24.963 Linguistic Phonetics
Fall 2005

The effect of stop voicing on vowel formants

Terminology:
Following Kwong and Stevens, I refer to the latter portion of a tense vowel or diphthong as an 'offglide'. The variable names for F1 and F2 at the offset of a vowel are **f1offglide** and **f2offglide**.

Kwong and Stevens predict different effects of voicing depending on the nature of the offglide:
High front offglide /i, eɪ, aɪ/: F1 lower and F2 higher before voiceless stop.
High back offglide /oʊ, u, aʊ/: F1 and F2 lower before voiceless stop.
No high offglide (lax or low) /ɪ, ɛ, æ/: No difference in F1, F2 before voiced and voiceless stops.

In the analyses below, these three classes of vowels are referred to in terms of their offglide: 'front', 'back' or 'none'.

Means and standard deviations (in parentheses):

F2 offglide (Hz)

offglide	voiced	voiceless
front	2342 (213)	2607 (145)
back	1695 (166)	1450 (210)
none	2035 (160)	2028 (176)

F1 offglide (Hz)

offglide	voiced	voiceless
front	359 (58)	353 (66)
back	370 (67)	404 (57)
none	473 (113)	648 (184)

I leave it to you to look at means for individual pairs – this may well be instructive.
I also leave it to you to look at the steady states of the vowels.

Some complications:

- I have one report that a speaker read 'bowed' as homophonous with 'bode' rather than as [baʊd] one one occasion. Judging from measurements, this may have happened with other speakers. Since this leaves the vowel in the same offglide class (back), I haven't made any corrections in the analysis below, but this probably interferes with some of the interactions, and would certainly complicate analysis of the steady states.
- Final /t/s were probably glottalized in many cases. It would be useful to know what you observed with your speaker. This may have affected formant measurements – if glottal closure precedes the oral closure, then formants will show less effect of the oral closure at vowel offset. This might account for some very high F1 offsets before voiceless stops (check your spectrograms to see of this is plausible). More importantly, voiced stops are not glottalized, so apparent effects of stop voicing might be due to glottalization rather than being related to sustaining vocal fold vibration. Bear this in mind in interpreting the results.

- A number of people have noted that diphthongs like /aɪ/ often contain two stationary points, e.g. a maximum in F1 early in the diphthong and a maximum in F2 later, with F2 falling into the consonant. Where people submitted formant measurements at both points, I've only included the former in the spreadsheet on the grounds that this represents the 'nucleus' as opposed to the 'offglide'.

ANOVAs

Since Kwong and Stevens predict different effects of voicing for the three classes of vowels, I have carried out separate ANOVAs for each class of offglide (front, back, none). For each class F1 and F2 offset are analyzed separately, making for 6 ANOVAs in all.

Each analysis is a repeated measures ANOVA. I've given the full results from Stata for each analysis below. The first analysis is accompanied by comments that should provide enough information for you to extract the relevant information from the remaining analyses.

First formant offset

<u>F1 offset (f1offglide) for vowels with front offglides /i, eɪ, aɪ/</u>
This is the stata command, giving the model specification. The factors are Subject, Voicing (voiced or voiceless), and Pair (one value for each minimal pair)

```
. anova f1offglide subject voicing/subject*voicing pair/subject*pair
voicing*pair/subject*voicing*pair if glidestr=="front", repeated(voicing pair)
```

This is the summary results table:

```
                          Number of obs =      30     R-squared     =  1.0000
                          Root MSE      =       0     Adj R-squared =

               Source |  Partial SS    df       MS           F     Prob > F
----------------------+----------------------------------------------------
                Model |  108273.467    29  3733.56782
                      |
              subject |  70051.1333     4  17512.7833      15.36     0.0108
              voicing |       235.2     1       235.2       0.21     0.6732
     subject*voicing |  4560.46667     4  1140.11667
----------------------+----------------------------------------------------
                 pair |  18298.0667     2  9149.03333      10.49     0.0058
         subject*pair |  6977.26667     8  872.158333
----------------------+----------------------------------------------------
        voicing*pair |      4308.2     2      2154.1       4.48     0.0494
subject*voicing*pair |  3843.13333     8  480.391667
----------------------+----------------------------------------------------
                      |
             Residual |           0     0
----------------------+----------------------------------------------------
                Total |  108273.467    29  3733.56782
```

In this table, each effect is followed on the next row by the 'error term' used to estimate the within groups variance for that F-test. That is, the interaction subject*voicing is the error term in the F-test of voicing.

We're mainly interested in the significance of 'voicing', but if there is a significant interaction with Pair, then it's necessary to examine the nature of the interaction – is it just the magnitude of the effect that varies between pairs, or is the effect reversed in some cases?

The probabilities in the table above are not the final word on effects other than Voicing because they do not correct for violations of the assumption of sphericity (equal variances for the difference scores for each pair of conditions – see Max and Onghena 1999 (week 5) for more information). A standard approach to sphericity violations is to adjust the degrees of freedom of the F ratio. One method is the Huynh-Feldt correction, given in the detailed reports on repeated variables (below). The Huynh-Feldt epsilon is given above each table. This a factor by which the degrees of freedom are multiplied before calculating the probability of the calculated *F*-ratio. The corrected *p* value is given in the table under H-F.

For example, Voicing has only two levels, so a sphericity violation is not possible, so the Huynh-Feldt epsilon is 1, and the *p* value under the H-F column is 0.6732, as in the table above. However, the H-F corrected *p* value for the Voicing*Pair interaction is 0.0692, whereas the uncorrected *p* value is < 0.05. To report an H-F corrected *F*-test, you need to calculate the corrected degrees of freedom by multiplying the listed degrees of freedom by the H-F epsilon, e.g. $F(1.53, 6.11) = 4.48$, $p = 0.07$ (H-F epsilon = 0.7633, 2×0.7633 = 1.53, 8×0.7633 = 6.11).

```
Between-subjects error term:  subject
                     Levels:  5         (4 df)
     Lowest b.s.e. variable:  subject

Repeated variable: voicing
                                          Huynh-Feldt epsilon        =  1.0000
                                          Greenhouse-Geisser epsilon =  1.0000
                                          Box's conservative epsilon =  1.0000

                                            ------------ Prob > F ------------
                  Source |     df      F    Regular    H-F      G-G      Box
    ---------------------+--------------------------------------------------
                 voicing |      1     0.21   0.6732   0.6732   0.6732   0.6732
          subject*voicing |      4
    ---------------------+--------------------------------------------------

Repeated variable: pair
                                          Huynh-Feldt epsilon        =  1.1717
                                          *Huynh-Feldt epsilon reset to 1.0000
                                          Greenhouse-Geisser epsilon =  0.7744
Box's conservative epsilon =  0.5000

                                            ------------ Prob > F ------------
                  Source |     df      F    Regular    H-F      G-G      Box
    ---------------------+--------------------------------------------------
                    pair |      2    10.49   0.0058   0.0058   0.0124   0.0317
            subject*pair |      8
    ---------------------+--------------------------------------------------
```

```
Repeated variables: voicing*pair
                                        Huynh-Feldt epsilon        =  0.7633
                                        Greenhouse-Geisser epsilon =  0.6210
                                        Box's conservative epsilon =  0.5000

                                          ------------ Prob > F ------------
                Source |     df      F    Regular    H-F      G-G      Box
    -------------------+-----------------------------------------------------
           voicing*pair |      2     4.48   0.0494   0.0692   0.0850   0.1016
    subject*voicing*pair |      8
    -------------------+-----------------------------------------------------
```

F1 offset (f1offglide) for vowels with back offglides /oʊ, u, aʊ/

```
. anova f1offglide subject voicing/subject*voicing pair/subject*pair
voicing*pair/subject*voicing*pair if glidestr=="back", repeated(voicing pair)

                           Number of obs =      30     R-squared     =  1.0000
                           Root MSE      =       0     Adj R-squared =

                  Source |  Partial SS    df       MS           F     Prob > F
    -------------------+-----------------------------------------------------
                   Model |  117693.467    29   4058.3954
                         |
                 subject |  48082.4667     4  12020.6167      21.82     0.0056
                 voicing |  8806.53333     1  8806.53333      15.99     0.0161
         subject*voicing |  2203.13333     4  550.783333
    -------------------+-----------------------------------------------------
                    pair |  35796.0667     2  17898.0333      22.95     0.0005
            subject*pair |  6238.93333     8  779.866667
    -------------------+-----------------------------------------------------
           voicing*pair |  3954.06667     2  1977.03333       1.25     0.3359
    subject*voicing*pair |  12612.2667     8  1576.53333
    -------------------+-----------------------------------------------------
                         |
                Residual |           0     0
    -------------------+-----------------------------------------------------
                   Total |  117693.467    29   4058.3954


Between-subjects error term:  subject
                     Levels:  5         (4 df)
     Lowest b.s.e. variable:  subject

Repeated variable: voicing
                                        Huynh-Feldt epsilon        =  1.0000
                                        Greenhouse-Geisser epsilon =  1.0000
                                        Box's conservative epsilon =  1.0000

                                          ------------ Prob > F ------------
                Source |     df      F    Regular    H-F      G-G      Box
    -------------------+-----------------------------------------------------
                voicing |      1    15.99   0.0161   0.0161   0.0161   0.0161
         subject*voicing |      4
    -------------------+-----------------------------------------------------
```

```
Repeated variable: pair
                                     Huynh-Feldt epsilon        =  1.8015
                                     *Huynh-Feldt epsilon reset to 1.0000
                                     Greenhouse-Geisser epsilon =  0.9538
                                     Box's conservative epsilon =  0.5000

                                       ------------ Prob > F ------------
              Source |     df      F    Regular    H-F      G-G      Box
    -----------------+------------------------------------------------------
                pair |      2    22.95   0.0005   0.0005   0.0006   0.0087
        subject*pair |      8
    -----------------+------------------------------------------------------

Repeated variables: voicing*pair
                                     Huynh-Feldt epsilon        =  1.0987
                                     *Huynh-Feldt epsilon reset to 1.0000
                                     Greenhouse-Geisser epsilon =  0.7496
                                     Box's conservative epsilon =  0.5000

                                       ------------ Prob > F ------------
              Source |     df      F    Regular    H-F      G-G      Box
    -----------------+------------------------------------------------------
        voicing*pair |      2     1.25   0.3359   0.3359   0.3336   0.3255
    subject*voicing*pair |      8
    -----------------+------------------------------------------------------
```

F1 offset (f1offglide) for vowels with no offglides /ɪ, ɛ, æ/

```
. anova f1offglide subject voicing/subject*voicing pair/subject*pair
voicing*pair/subject*voicing*pair if glidestr=="none", repeated(voicing pair)

                          Number of obs =      30     R-squared     =  1.0000
                          Root MSE      =       0     Adj R-squared =

              Source |  Partial SS    df       MS           F     Prob > F
    -----------------+----------------------------------------------------
               Model |    884378.8    29  30495.8207
                     |
             subject |    180694.8     4     45173.7       4.82     0.0785
             voicing |  228813.333     1  228813.333      24.39     0.0078
     subject*voicing |  37525.3333     4  9381.33333
    -----------------+----------------------------------------------------
                pair |    333637.8     2    166818.9      30.55     0.0002
        subject*pair |     43680.2     8    5460.025
    -----------------+----------------------------------------------------
        voicing*pair |  28534.8667     2  14267.4333       3.62     0.0758
    subject*voicing*pair |  31492.4667     8  3936.55833
    -----------------+----------------------------------------------------
                     |
            Residual |           0     0
    -----------------+----------------------------------------------------
               Total |    884378.8    29  30495.8207
```

```
Between-subjects error term:  subject
                     Levels:  5         (4 df)
     Lowest b.s.e. variable:  subject

Repeated variable: voicing
                                          Huynh-Feldt epsilon        =  1.0000
                                          Greenhouse-Geisser epsilon =  1.0000
                                          Box's conservative epsilon =  1.0000

                                            ------------ Prob > F ------------
                Source |     df      F    Regular    H-F      G-G      Box
    -------------------+-----------------------------------------------------
               voicing |      1    24.39   0.0078   0.0078   0.0078   0.0078
     subject*voicing |      4
    -------------------+-----------------------------------------------------

Repeated variable: pair
                                          Huynh-Feldt epsilon        =  0.6247
                                          Greenhouse-Geisser epsilon =  0.5598
                                          Box's conservative epsilon =  0.5000

                                            ------------ Prob > F ------------
                Source |     df      F    Regular    H-F      G-G      Box
    -------------------+-----------------------------------------------------
                  pair |      2    30.55   0.0002   0.0022   0.0035   0.0052
         subject*pair |      8
    -------------------+-----------------------------------------------------

Repeated variables: voicing*pair
                                          Huynh-Feldt epsilon        =  0.9471
                                          Greenhouse-Geisser epsilon =  0.6946
                                          Box's conservative epsilon =  0.5000

                                            ------------ Prob > F ------------
                Source |     df      F    Regular    H-F      G-G      Box
    -------------------+-----------------------------------------------------
         voicing*pair |      2     3.62   0.0758   0.0801   0.1049   0.1297
    subject*voicing*pair |      8
    -------------------+-----------------------------------------------------
```

Second formant offset

F2 offset (f2offglide) for vowels with front offglides /i, eɪ, aɪ/

```
. anova f2offglide subject voicing/subject*voicing pair/subject*pair
voicing*pair/subject*voicing*pair if glidestr=="front", repeated(voicing pair)

                          Number of obs =      30     R-squared     =  1.0000
                          Root MSE      =       0     Adj R-squared =

                 Source |  Partial SS    df       MS           F     Prob > F
    --------------------+-----------------------------------------------------
                  Model |  1456841.87    29  50235.9264
                        |
                subject |  559051.867     4  139762.967      10.30     0.0221
                voicing |  528013.333     1  528013.333      38.91     0.0034
         subject*voicing |  54280.6667     4  13570.1667
    --------------------+-----------------------------------------------------
                   pair |  168531.267     2  84265.6333       8.54     0.0103
           subject*pair |  78915.7333     8  9864.46667
    --------------------+-----------------------------------------------------
           voicing*pair |  26939.2667     2  13469.6333       2.62     0.1332
    subject*voicing*pair |  41109.7333     8  5138.71667
    --------------------+-----------------------------------------------------
                        |
                Residual |           0     0
    --------------------+-----------------------------------------------------
                  Total |  1456841.87    29  50235.9264


Between-subjects error term:  subject
                     Levels:  5         (4 df)
      Lowest b.s.e. variable:  subject

Repeated variable: voicing
                                          Huynh-Feldt epsilon        =  1.0000
                                          Greenhouse-Geisser epsilon =  1.0000
                                          Box's conservative epsilon =  1.0000

                                            ------------ Prob > F ------------
                 Source |     df      F    Regular    H-F      G-G       Box
    --------------------+-----------------------------------------------------
                voicing |      1    38.91   0.0034   0.0034   0.0034   0.0034
         subject*voicing |      4
    --------------------+-----------------------------------------------------

Repeated variable: pair
                                          Huynh-Feldt epsilon        =  1.3778
                                          *Huynh-Feldt epsilon reset to 1.0000
                                          Greenhouse-Geisser epsilon =  0.8395
                                          Box's conservative epsilon =  0.5000

                                            ------------ Prob > F ------------
                 Source |     df      F    Regular    H-F      G-G       Box
    --------------------+-----------------------------------------------------
                   pair |      2     8.54   0.0103   0.0103   0.0162   0.0431
           subject*pair |      8
    --------------------+-----------------------------------------------------
```

```
Repeated variables: voicing*pair
                                        Huynh-Feldt epsilon        =  0.6538
                                        Greenhouse-Geisser epsilon =  0.5732
                                        Box's conservative epsilon =  0.5000

                                          ------------ Prob > F ------------
                Source |     df      F    Regular    H-F      G-G      Box
    -------------------+-----------------------------------------------------
           voicing*pair |      2     2.62   0.1332   0.1647   0.1730   0.1808
    subject*voicing*pair |      8
    -------------------+-----------------------------------------------------
```

F2 offset (f2offglide) for vowels with back offglides /oʊ, u, aʊ/

```
. anova f2offglide subject voicing/subject*voicing pair/subject*pair
voicing*pair/subject*voicing*pair if glidestr=="back", repeated(voicing pair)

                           Number of obs =      30     R-squared     =  1.0000
                           Root MSE      =       0     Adj R-squared =

                  Source |  Partial SS    df       MS           F     Prob > F
    ---------------------+------------------------------------------------------
                   Model |  1449615.87    29   49986.754
                         |
                 subject |    292770.2     4    73192.55      12.81     0.0150
                 voicing |  450432.533     1  450432.533      78.81     0.0009
         subject*voicing |  22861.1333     4  5715.28333
    ---------------------+------------------------------------------------------
                    pair |  377628.067     2  188814.033       7.09     0.0169
            subject*pair |    213169.6     8     26646.2
    ---------------------+------------------------------------------------------
           voicing*pair |  20231.6667     2  10115.8333       1.12     0.3737
    subject*voicing*pair |  72522.6667     8  9065.33333
    ---------------------+------------------------------------------------------
                         |
                Residual |           0     0
    ---------------------+------------------------------------------------------
                   Total |  1449615.87    29   49986.754


Between-subjects error term:  subject
                     Levels:  5         (4 df)
     Lowest b.s.e. variable:  subject

Repeated variable: voicing
                                        Huynh-Feldt epsilon        =  1.0000
                                        Greenhouse-Geisser epsilon =  1.0000
                                        Box's conservative epsilon =  1.0000

                                          ------------ Prob > F ------------
                Source |     df      F    Regular    H-F      G-G      Box
    -------------------+-----------------------------------------------------
                voicing |      1    78.81   0.0009   0.0009   0.0009   0.0009
         subject*voicing |      4
    -------------------+-----------------------------------------------------
```

```
Repeated variable: pair
                                        Huynh-Feldt epsilon        =  0.7239
                                        Greenhouse-Geisser epsilon =  0.6042
                                        Box's conservative epsilon =  0.5000

                                          ------------ Prob > F ------------
                Source |     df      F    Regular    H-F      G-G      Box
    -------------------+-----------------------------------------------------
                  pair |      2     7.09   0.0169   0.0326   0.0436   0.0563
           subject*pair |      8
    -------------------+-----------------------------------------------------

Repeated variables: voicing*pair
                                        Huynh-Feldt epsilon        =  0.7387
                                        Greenhouse-Geisser epsilon =  0.6105
                                        Box's conservative epsilon =  0.5000

                                          ------------ Prob > F ------------
                Source |     df      F    Regular    H-F      G-G      Box
    -------------------+-----------------------------------------------------
           voicing*pair |      2     1.12   0.3737   0.3651   0.3584   0.3504
    subject*voicing*pair |      8
    -------------------+-----------------------------------------------------
```

F1 offset (f1offglide) for vowels with no offglides /ɪ, ɛ, æ/

```
. anova f2offglide subject voicing/subject*voicing pair/subject*pair
voicing*pair/subject*voicing*pair if glidestr=="none", repeated(voicing pair)

                          Number of obs =      30     R-squared     =  1.0000
                          Root MSE      =       0     Adj R-squared =

                Source |  Partial SS    df       MS           F     Prob > F
    -------------------+-----------------------------------------------------
                 Model |  792138.167    29  27315.1092
                       |
               subject |  292710.333     4  73177.5833       8.49     0.0310
               voicing |  353.633333     1  353.633333       0.04     0.8494
            subject*voicing |  34492.8667     4  8623.21667
    -------------------+-----------------------------------------------------
                  pair |  430223.267     2  215111.633     198.46     0.0000
           subject*pair |  8671.06667     8  1083.88333
    -------------------+-----------------------------------------------------
           voicing*pair |  5258.06667     2  2629.03333       1.03     0.4001
    subject*voicing*pair |  20428.9333     8  2553.61667
    -------------------+-----------------------------------------------------
                       |
                Residual |           0     0
    -------------------+-----------------------------------------------------
                 Total |  792138.167    29  27315.1092


Between-subjects error term:  subject
                     Levels:  5         (4 df)
      Lowest b.s.e. variable:  subject
```

Repeated variable: voicing

```
                                        Huynh-Feldt epsilon        =  1.0000
                                        Greenhouse-Geisser epsilon =  1.0000
                                        Box's conservative epsilon =  1.0000

                                          ------------ Prob > F ------------
              Source |     df      F    Regular    H-F      G-G      Box
    -----------------+-----------------------------------------------------
             voicing |      1     0.04   0.8494   0.8494   0.8494   0.8494
     subject*voicing |      4
    -----------------+-----------------------------------------------------
```

Repeated variable: pair

```
                                        Huynh-Feldt epsilon        =  1.3177
                                        *Huynh-Feldt epsilon reset to 1.0000
                                        Greenhouse-Geisser epsilon =  0.8213
Box's conservative epsilon =  0.5000

                                          ------------ Prob > F ------------
              Source |     df      F    Regular    H-F      G-G      Box
    -----------------+-----------------------------------------------------
                pair |      2   198.46   0.0000   0.0000   0.0000   0.0001
        subject*pair |      8
    -----------------+-----------------------------------------------------
```

Repeated variables: voicing*pair

```
                                        Huynh-Feldt epsilon        =  0.5886
                                        Greenhouse-Geisser epsilon =  0.5430
                                        Box's conservative epsilon =  0.5000

                                          ------------ Prob > F ------------
              Source |     df      F    Regular    H-F      G-G      Box
    -----------------+-----------------------------------------------------
        voicing*pair |      2     1.03   0.4001   0.3762   0.3720   0.3677
    subject*voicing*pair |      8
    -----------------+-----------------------------------------------------
```