Analyses

TODO discussion 2021-09-20

- why negative: because inverted u and a in 2 places, see chunk calc-space
- why different dfs: I wasn't doing the same analysis, now I am
- why different Fs: because the vss data was different (due to formula switch)
- explain about 2 rows that are excluded, & possible mismatch
- do we have pitch.txt & formant.txt outputted by script?
  - no, but we have output by mom! Alex to integrate
- we can also leave that be, since by now we are aligned in terms of means for point vowels?
  - no, because duration & pitch are incorrectly attributed
- not sure what to do about the fact that our vowel triangles don't look the same – how were those generated? perhaps they are not over the same tokens?
  - mystery solved! i and u based on all moms in both figures
  - what shall we do, single figure for all moms?
- "REAL" decisions:
  - removing outliers? – yes
  - removing items from f1/2 when no f0 – keeping them because they are not outliers
  - doing ANOVA first, then mixed model- agreed
  - (I don't think software matters!) - ANOVA in spss, lmer in r

interim done:

- switched input file, results are stable
- add confidence intervals to vowel space, incorporate results to paper

Replication

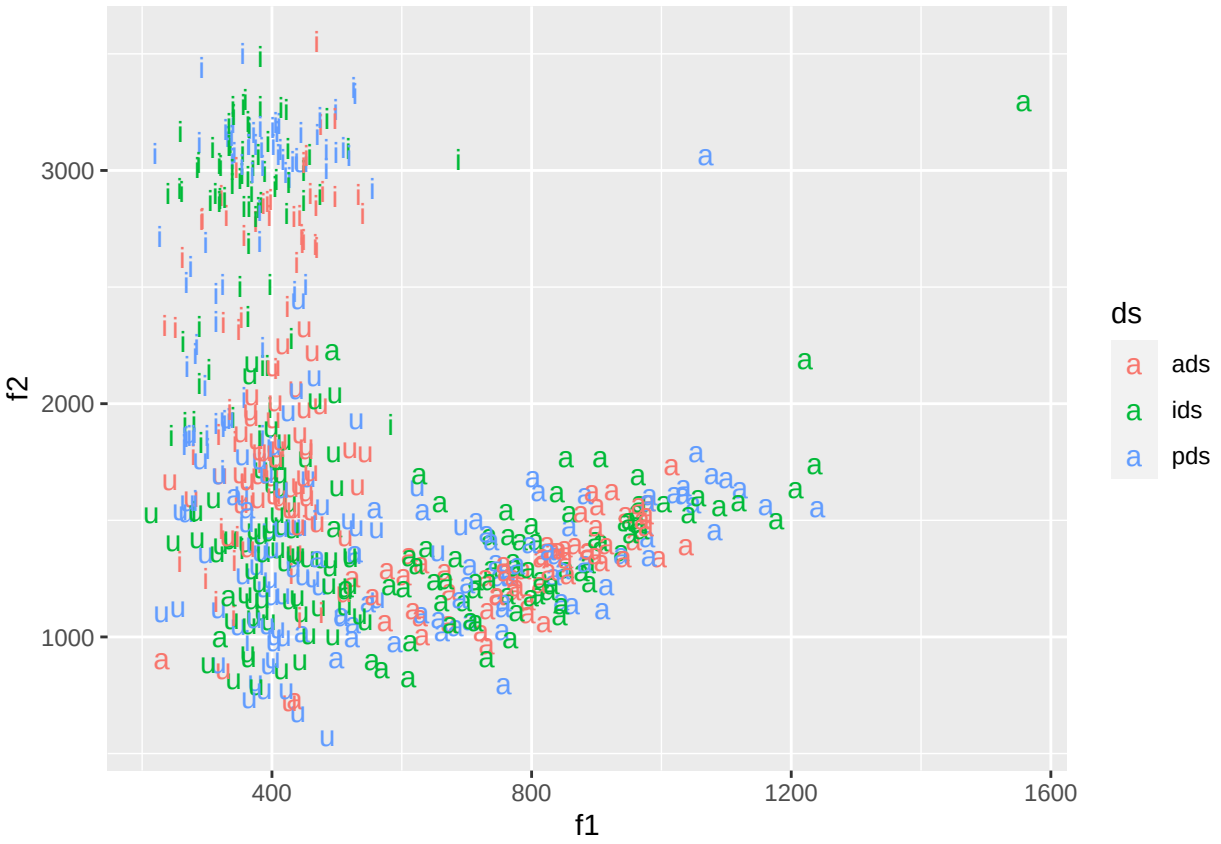
In this section, I'm just doing again the analyses that are already in the manuscript. Those were done with SPSS, so this is more replication than reproduction.

```
## [1] 614   7
```

Checking that we are getting the same means against the means in the xls file – we are.

Checking that we are getting the same means against the means in the sav file – we are.

Note there are some likely outliers:



we may want to kick them out:

```
##                  Mom Word Listener   f1   f2 duration pitch motID  ds v
## 338 mom1idsbox5v.wav    3        1 1221 2194     0.27   137  mom1 ids a
##     stim_group target
## 338        box    box
```

```
##                  Mom Word Listener   f1   f2 duration pitch motID  ds v
## 334 mom1idsbox1v.wav    3        1 1558 3298    0.383   154  mom1 ids a
##     stim_group target
## 334        box    box
```

```
##                  Mom Word Listener   f1   f2 duration pitch motID  ds v
## 80 mom9pdsball7v.wav    4        2 1068 3068    0.146   213  mom9 pds a
##    stim_group target
## 80       ball   ball
```

```
## [1] 609  12
```

```
## `summarise()` has grouped output by 'motID', 'ds', 'v', 'stim_group'. You can override using the `.g
```

```
## `summarise()` has grouped output by 'ds', 'v'. You can override using the `.groups` argument.
```

```
## `summarise()` has grouped output by 'motID', 'ds'. You can override using the `.groups` argument.
```

```
## `summarise()` has grouped output by 'ds'. You can override using the `.groups` argument.
## `summarise()` has grouped output by 'ds'. You can override using the `.groups` argument.
## `summarise()` has grouped output by 'ds'. You can override using the `.groups` argument.
```

Figs 1-2

Get average point vowels and draw figures 1 & 2.

Note that they don't look identical. That is because Fig 1 and 2 were using all 10 moms for the /i/ and /u/ points, but only the relevant moms for the /a/ points. We decide to just collapse across groups and have a single figure with all 10 moms.

First check whether there are significant differences for each vowel on F1 and F2. Differences on F2 for both /i/ and /u/; n/s differences on F1 for either of those; neither F1 nor F2 differ across registers for /a/.

```
## boundary (singular) fit: see ?isSingular
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: f1 ~ ds + (1 + ds | motID)
##    Data: dat
##  Subset: c(v == "i")
##
## REML criterion at convergence: 2343.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.7944 -0.6174 -0.0394  0.4882  4.8783
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  motID    (Intercept) 2882     53.69
##           dsids       2068     45.48    -0.89
##           dspds       1030     32.10    -0.96  0.73
##  Residual             4604     67.85
## Number of obs: 208, groups:  motID, 10
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  393.716     19.447   9.529  20.245 3.75e-09 ***
## dsids        -26.434     18.996  11.090  -1.392    0.191
## dspds        -16.505     15.907  12.758  -1.038    0.319
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##       (Intr) dsids
## dsids -0.829
## dspds -0.823  0.640
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: f2 ~ ds + (1 + ds | motID)
##    Data: dat
##  Subset: c(v == "i")
##
## REML criterion at convergence: 3149.4
##
```

```
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.3172 -0.4745  0.1470  0.6137  2.1658
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  motID    (Intercept)  87758   296.2
##           dsids       204492   452.2    -0.68
##           dspds       231394   481.0    -0.70  0.90
##  Residual             219986   469.0
## Number of obs: 208, groups:  motID, 10
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept) 2399.491    114.444    8.691  20.967 9.58e-09 ***
## dsids        381.948    167.508    8.142   2.280   0.0515 .
## dspds        334.700    175.154    8.776   1.911   0.0892 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##       (Intr) dsids
## dsids -0.703
## dspds -0.716  0.817
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: f1 ~ ds + (1 + ds | motID)
##    Data: dat
##  Subset: c(v == "u")
##
## REML criterion at convergence: 2231.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.7054 -0.5835 -0.1209  0.5959  3.4969
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  motID    (Intercept) 1260.2   35.50
##           dsids        937.9   30.62    -0.57
##           dspds       1945.7   44.11    -0.02 -0.05
##  Residual             3888.6   62.36
## Number of obs: 200, groups:  motID, 10
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept) 411.2328    13.9247   8.8778  29.533 3.58e-10 ***
## dsids         0.4875    14.7855   8.6008   0.033    0.974
## dspds         3.5278    17.9069   9.1415   0.197    0.848
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
```

```
##       (Intr) dsids
## dsids -0.631
## dspds -0.284  0.230
## boundary (singular) fit: see ?isSingular
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: f2 ~ ds + (1 + ds | motID)
##    Data: dat
##  Subset: c(v == "u")
##
## REML criterion at convergence: 2790.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2048 -0.5698 -0.0336  0.5940  4.0793
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  motID    (Intercept) 34561    185.9
##           dsids       15055    122.7    -0.13
##           dspds       19006    137.9     0.79  0.50
##  Residual             66744    258.3
## Number of obs: 200, groups:  motID, 10
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept) 1722.644     67.941    8.931  25.355 1.25e-09 ***
## dsids       -361.183     60.333    6.477  -5.987 0.000736 ***
## dspds       -346.383     63.625   11.945  -5.444 0.000152 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##       (Intr) dsids
## dsids -0.355
## dspds  0.203  0.526
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
## boundary (singular) fit: see ?isSingular
## Warning: Model failed to converge with 1 negative eigenvalue: -1.7e+00
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: f1 ~ ds + (1 + ds | motID)
##    Data: dat
##  Subset: c(v == "a")
##
## REML criterion at convergence: 2593.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.4122 -0.5431  0.0688  0.5354  2.8205
```

```
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  motID    (Intercept)  9235     96.10
##           dsids        5140     71.69   -0.19
##           dspds        1822     42.69   -0.61 -0.67
##  Residual             23557    153.48
## Number of obs: 201, groups:  motID, 10
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  808.886     36.240   9.484  22.320 1.62e-09 ***
## dsids          6.247     35.223   9.837   0.177    0.863
## dspds         -7.171     30.879  19.138  -0.232    0.819
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##       (Intr) dsids
## dsids -0.398
## dspds -0.563  0.152
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: f2 ~ ds + (1 + ds | motID)
##    Data: dat
##  Subset: c(v == "a")
##
## REML criterion at convergence: 2671.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.5532 -0.5246  0.0408  0.5580  5.2935
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  motID    (Intercept) 20005    141.44
##           dsids       24277    155.81   -0.72
##           dspds        4447     66.68    0.01  0.15
##  Residual             33732    183.66
## Number of obs: 201, groups:  motID, 10
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept) 1318.702     50.683    8.577  26.019 1.84e-09 ***
## dsids         45.277     59.204    8.710   0.765    0.465
## dspds         53.666     39.777    7.216   1.349    0.218
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##       (Intr) dsids
```

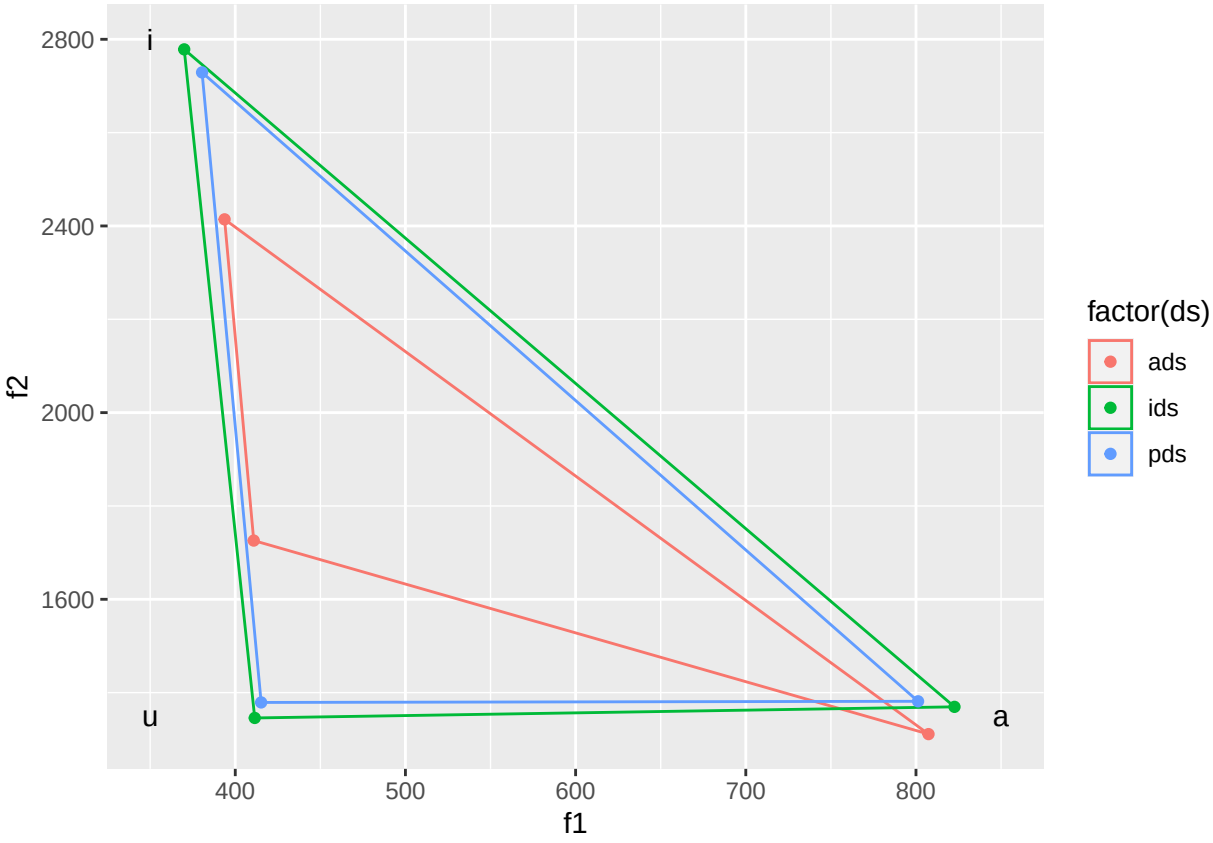
```
## dsids -0.721
## dspds -0.270  0.306
```

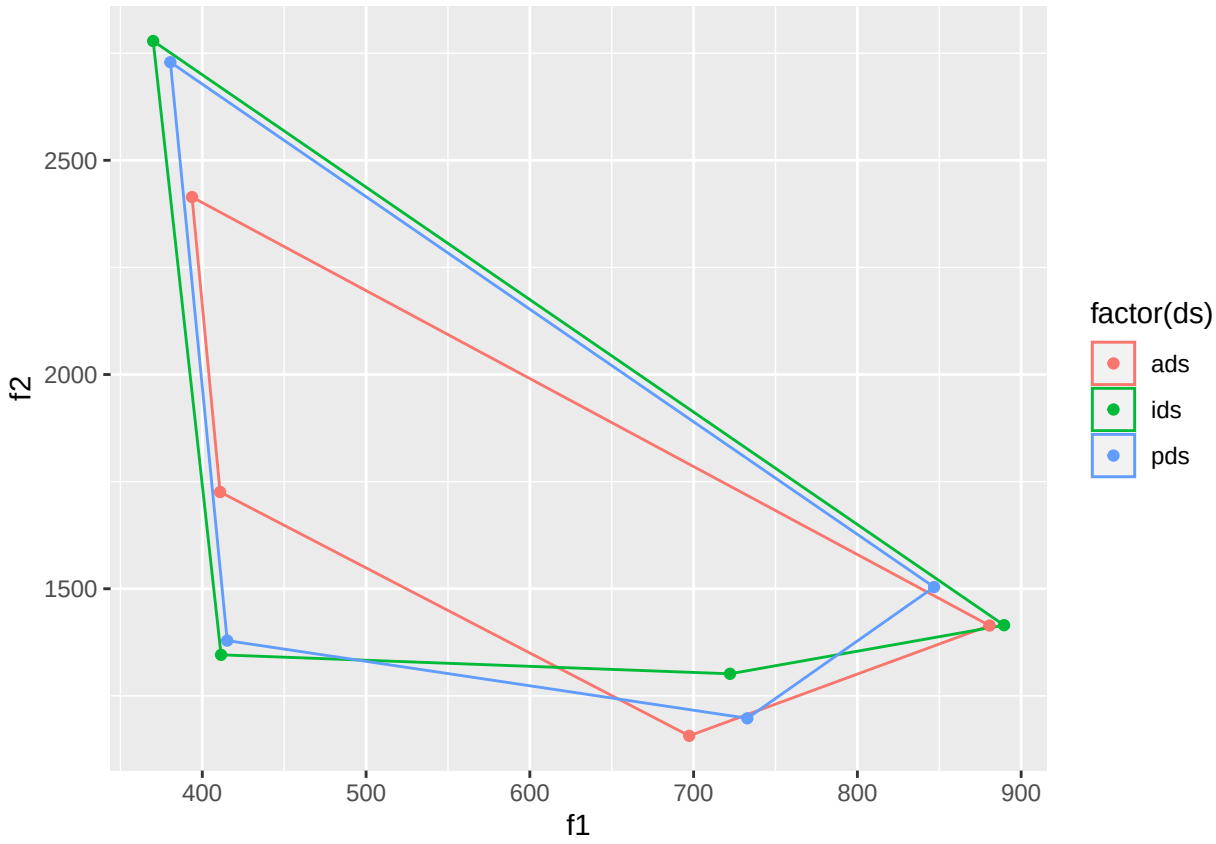
Last check before merging across groups: whether F1 and F2 differ across the two mom subgroups: They do...

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: f1 ~ ds + stim_group + (1 | motID)
##    Data: dat
##  Subset: c(v == "a")
##
## REML criterion at convergence: 2579.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.3121 -0.5389  0.0697  0.5698  2.5173
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  motID    (Intercept)   714.6   26.73
##  Residual             25722.5  160.38
## Number of obs: 201, groups:  motID, 10
##
## Fixed effects:
##               Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)    712.391     27.202  12.859  26.189 1.54e-12 ***
## dsids            6.728     27.491 190.644   0.245   0.8069
## dspds            5.861     28.712 192.482   0.204   0.8385
## stim_groupbox  154.411     28.723   5.899   5.376   0.0018 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) dsids  dspds
## dsids       -0.571
## dspds       -0.550  0.524
## stim_gropbx -0.587  0.041  0.042
```

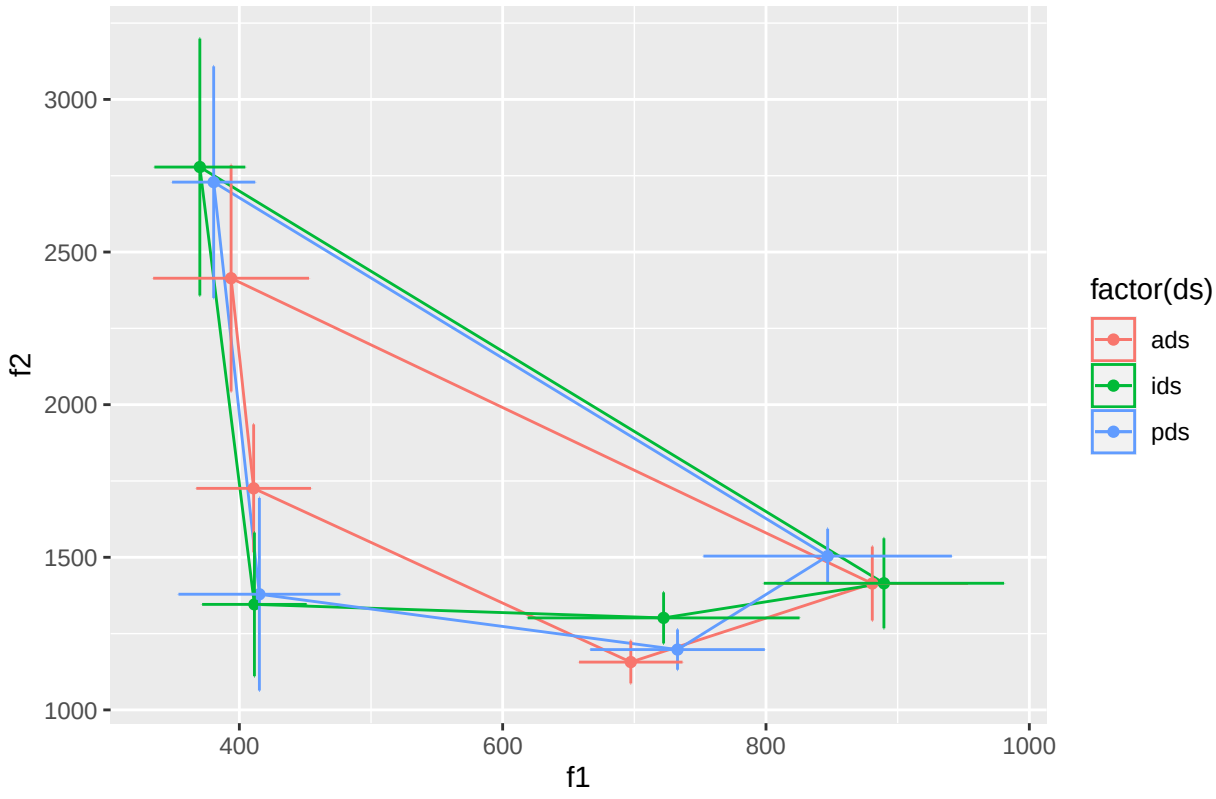
```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: f2 ~ ds + stim_group + (1 | motID)
##    Data: dat
##  Subset: c(v == "a")
##
## REML criterion at convergence: 2660.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9578 -0.5666  0.0530  0.5995  4.9663
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  motID    (Intercept)   220.9   14.86
##  Residual             39390.2  198.47
```

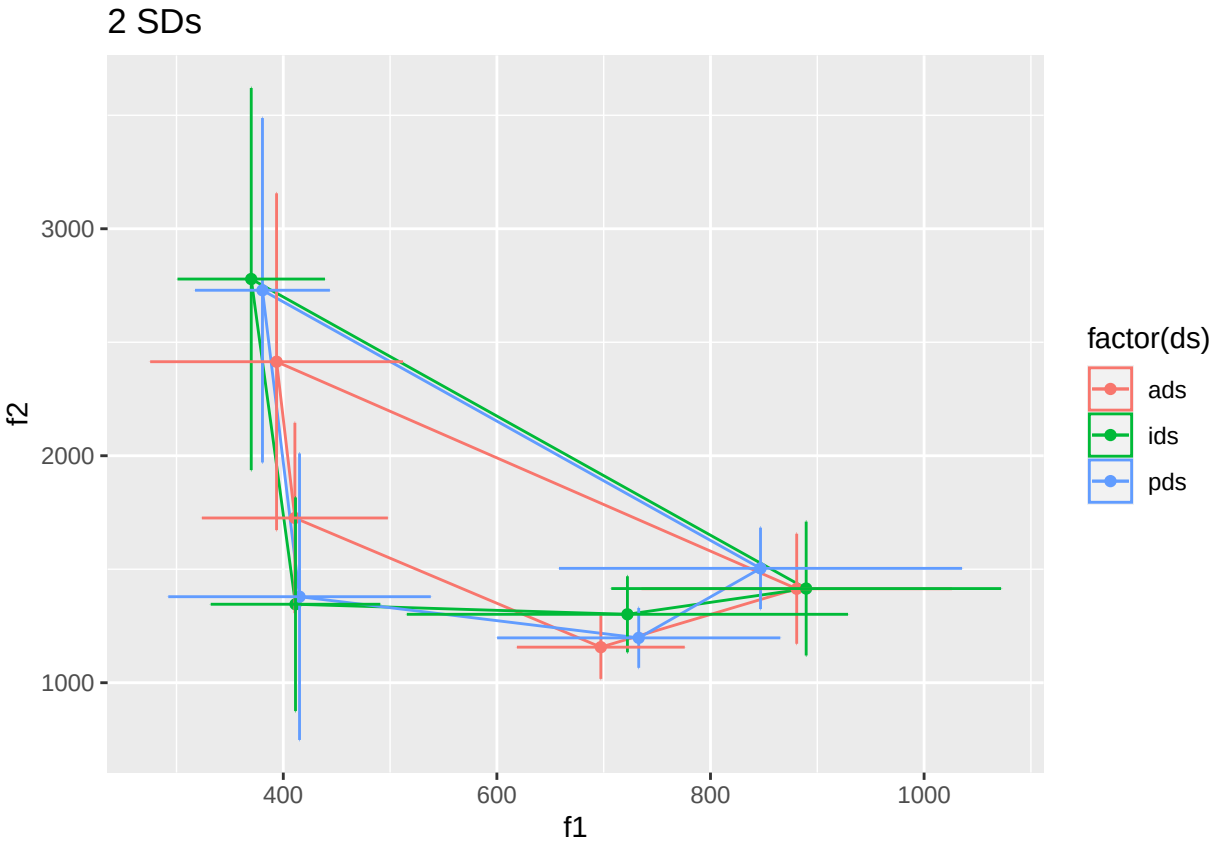
```
## Number of obs: 201, groups:  motID, 10
##
## Fixed effects:
##                Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)   1195.531     30.028   21.595  39.814  < 2e-16 ***
## dsids           54.946     33.974  193.361   1.617 0.107447
## dspds           49.934     35.443  195.098   1.409 0.160465
## stim_groupbox  207.508     29.747    8.112   6.976 0.000108 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) dsids  dspds
## dsids       -0.637
## dspds       -0.614  0.523
## stim_gropbx -0.529  0.046  0.048
```




2500
2000
1500
f2
400
500
600
700
800
900
f1
factor(ds)
ads
ids
pds

1 SDs


3000
2500
2000
1500
1000
f2
400
600
800
1000
f1
factor(ds)
ads
ids
pds



Vowel space calculations

Next we calculate vowel space for each mom and register.

ANOVA

A within-subject analysis of variance (ANOVA) on average vowel space across listener conditions (3=IDS, PDS, ADS) revealed a significant main effect ($F(2,18) = 6.04$, $p < .02$), with paired contrasts indicating that the average vowel spaces for IDS and PDS were not significantly different, $t(9) = .98$, $p > .05$, but were both greater than the vowel space of ADS (IDS v. ADS, $t(9) = 3.79$, $p < .005$) and (PDS v. ADS, $t(9) = 2.24$, $p = .05$; see Figure 1a and b).

ANOVA: not the same F or dfs

explanations from Robin: "To conduct a repeated-measures ANOVA in SPSS, we do not specify the repeated-measures factor and the dependent variable in the SPSS data file. Instead, the SPSS data file contains several quantitative variables. The number of quantitative variables is equal to the number of levels of the within-subjects factor. The scores on any one of these quantitative variables are the scores on the dependent variable for a single level of the within-subjects factor. Although we do not define the within-subjects factor in the SPSS data file, we specify it in the dialog box for the General Linear Model Repeated-Measures procedure. To define the factor, we give a name to the within-subjects factor, specify the number of levels of this factor, and indicate the quantitative variables in the data set associated with the levels of the within-subjects factor." and "the error term in the SPSS repeated measures model is (n-1) for sample (10-1=9) x (k-1) for the number of measures per sample (3-1=2); so 9 x 2 = 18 (for the denominator)."

```
## Anova Table (Type III tests)
```

```
##
## Response: vss
##                 Sum Sq Df F value    Pr(>F)
## (Intercept) 3.4849e+12  1 52.0075 9.342e-08 ***
## ds          4.8711e+11  2  3.6347   0.04001 *
## Residuals   1.8092e+12 27
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Warning: Converting "mom" to factor for ANOVA.

## Warning: Converting "ds" to factor for ANOVA.

## Tables of means
## Grand mean
##
## 766822.1
##
##  ds
## ds
##    ads    ids    pds
## 590331 886587 823548
```

Paired contrasts – same pattern of results but different t values.

```
##
##  Paired t-test
##
## data:  vs$vss[vs$ds == "pds"] and vs$vss[vs$ds == "ids"]
## t = -0.71538, df = 9, p-value = 0.4925
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -262381.7  136302.2
## sample estimates:
## mean of the differences
##               -63039.71

##
##  Paired t-test
##
## data:  vs$vss[vs$ds == "ids"] and vs$vss[vs$ds == "ads"]
## t = 2.4563, df = 9, p-value = 0.03638
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   23411.75 569100.76
## sample estimates:
## mean of the differences
##                296256.3

##
##  Paired t-test
##
## data:  vs$vss[vs$ds == "pds"] and vs$vss[vs$ds == "ads"]
## t = 1.9358, df = 9, p-value = 0.08487
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -39313.28 505746.37
```

```
## sample estimates:
## mean of the differences
##                233216.5
```

Duration and pitch can be done at the item level

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: duration ~ ds + (1 | motID)
##    Data: dat
##
## REML criterion at convergence: -1115
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8220 -0.6084 -0.1208  0.3386  6.0874
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  motID    (Intercept) 0.002933 0.05416
##  Residual             0.008659 0.09305
## Number of obs: 609, groups:  motID, 10
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept) 1.913e-01  1.853e-02 1.109e+01  10.321 5.02e-07 ***
## dsids       5.765e-02  9.422e-03 5.975e+02   6.119 1.71e-09 ***
## dspds       3.072e-02  9.561e-03 5.973e+02   3.213  0.00138 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##       (Intr) dsids
## dsids -0.281
## dspds -0.279  0.556
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Pitch ~ ds + (1 | motID)
##    Data: dat
##
## REML criterion at convergence: 6825.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8254 -0.7366 -0.1510  0.4697  8.6634
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  motID    (Intercept)  860.7   29.34
##  Residual             4984.8   70.60
## Number of obs: 601, groups:  motID, 10
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
```

```
## (Intercept)  199.642     10.733  12.788  18.601 1.23e-10 ***
## dsids         49.567      7.204 589.761   6.881 1.53e-11 ***
## dspds         52.876      7.292 589.285   7.251 1.31e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##       (Intr) dsids
## dsids -0.370
## dspds -0.368  0.555
```

Variability analyses

Get standard deviation of f1 and f2, separating by mom and register.

```
## `summarise()` has grouped output by 'motID', 'ds'. You can override using the `.groups` argument.
```

The following mixed model shows register effects on standard deviation of F1: highest for PDS, intermediate for IDS, lowest for ADS.

```
## boundary (singular) fit: see ?isSingular
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: f1.sd ~ ds + v + (1 | motID)
##    Data: sds
##
## REML criterion at convergence: 925.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.5694 -0.5696 -0.1199  0.4101  5.8709
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  motID    (Intercept)    0      0.00
##  Residual             2596     50.95
## Number of obs: 90, groups:  motID, 10
##
## Fixed effects:
##             Estimate Std. Error     df t value Pr(>|t|)
## (Intercept)   113.36      12.01  85.00   9.440 6.94e-15 ***
## dsids          25.26      13.15  85.00   1.920   0.0582 .
## dspds          55.54      13.15  85.00   4.222 6.04e-05 ***
## vi            -79.52      13.15  85.00  -6.045 3.86e-08 ***
## vu            -86.71      13.15  85.00  -6.591 3.48e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##       (Intr) dsids  dspds  vi
## dsids -0.548
## dspds -0.548  0.500
## vi    -0.548  0.000  0.000
## vu    -0.548  0.000  0.000  0.500
```

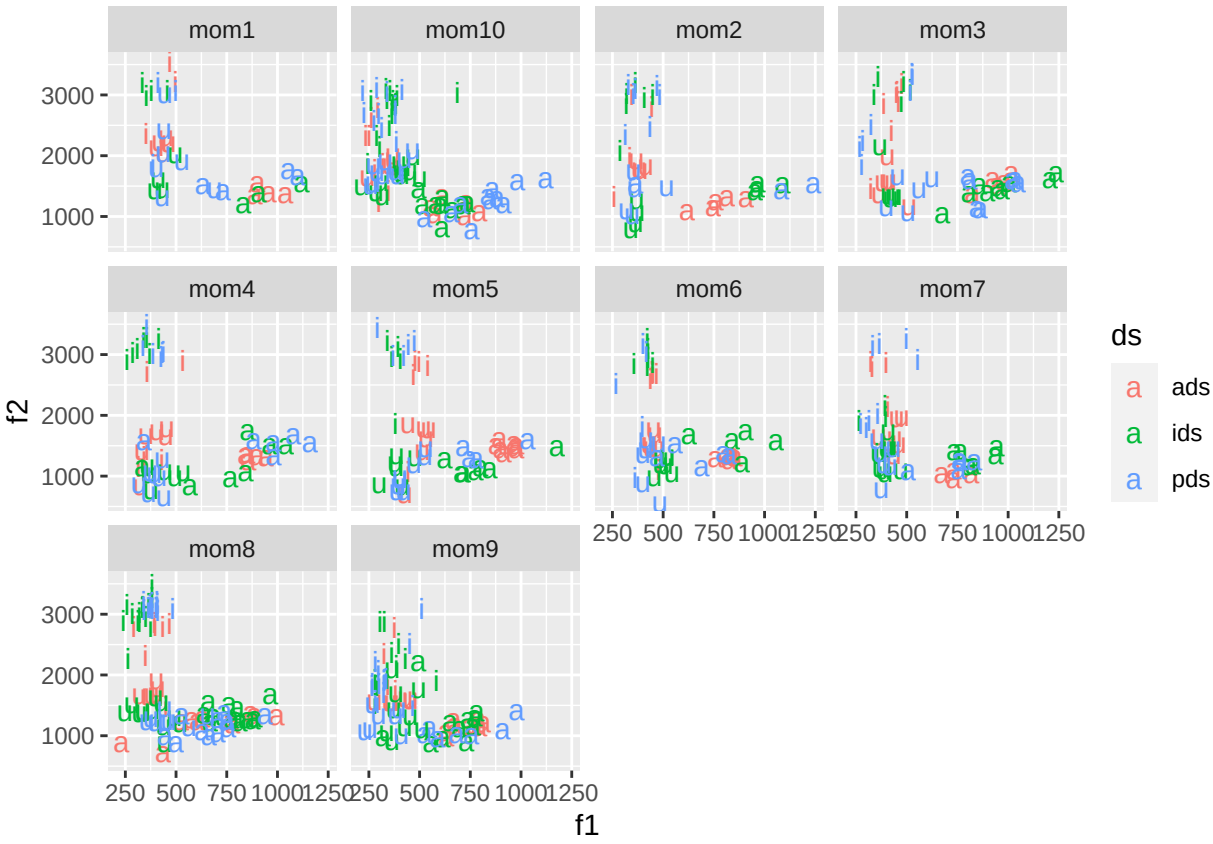
```
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

No such effect for SD of F2.

```
## boundary (singular) fit: see ?isSingular
```

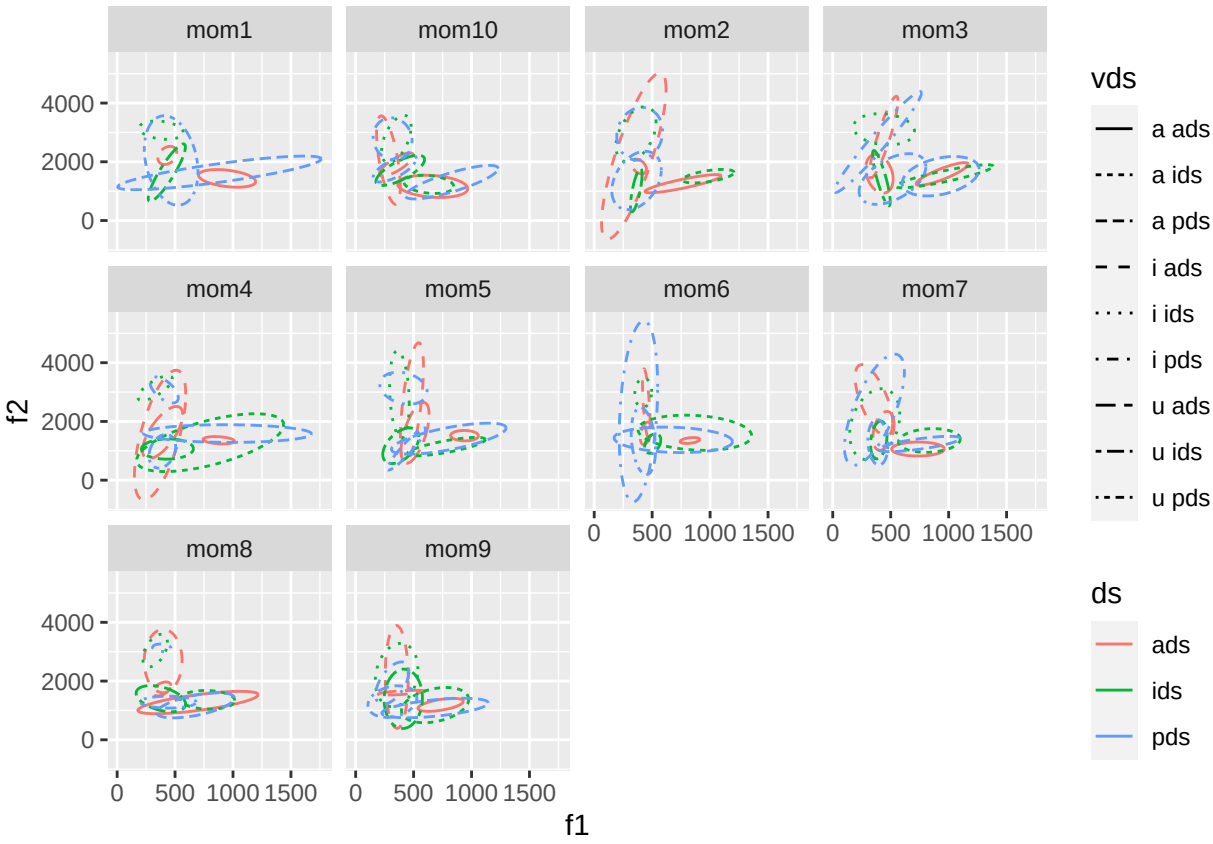
```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: f2.sd ~ ds + v + (1 | motID)
##    Data: sds
##
## REML criterion at convergence: 1120.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.56967 -0.58443 -0.02158  0.49246  2.82118
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  motID    (Intercept)     0      0.0
##  Residual             25807    160.6
## Number of obs: 90, groups:  motID, 10
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  167.804     37.865  85.000   4.432 2.77e-05 ***
## dsids        -57.018     41.479  85.000  -1.375   0.1729
## dspds         -8.436     41.479  85.000  -0.203   0.8393
## vi           303.774     41.479  85.000   7.324 1.27e-10 ***
## vu            88.246     41.479  85.000   2.127   0.0363 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##       (Intr) dsids  dspds  vi
## dsids -0.548
## dspds -0.548  0.500
## vi    -0.548  0.000  0.000
## vu    -0.548  0.000  0.000  0.500
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

We cannot really see variability easily in any of these graphs...




```
## Too few points to calculate an ellipse
## Too few points to calculate an ellipse
## Too few points to calculate an ellipse
## Too few points to calculate an ellipse
```

```
## Warning: Removed 3 row(s) containing missing values (geom_path).
```


mom1
mom10
mom2
mom3
mom4
mom5
mom6
mom7
mom8
mom9
f2
f1
4000
2000
0
0
500
1000
1500
vds
a ads
a ids
a pds
i ads
i ids
i pds
u ads
u ids
u pds
ds
ads
ids
pds