Here we provide a worked example of one of the subsets of the data for which we provide summaries in Tables 2 and 3, namely the older Scilly-educated males producing mouth vowels.

In the linear mixed-effect models, reference levels for estimates are chosen simply by alphabetical order, so for natural class, “obs” (= obstruent) is the reference level with an estimate for “son” (= sonorant), and for voicing, “h” (= voiceless) is the reference level with an estimate for “v” (= voiced).

A linear mixed-effect model was fitted by maximum likelihood, with natural class as a fixed effect and random intercept and slope for natural class per speaker.

Natural class appears in the model with ß = .12094, SE = .01553, *t* = 7.79.

Our second model included voicing as an additional fixed effect, with a view to testing the improvement in goodness of fit of the model when voicing is added over and above natural class as a predictor. In this second model, natural class has ß = .05186, SE = .01788, *t* = 2.90, and voicing has ß = .08985, SE = .01672, *t* = 5.37.

The second model (which includes voicing) has a significant improvement in goodness of fit over the first model (which excludes voicing), as demonstrated by a log-likelihood comparison between the two models: χ2(1) = 27.381, *p* < .0001. This *p*-value is the value reported under “LME model” in Table 2.

A simple conditional inference decision tree using only voicing to predict maximum F1 is shown in Figure A1. The p-value for the split on voicing is the value reported under “Decision tree” in Table 2.

A more complex tree that also includes natural class, stress, manner, whether the rhyme was open or closed, and utterance position (initial, final, or medial) is shown in Figure A2. Note that the absence of some predictors from the tree is evidence that they have no significant effect on maximum F1. The fact that there is a significant split on voicing is represented in Table 2 by the asterisk under “Decision tree.”

We report here on the result of conditional variable importance modeling from a random forest analysis with a small set of predictors (voicing, natural class, and duration). Note that the absolute values of this measure are not comparable across different models. We stepped up the number of trees in each forest (modeling five forests at a time and starting with 500 trees per forest) until the model became robust, where we define robust as having the same ordering of variables across the five forests. Each tree in the forest had two randomly selected variables. This model became robust with 1000 trees in each forest. The mean importance values averaged across those five forests are as follows, with duration being more important than voicing, and voicing being more important than natural class:

<2CD>

Duration 5.253 × 10−4
Voicing 3.809 × 10−4
Natural class 3.605 × 10−4
</2CD>

In a model with a larger set of predictors containing more potential confounds, the predictors were voicing, natural class, the duration of the vowel, the stress status of the syllable, the manner of articulation of the following consonant, whether the rhyme was open or closed, and the position of the syllable in the utterance. For this subset of the data, we stopped increasing the number of trees in each random forest on reaching 8000. Manner was clearly the most important variable in predicting F1 in the vowel. Next came voicing, natural class, and duration (it proved impossible to differentiate conclusively between the relative importance of these three factors in this model), with stress being less important and the remaining two predictors (rhyme and utterance position) close to zero importance. The following are mean importance values averaged across five forests, each of 8000 trees:

<2CD>

Manner 8.974 × 10−4
Voicing 3.864 × 10−4
Duration 3.841 × 10−4
Natural class 3.815 × 10−4
Stress 1.086 × 10−4
Rhyme 8.653 × 10−6
Utterance position 5.350 × 10−6
</2CD>

Although voicing is not the most important variable in predicting F1 in the vowel, the purpose of our tests was to see whether voicing had a role to play; in all these models, that is shown to be the case.

figure a1. A simple conditional inference decision tree using only voicing to predict maximum F1 for older Scilly-educated males producing mouth vowels.

figure a2. A more complex decision tree using voicing, natural class, stress, manner, whether the rhyme was open or closed and utterance position (initial, final, or medial) to predict maximum F1 for older Scilly-educated males producing mouth vowels.