Supplemental Material S1. Detailed explanation of statistical models for Research Questions 1–4.

Research Question 1

The model analyzed the number of correct productions by person *i* on each word *j* as a binomial distribution governed by two parameters, n_j, the number of consonants in word *j*, and p_{ij}, the proportion of consonants produced correctly by person *i* on word *j*. The parameter p_{ij}, which is bounded by 0 and 1, is then subjected to a log-odds transformation, which places it on an open-ended scale, and modeled as a function of a linear combination of the predictor variables. In this model, we coded the cluster vs. singleton contrast as a dummy variable with singletons as the reference category. With this formulation, the model intercept represents the log odds of correct singleton production, and the coefficient for the contrast variable represents the change in the log odds of correct production associated with clusters.

The model permitted intercepts to vary across both persons and words, and also permitted the slope of the cluster vs. singleton contrast to vary across persons. The by-subject intercepts and slopes were permitted to correlate with each other. The means, variances, and covariances of the person-level intercepts and slopes were estimated separately for the two groups. The mean of the word-level intercepts was fixed to zero in both groups in order to identify the model, and the variances were estimated separately for the two groups.

Non- or minimally informative priors were used for all parameters. Following the Birats example in the OpenBUGS user manual, the subject-varying intercepts and slopes (representing performance on singletons and the cluster vs. singleton contrast, respectively) were modeled as a multivariate normal distribution with independent univariate normal priors (mean = 0, precision = 10-6) on their means. OpenBUGS parameterizes the variance of a normal distribution by its inverse, precision. Accordingly, the prior for the variance-covariance matrix of the regression coefficients was specified as a precision matrix and given a Wishart prior with two degrees of freedom, equal to the rank of the precision matrix, and a scale matrix representing prior beliefs about the order of magnitude of the elements of the covariance matrix. The word-varying intercepts were modeled as a normal distribution with a mean fixed to zero and a gamma prior (shape = 0.1, rate = 0.1) on the estimated precision. Attempts to use less informative priors (e.g., gamma (0.001, 0.001) on the precision or a uniform prior on the standard deviation) resulted in poor mixing of the MCMC chains for these parameters. All of these parameters and their priors were defined separately for group A and group P.

The model was estimated in OpenBUGS 3.2.2. Two Markov chain Monte Carlo (MCMC) chains were run with 5000 iterations discarded as burn-in and an additional 50,000 iterations from each chain saved for parameter estimation. Overdispersed starting values determined from preliminary estimates were used. Non- or minimally informative priors were used for all parameters, with details provided in the supplementary materials.

Convergence and Model Fit

MCMC convergence was evaluated using trace and autocorrelation plots produced by OpenBUGS and the Gelman-Rubin potential scale reduction factor (PSRF) computed using the R package coda. The plots suggested adequate convergence and mixing of chains for all parameters, the PSRF upper 95% CI bound was one for all parameters, and the multivariate PSRF was 1.01. Posterior parameter distributions were summarized by their means and 95% credible intervals (CIs). The mean (95%CI) of the standardized residuals was -0.001

(-0.06, 0.06), and the standard deviation was 1.06 (1.01, 1.14). The fact that the standard deviation was credibly larger than one indicates that the residuals were mildly overdispersed relative to the expectations of the binomial distribution, and this was confirmed by comparing the sum of squares of the standardized residuals to a chi-square distribution with degrees of freedom equal to the number of observations minus the number of estimated parameters ($\chi^2 = 1201$, df = 1060, $p < 0.001$). Attempts to add an observation-level error term to account for this overdispersion were unsuccessful, resulting in poor mixing and convergence of the MCMC chains. As an alternative, we computed a correction factor equal to the square root of the sum of squares of the standardized residuals divided by the degrees of freedom (Gelman & Hill, 2007). The value of this correction factor was 1.06 and we multiplied the standard deviations of the posterior distributions (analogous to the standard errors of frequentist parameter estimates) by it to adjust for the overdispersion. The CIs reported in the manuscript are based on the adjusted values for the predicted error rates.

Research Question 2

This model was identical to the model for Research Question 1, except that there were two subject-varying slopes (for the medial vs. initial and final vs. medial contrasts). Accordingly, the Wishart prior on the precision matrix for these regression coefficients had three degrees of freedom. The estimation procedures were also identical, with the exception that additional MCMC iterations were required to achieve satisfactory convergence and mixing. Following burn-in, 150,000 iterations were run for each chain and the chains were thinned by a factor of 3 to reduce computer memory demands, saving 50,000 iterations from each chain for parameter estimation.

Convergence and Model Fit

There was adequate convergence and mixing of chains for all parameters, with PSRFs of 1 for all parameters and a multivariate PSRF of 1.01. As with question 1, the standardized residuals were overdispersed (sd = 1.09, 95%CI: 1.02, 1.2), the chi-square test was significant ($\chi2 = 1923$, df = 1597, $p < 0.001$), and a model with an observation-level error term did not converge. As above, a correction factor (1.10) was computed and used to adjust the posterior standard deviations and the credible intervals reported in the manuscript.

Research Question 3

This model was identical to the model for Question 2, except that the precision parameter for the word-varying intercepts was given a gamma (0.001, 0.001) prior, and it included a trial-level error term to account for overdispersion of the residuals. Adequate convergence and mixing were obtained with 50,000 post-burn-in iterations of each of the two MCMC chains.

Convergence and Model Fit

Trace and autocorrelation plots suggested adequate convergence and mixing of chains for all parameters, and the PSRF 95%CI upper bounds were ≤ 1.01 for all parameters and the multivariate PSRF was 1.01. The standard deviation of the residuals obtained a point estimate of 1.02 with a 95%CI (0.94, 1.16) that included zero and the sum of the squared standardized residuals was credibly drawn from a chi-square distribution with the appropriate degrees of freedom ($\chi2 = 638$, df = 587, $p = 0.07$), suggesting that the observation-level error term adequately accounted for overdispersion in the residuals.

Research Question 4a

The analysis model for question 4a differed in a few key aspects from the previous models. First, the model was simplified by removing the covariances between the subject-varying model intercept, which was coded to represent performance on stop consonants and the varying dummy-coded contrasts for affricates, fricatives, liquids, and nasals were removed. This was necessary because a model including these covariances failed to converge to stable posterior distributions with adequate mixing of the MCMC chains. We suspect that there were insufficient numbers of observations of the individual manners of articulation to support estimation of the covariances. The same issue also required the variances of the contrast effects to be estimated with a single common parameter and a minimally informative prior distribution. As with the models for questions 1 and 2, we were unable to include an observation-level error term to correct for overdispersion. The means of the intercepts and slopes were given normal priors with mean zero and precisions of 1E-6. The intercept variance parameter was given a gamma (0.001, 0.001) prior. The precision of the slopes was modeled with a single parameter with a gamma (0.1, 0.1) prior. The word-varying intercept parameter was given a gamma (0.1, 0.1) prior. With these modifications, 50,000 post-burn-in MCMC iterations were adequate for convergence.

Convergence and Model Fit

Trace and autocorrelation plots suggested adequate convergence and mixing of chains for all parameters, and the PSRF 95%CI upper bounds were ≤ 1.01 for all parameters and the multivariate PSRF was 1.01. The standard deviation of the standardized residuals was 1.03 with a 95%CI (0.97, 1.12), that included zero. However, comparison of the squared standardized residuals to the chi-square distribution ($\chi 2 = 2229$, df = 2084, $p = 0.006$) suggested mild overdisperson and a correction factor of 1.04 was used to adjust the posterior standard deviations and credible intervals.

Research Question 4b

In this model, the subject-varying intercepts were modeled as above in Question 4a. The most frequently observed place of articulation, alveolar, was coded as a subject-varying intercept. The contrasts for the remaining places of articulation, bilabial, labiodental, dental, palatal, and velar, were estimated as constant across subjects with non-informative normal distribution priors on their means. The mean of each slope parameter was given a normal prior with mean zero and precision 1E-6. It was also necessary in this model to estimate the variance of the varying intercept for the word stimuli as a single common parameter rather than as a separate parameter for the two groups. The MCMC chains were run for 50,0000 post-burn-in iterations.

Results

Trace and autocorrelation plots suggested adequate convergence and mixing of chains for all parameters, and the PSRF 95%CI upper bounds were ≤ 1.01 for all parameters and the multivariate PSRF was 1.01. The residuals were overdispersed (sd = 1.09, 95%CI: 1.03, 1.18; $\chi 2 = 2119$, df = 1785, $p < 0.001$), and the posterior standard deviations were corrected by a factor of 1.09.

Research Question 4c

This model was identical to the model for Research Question 1. The voiced vs. voiceless contrast was coded with performance on voiced consonants as the reference category and the model intercept. The MCMC chains were run for 50,0000 post-burn-in iterations.

Results

Trace and autocorrelation plots suggested adequate convergence and mixing of chains for all parameters, and the PSRF 95%CI upper bounds were 1.00 for all parameters and the multivariate PSRF was 1.01. Both the standard deviation of the residuals (1.09, 95%CI: 1.03, 1.19) and comparison of the sum of squared residuals to the chi-square distribution ($\chi 2 = 1403$, $df = 1165$, $p > 0.001$) suggested overdispersion, and a correction factor of 1.1 was used.