.262  35.915
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)               2.7954     0.3191   8.759   <2e-16 ***
## no_outliers[, thispred]  60.1262     2.4989  24.061   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.614 on 435 degrees of freedom
##   (39 observations deleted due to missingness)
## Multiple R-squared:  0.571,  Adjusted R-squared:   0.57
## F-statistic:   579 on 1 and 435 DF,  p-value: < 2.2e-16
```




```
##
## Call:
## lm(formula = no_outliers[, thismet] ~ no_outliers[, thispred])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -12.355  -4.703  -3.162   2.793  31.843
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)               2.7502     0.8186   3.360  0.00085 ***
## no_outliers[, thispred]   1.3721     0.2346   5.849 9.71e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.252 on 435 degrees of freedom
##   (39 observations deleted due to missingness)
## Multiple R-squared:  0.07292,    Adjusted R-squared:  0.07079
## F-statistic: 34.22 on 1 and 435 DF,  p-value: 9.706e-09
##
## [1] "removing 2 outliers in conf.pc specifically the following files:"
## [1] "namibia_uebn_20170309_1980"  "namibia_uebn_20170309_30780"
```




```
##
## Call:
## lm(formula = no_outliers[, thismet] ~ no_outliers[, thispred])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.353 -11.394  -0.216  12.273  37.729
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)              32.3533     0.8945  36.169  < 2e-16 ***
## no_outliers[, thispred] -33.4521     5.2262  -6.401 3.86e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.21 on 454 degrees of freedom
##   (139 observations deleted due to missingness)
## Multiple R-squared:  0.08278,    Adjusted R-squared:  0.08075
## F-statistic: 40.97 on 1 and 454 DF,  p-value: 3.857e-10
```




```
##
## Call:
## lm(formula = no_outliers[, thismet] ~ no_outliers[, thispred])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -34.235 -10.934   1.152  10.209  45.606
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)              12.1577     1.4027   8.667   <2e-16 ***
## no_outliers[, thispred]   5.6567     0.4071  13.896   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.18 on 454 degrees of freedom
##   (139 observations deleted due to missingness)
## Multiple R-squared:  0.2984, Adjusted R-squared:  0.2968
## F-statistic: 193.1 on 1 and 454 DF,  p-value: < 2.2e-16
```

Messages I take away from this:

For FA rate capped

(conclusions to be taken with a grain of salt, given that capping shifts the distribution)

- significantly lower FA with higher proportion of overlapping speech
- no sig different FA as a function of number of different speakers

For miss rate

Attention that most SRI files get excluded as outliers for miss rate…

- sig higher miss for files with higher prop overlapping speech
- sig higher miss when more speakers

For confusion

- sig lower confusion for files with higher prop overlapping speech
- sig higher confusion when more speakers