Bayesian meta-analysis of the voicing effect in English

Stefano Coretta

A Bayesian meta-analysis of the English voicing effect has been run on the basis of 11 estimated posterior distributions extracted from 9 different publications, following the procedures discussed in Nicenboim et al. (2018). The studies were selected by scraping the first 100 results on Google Scholar with the keywords "vowel duration voicing English". Other studies which were known to the author but not present among the Google Scholar results were also included. Since two publications (Sharf 1962 and Klatt 1973) tested both monosyllabic and disyllabic words, two separate posterior distributions were estimated for each word type. This leads to a total of 11 posterior distribution of the effect of consonant voicing on vowel duration in English (7 estimated posteriors from 7 publications plus 2 each from 2 publications).

The posterior distributions of each study have been obtained by fitting a Bayesian linear model to the summary data (the means of vowel duration before voiceless and voiced stops) provided by the respective publications. These models had the mean vowel durations as outcome and consonant voicing (voiceless vs. voiced) as the only predictor. Three studies, Luce & Charles-Luce (1985), Davis & Van Summers (1989), and Ko (2018), reported measures of dispersion along with the means. Measurement error models were used to obtain the posterior distributions from these studies. The measurement error term in such models allows fitting to include information of the dispersion of the mean vowel durations, and hence of the uncertainty that comes with them. All the models for estimating the posterior of the individual studies were fitted with the following priors: a normal distribution with mean = 0 ms and SD = 300 for the intercept, and a normal distribution with mean = 0 ms and SD = 100 for the effect of consonant voicing. The simple models (without an error term) also included a prior for the residual variance as a half Cauchy distribution with location = 0 ms and scale = 25.

A data set with the mean estimates and estimated standard errors from these 11 posterior distributions has then been used to fit a further Bayesian measurement error model. In this model, the mean estimates with the estimated standard errors were included as the outcome, while a by-study random intercept was the only predictor. The models were fitted in R with brms using Markov Chain Monte Carlo simulations, with 4 chains, 2000 iterations of which 1000 for warm-up.

The following is the summary of the meta-analytical model (as output by `summary()` function). The population-level effects are the ones of interest. Figure 1 is a visual aid to the summury, and shows a variety of credible intervals of the estimates from the model. The blue-coloured bars represent (from darker to lighter blue) the 50%, 80%, and 95% credible intervals (CIs). THe black lines are the 66% (thick) and 98% (thin) CIs.

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: Estimate | se(Est.Error) ~ syl_pos + (1 | study)
##    Data: estimated_tbl (Number of observations: 15)
## Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
##          total post-warmup samples = 4000
##
## Group-Level Effects:
```
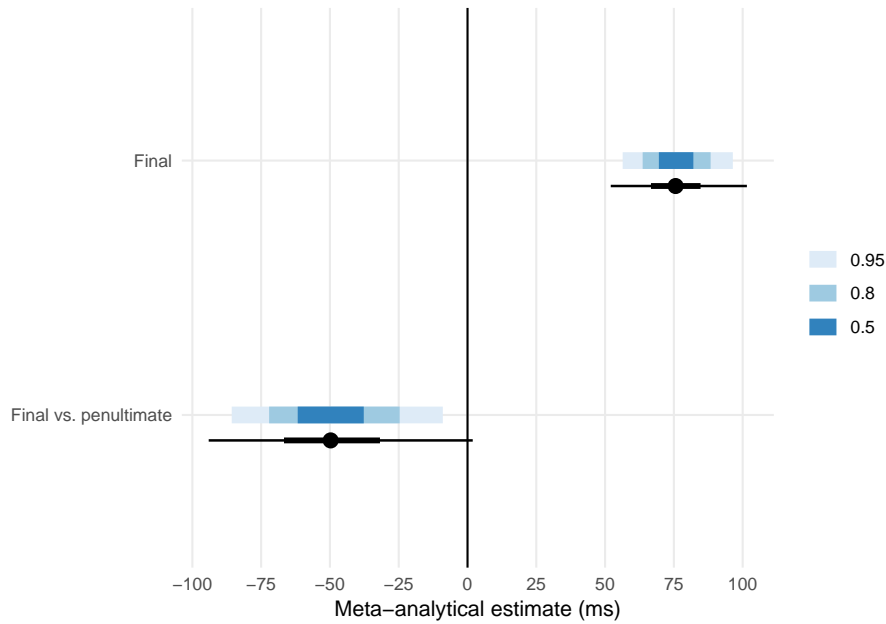
Figure 1: Credible intervals of the meta-analytical posterior distributions.

```
## ~study (Number of levels: 15)
##               Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat
## sd(Intercept)    23.36      8.88     9.80    44.34       1181 1.00
##
## Population-Level Effects:
##                  Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat
## Intercept           75.83     10.01    56.39    96.43       1542 1.00
## syl_posnonMfinal   -49.14     19.10   -85.69    -8.98       1629 1.00
##
## Samples were drawn using sampling(NUTS). For each parameter, Eff.Sample
## is a crude measure of effective sample size, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

The 95% credible interval (CI) of the model intercept (which corresponds to the estimated voicing effect in word-final syllables) is between 56.39 and 96.43 ms. This means that there is a 95% probability that the true effect lies between about 56 and 96 ms. The mean of the posterior distribution is 75.83 ms (SD = 10.01). Given the 95% CI of the meta-analytical posterior distribution, it can be inferred that the true effect of voicing in word-final syllables in English is positive and between 50 and 100 ms. However, note that the meta-anlytical estimate might suffer from publication bias (cf. below).

The posterior mean of the coefficient when the target syllable is in penultimate position is -49.14 ms (SD = 19.10, 95% CI = [-85.69, -8.98]). Note that the estimated error is double compared to that of the intercept, which means the there is greater uncertainty in this than the other estimate. We can argue that, on average, the mean voicing effect in penultimate syllables is about 50 ms smaller than the mean effect in monosyllabic words in the surveyed studies. The mean of the voicing effect in disyllabic words can thus be estimated to be around 25 ms (about 75 - 50 ms).

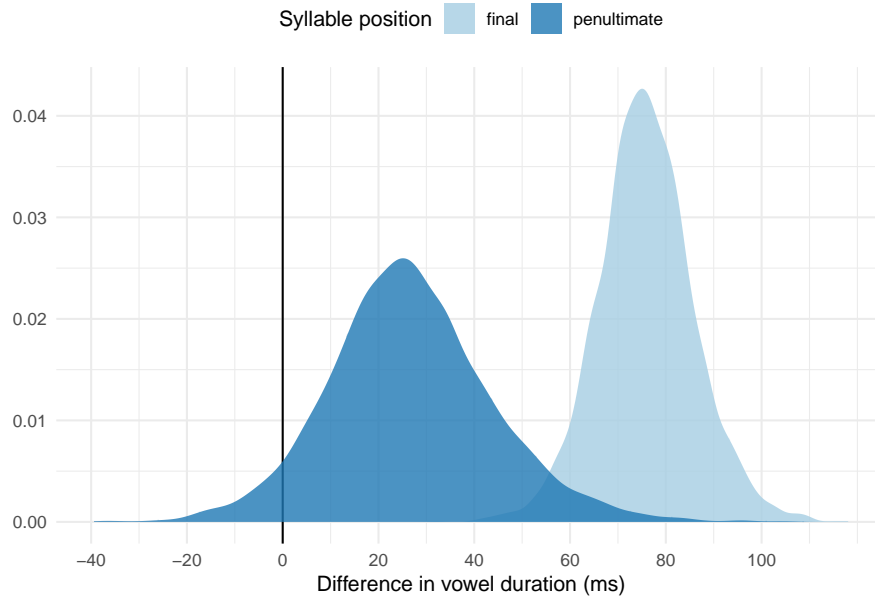A visual representation of the meta-analytical distributions is given in Figure 2. The plot shows the

Figure 2: Meta-analytical posterior distributions of the voicing effect in syllable-final and penultimate position.

full posterior distributions of the voicing effect in the word-final and penultimate contexts. Note how the posterior distribution in penultimate position is wider than the other.

Figure 3 shows the mean estimates (the points) of the voicing effect with 95% CIs (the horizontal segments) for each of the 11 studies. For each study, the plot gives both the original estimate (as obtained from the raw data summary of the study) and the estimate shrunk by the random effects in the meta-analytical model. The vertical lines indicate the meta-analytical 95% CI of the voicing effect in final (solid) and penultimate syllable position (dashed). Original estimates further away from the meta-analytical mean effect and those with greater uncertainty (wider errors) show greater shrinkage to the mean.

Figure 4 is a funnel plot, which can be used to visually check whether the sample suffers from publication bias. In this plot, the x-axis corresponds to the original estimated difference in vowel duration, while the y-axis is a measure of precision (calculated as 1 divided by the estimated error of the difference). The meta-analytical means are indicated by the thick and dashed vertical lines for syllable-final and penultimate postion respectively. The shaded areas indicate the 95% CI of the meta-analytical posterior of the voicing effect in final (light blue) and penultimate (dark blue) position. When there is no bias, the points with lower precision should be more spread out and symmetrically placed around the meta-analytical mean, while points with higher precision should cluster around the mean. This ideal situation is clearly not the case for the final syllable context. There seems to be a bias towards bigger effects (which also happen to have lower precision). This indicates that the estimate probably suffers from publication bias (i.e. bias towards publicating positive and significant results) and it is not representative of the true effect. It is not possible to assess bias with the effect in penultimate syllable position given the low number of studies.

The shaded areas indicate the 95% CI of the meta–analytical posterior
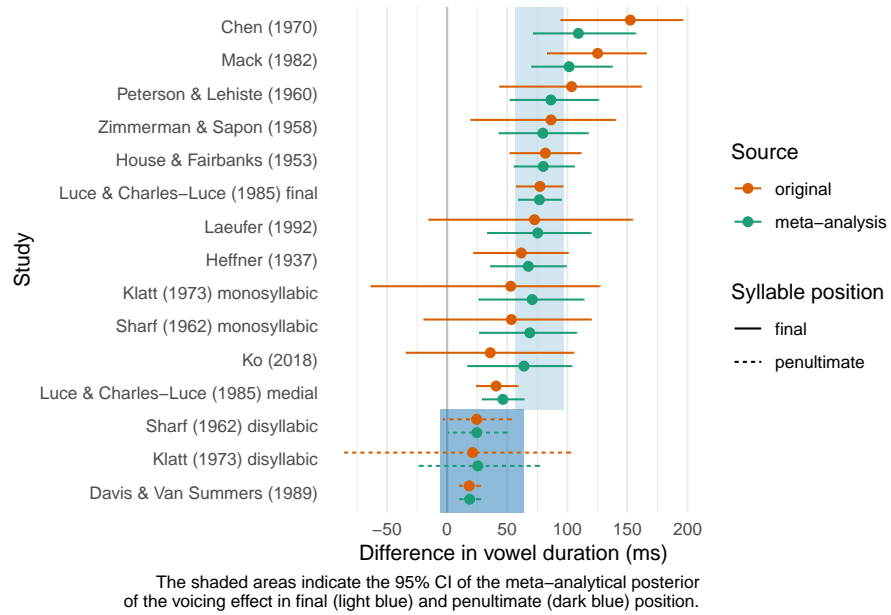of the voicing effect in final (light blue) and penultimate (dark blue) position.

Figure 3: Estimated voicing effect from the original source and from the meta-analysis.
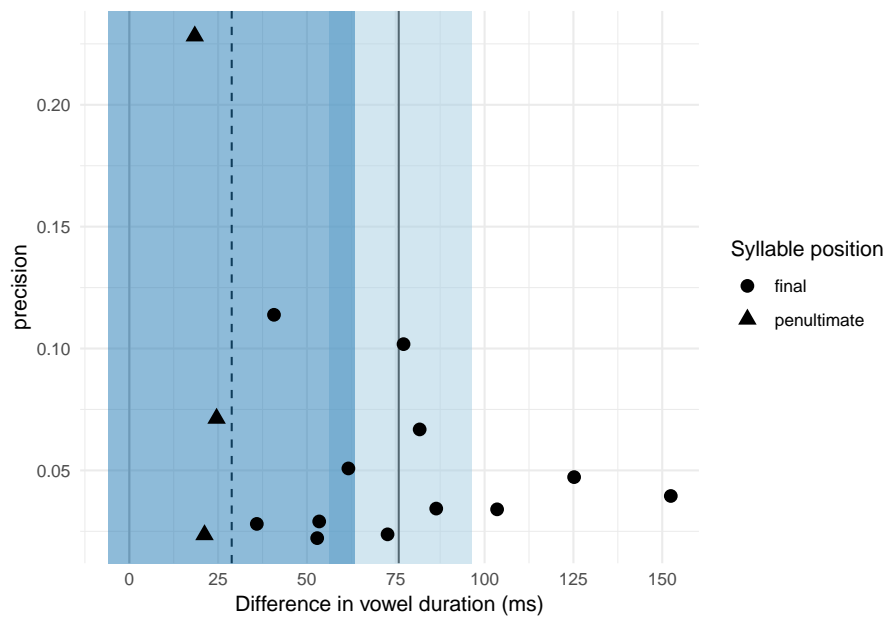

Figure 4: By-study funnel plot showing the estimate against the precision. The vertical thick and dashed lines are the meta-analytical means of the effect in final and penultimate position.

References

Davis, Stuart & W. Van Summers. 1989. Vowel length and closure duration in word-medial VC sequences. *Journal of Phonetics* 17. 339–353.

Klatt, Dennis H. 1973. Interaction between two factors that influence vowel duration. *The Journal of the Acoustical Society of America* 54(4). 1102–1104. doi:10.1121/1.1914322.

Ko, Eon-Suk. 2018. Asymmetric effects of speaking rate on the vowel/consonant ratio conditioned by coda voicing in English. *Phonetics and Speech Sciences* 10(2). 45–50. doi:10.13064/KSSS.2018.10.2.045.

Luce, Paul A & Jan Charles-Luce. 1985. Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *The Journal of the Acoustical Society of America* 78(6). 1949–1957.

Nicenboim, Bruno, Timo B. Roettger & Shravan Vasishth. 2018. Using meta-analysis for evidence synthesis: The case of incomplete neutralization in german. *Journal of Phonetics* 70. 39–55. doi:10.1016/j.wocn.2018.06.001.

Sharf, Donald J. 1962. Duration of post-stress intervocalic stops and preceding vowels. *Language and speech* 5(1). 26–30.