

PraatSauce/VoiceSauce comparison: White Hmong (female)

James Kirby

31/01/2018

Contents

Data	1
Plots	2
F0	2
Formants	3
Bandwidths	5
Uncorrected amplitudes	9
Corrected amplitudes	11
Corrected differences	21
Cepstral peak prominence	25
Harmonic to noise ratios	26
Distinguishing voice qualities	27
High vowels	27
Mid vowels	29
Low vowels	31

Data

This small dataset compares spectral measures generated by both PraatSauce v0.2.2 and VoiceSauce v1.31 at 1 msec intervals for 9 White Hmong lexical items spoken by a single female speaker. The original audio files can be found [here](#). For both scripts, 5 formants were estimated with a maximum formant frequency of 5500 Hz; minimum and maximum F0 values were set to 50 Hz and 600 Hz for all F0 estimators. For VoiceSauce, the STRAIGHT F0 estimate and Snack formant/bandwidth estimates were used for harmonic amplitude corrections.

The `method` column indicates whether the formant bandwidths were `estimated` using Praat (PraatSauce) or Snack (VoiceSauce), or whether the Hawks and Miller `formula` was used.

In Hmong orthography, final *-g* indicates a low-falling breathy tone, while *-m* indicates creaky tone.

```
head(df)
```

```
##      Filename Item Label seg_Start seg_End   t_ms      t method
## 1 25e-cab-w_Audio cab    a    585.504 868.719 585.504 0.000000000 formula
## 2 25e-cab-w_Audio cab    a    585.504 868.719 586.504 0.003546099 formula
## 3 25e-cab-w_Audio cab    a    585.504 868.719 587.504 0.007092199 formula
## 4 25e-cab-w_Audio cab    a    585.504 868.719 588.504 0.010638298 formula
## 5 25e-cab-w_Audio cab    a    585.504 868.719 589.504 0.014184397 formula
## 6 25e-cab-w_Audio cab    a    585.504 868.719 590.504 0.017730496 formula
##      script measure  value corrected
## 1 PraatSauce      pF0 247.514 uncorrected
## 2 PraatSauce      pF0 247.966 uncorrected
## 3 PraatSauce      pF0 248.418 uncorrected
## 4 PraatSauce      pF0 248.870 uncorrected
```

```
## 5 PraatSauce      pF0 249.322 uncorrected
## 6 PraatSauce      pF0 249.774 uncorrected
```

In the plots which follow, the PraatSauce measures are unsmoothed. If you want to compare to smoothed estimates, uncomment the two lines:

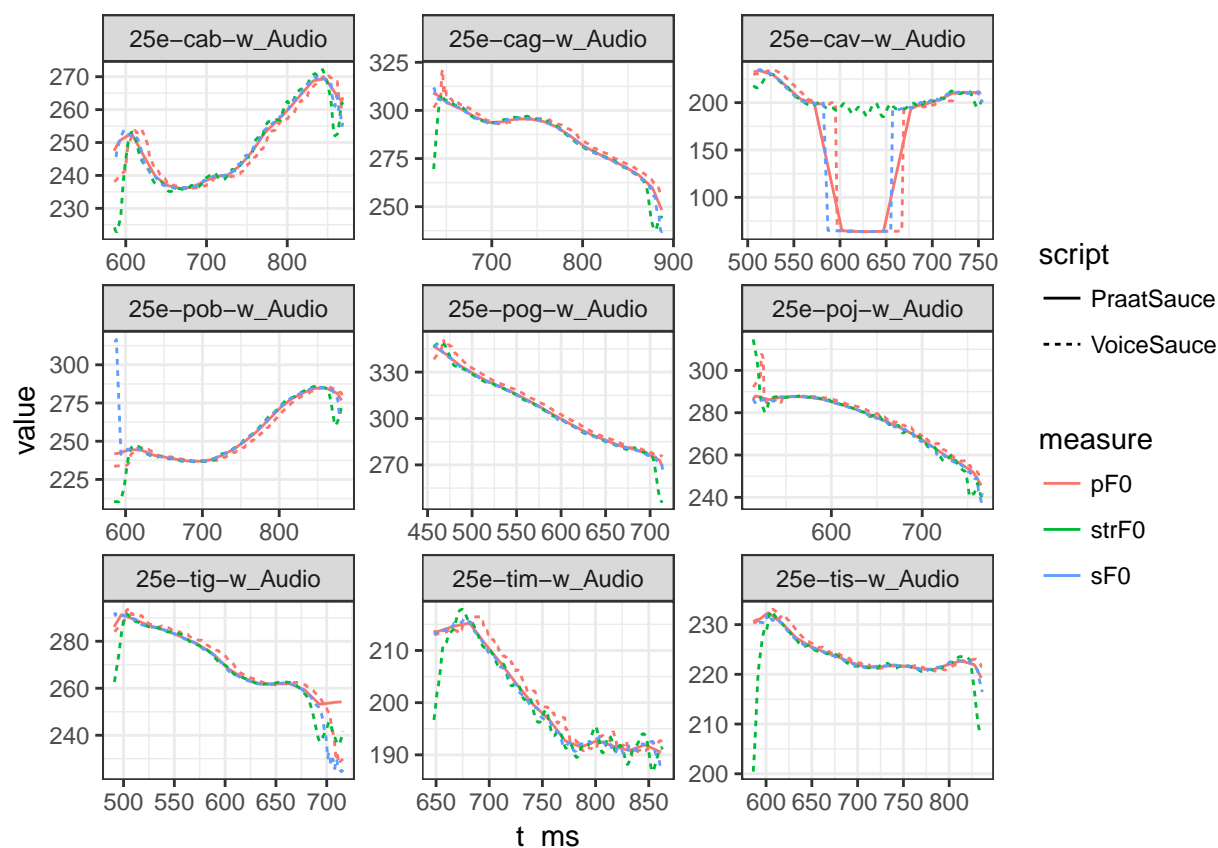
```
ps.fbw <- cbind(ps.fbw[1:6], apply(ps.fbw[7:43], 2, filter, filter=f21, sides=2))
ps.ebw <- cbind(ps.ebw[1:6], apply(ps.ebw[7:43], 2, filter, filter=f21, sides=2))
```

This implements a symmetric kernel filter. This is different from what VoiceSauce does. VoiceSauce uses the Matlab `filter()` function, by default a lag filter which pads with zeros. So while the smoothed value of sample 20 is equal to $\sum_{i=1}^{20} / 20$, the smoothed value of sample 19 is not undefined, but is calculated as $\sum_{i=1}^{19} / 20$, the smoothed value of sample 18 is $\sum_{i=1}^{18} / 20$, etc.

If you want to smooth the Matlab way, use the lag kernel by selecting `filter=f20` and set `sides=1`.

Plots

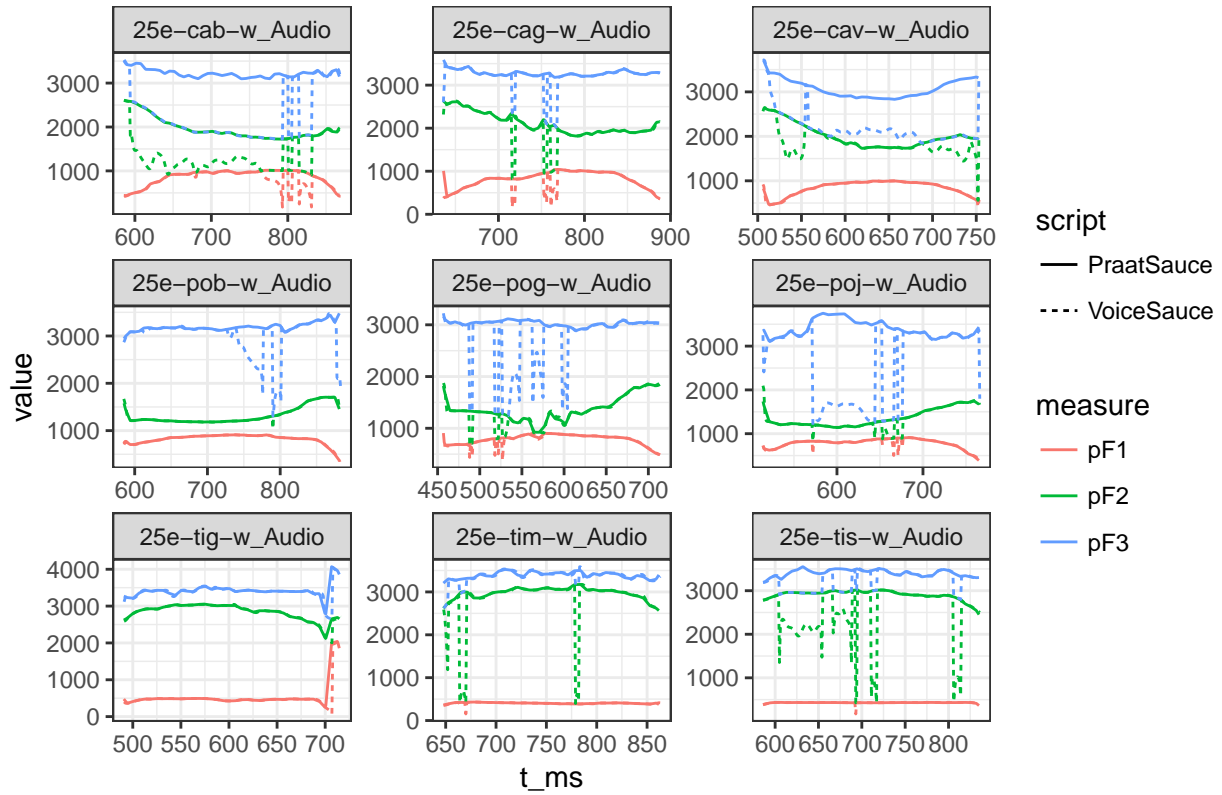
F0



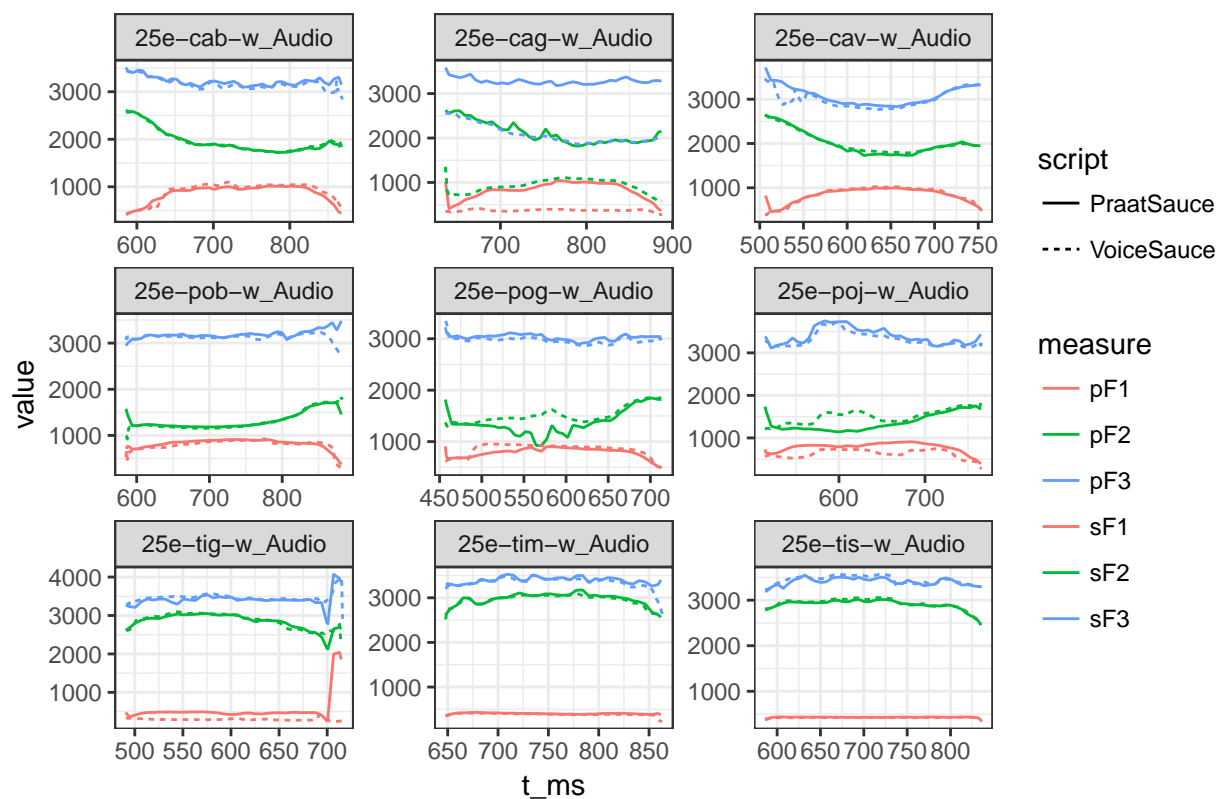
All F0 estimators except for STRAIGHT have difficulty with the somewhat constricted vowel quality of *cav* 'to argue'.

Formants

Formants estimate using Praat

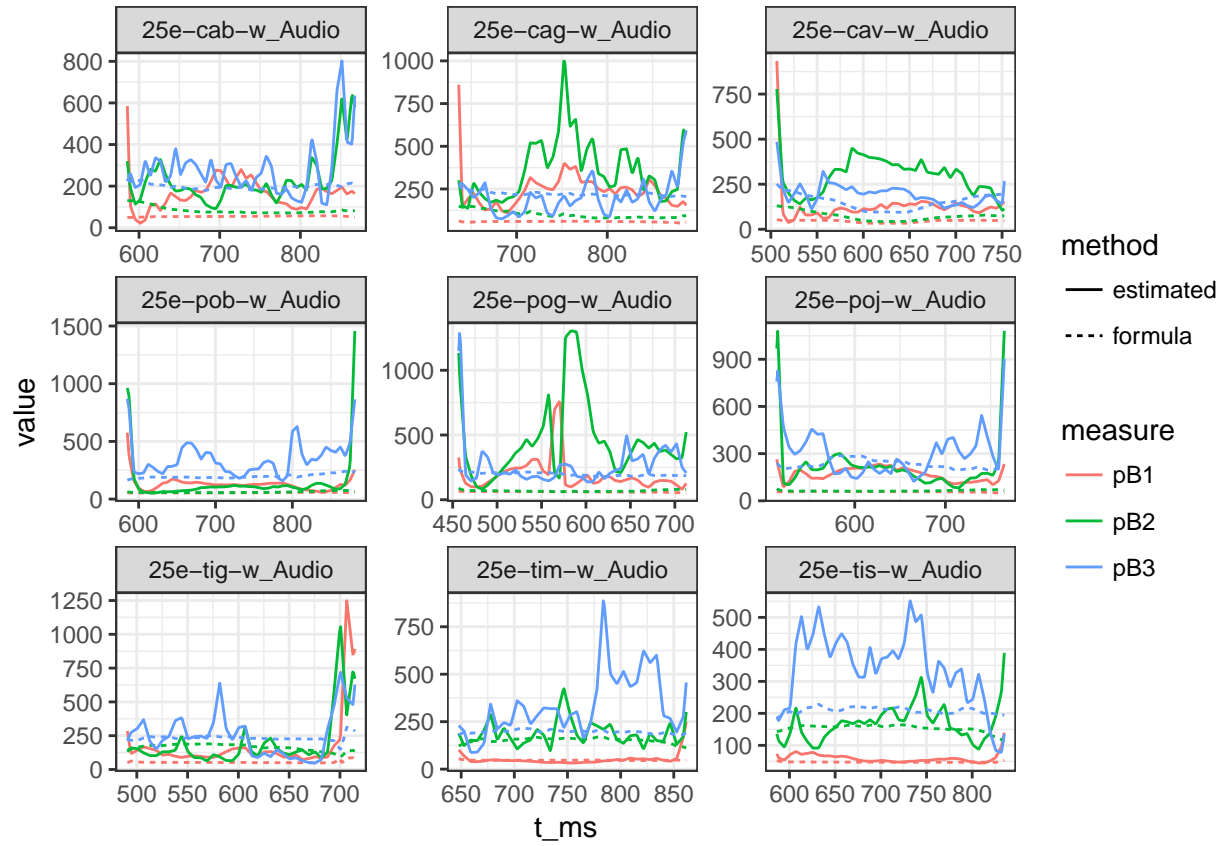


Formants estimated using Snack (VoiceSauce) vs. Praat (PraatSauce)



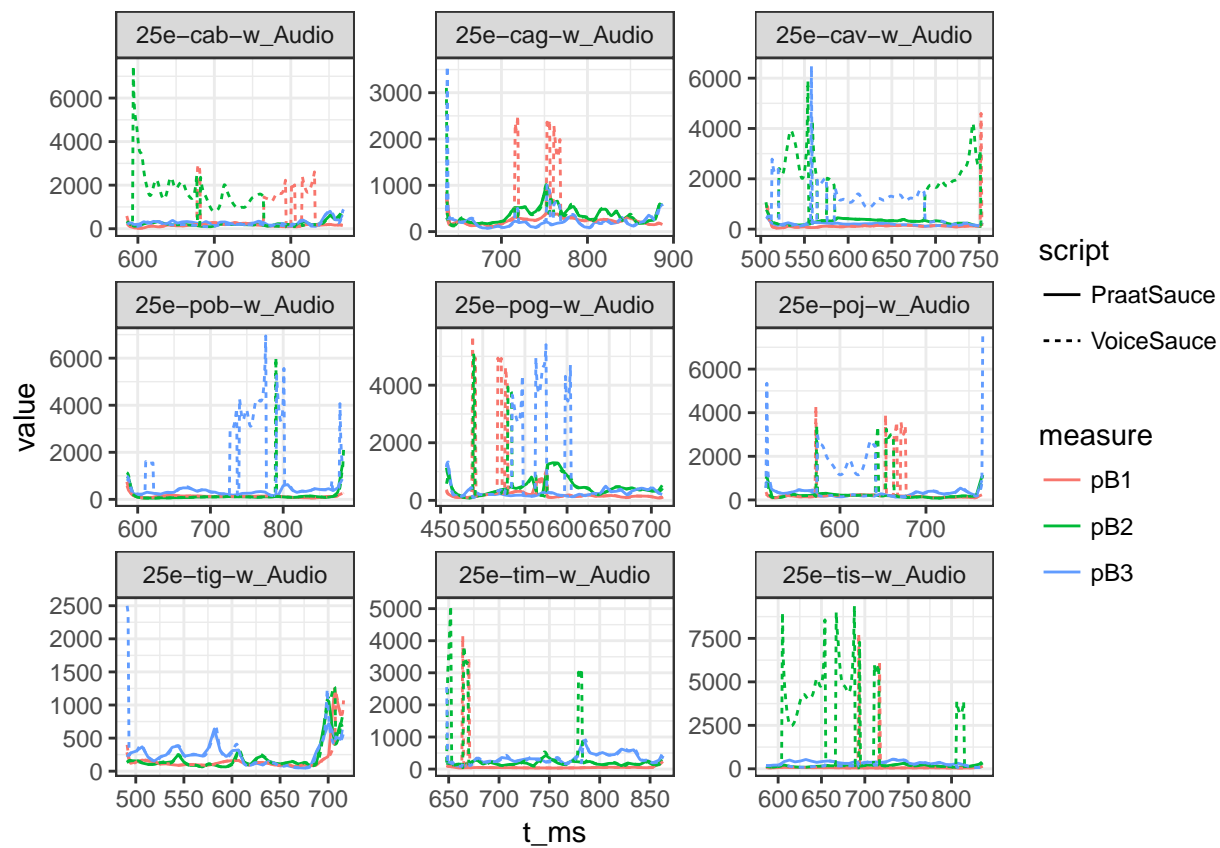
Bandwidths

PraatSauce estimated vs. formula bandwidths



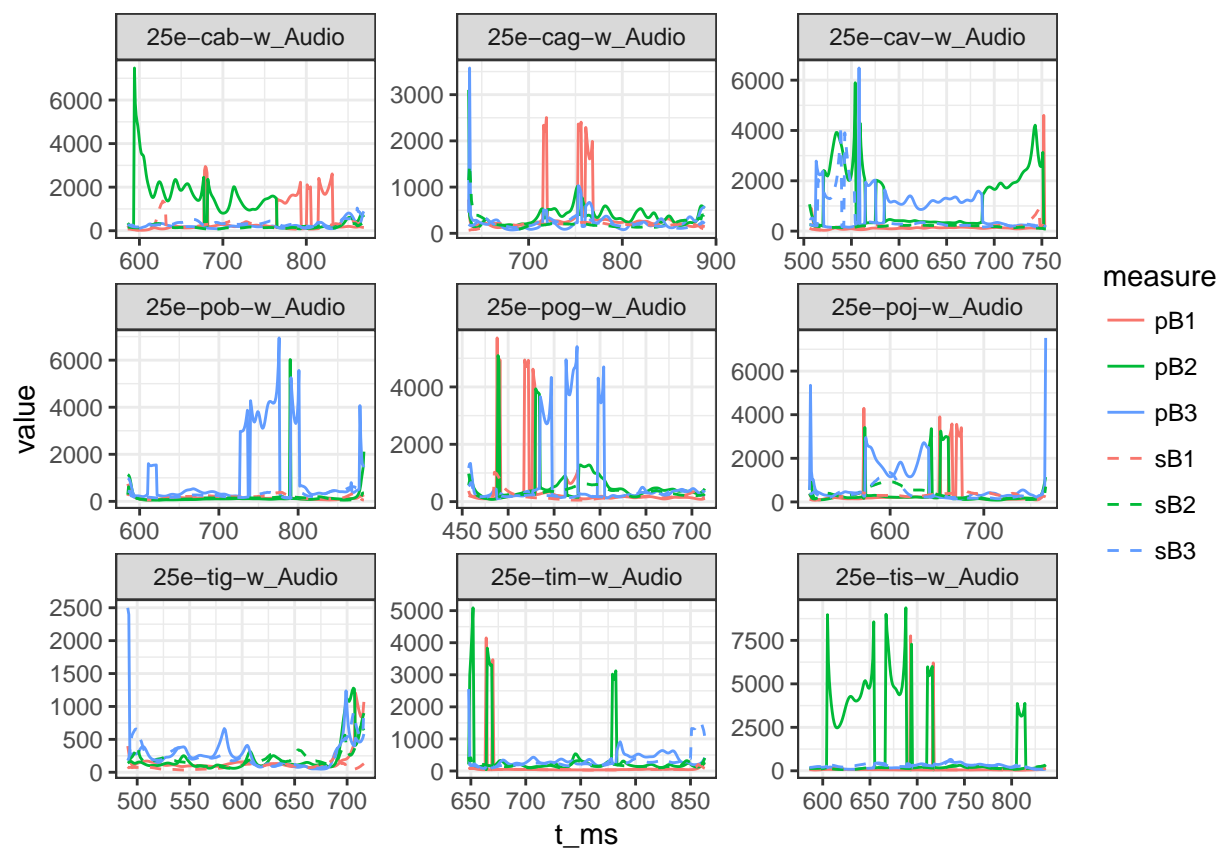
Compared to the formula estimates, PraatSauce estimated bandwidths are huge...

PraatSauce vs. VoiceSauce estimated bandwidths



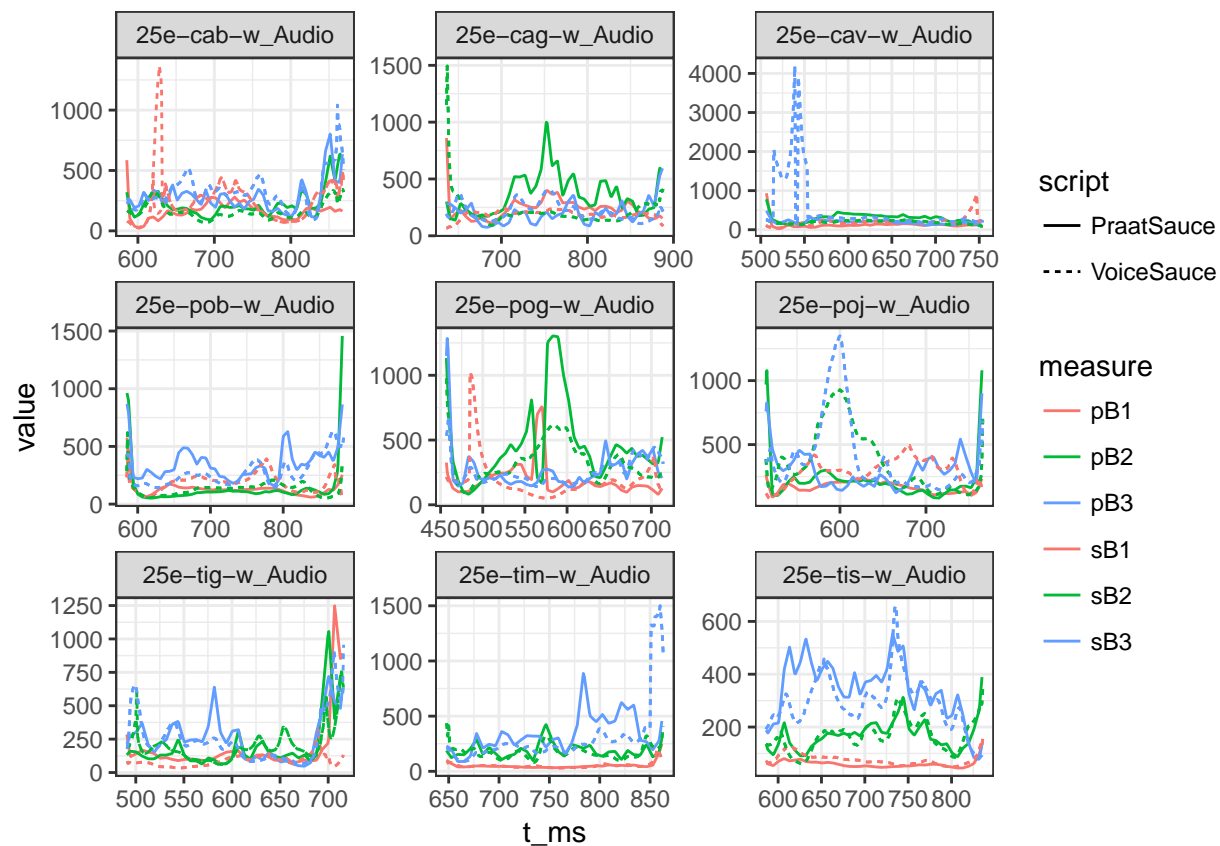
... but VoiceSauce Praat-estimated bandwidths are an order of magnitude huger.

VoiceSauce Praat vs. Snack estimated bandwidths



VoiceSauce's Snack estimates (if that's really what they are) look less erratic.

VoiceSauce Snack vs. PraatSauce estimated bandwidths

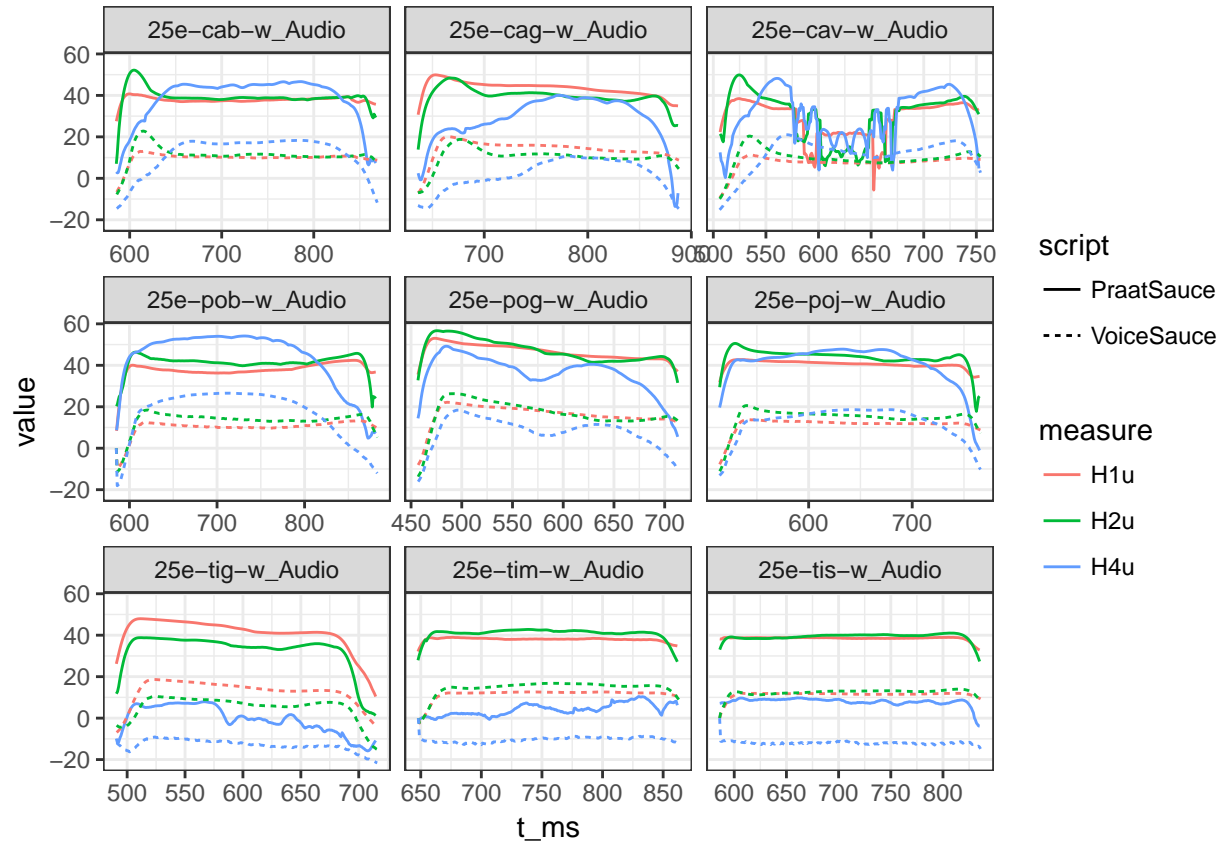


PraatSauce estimates not completely off from Snack's.

Uncorrected amplitudes

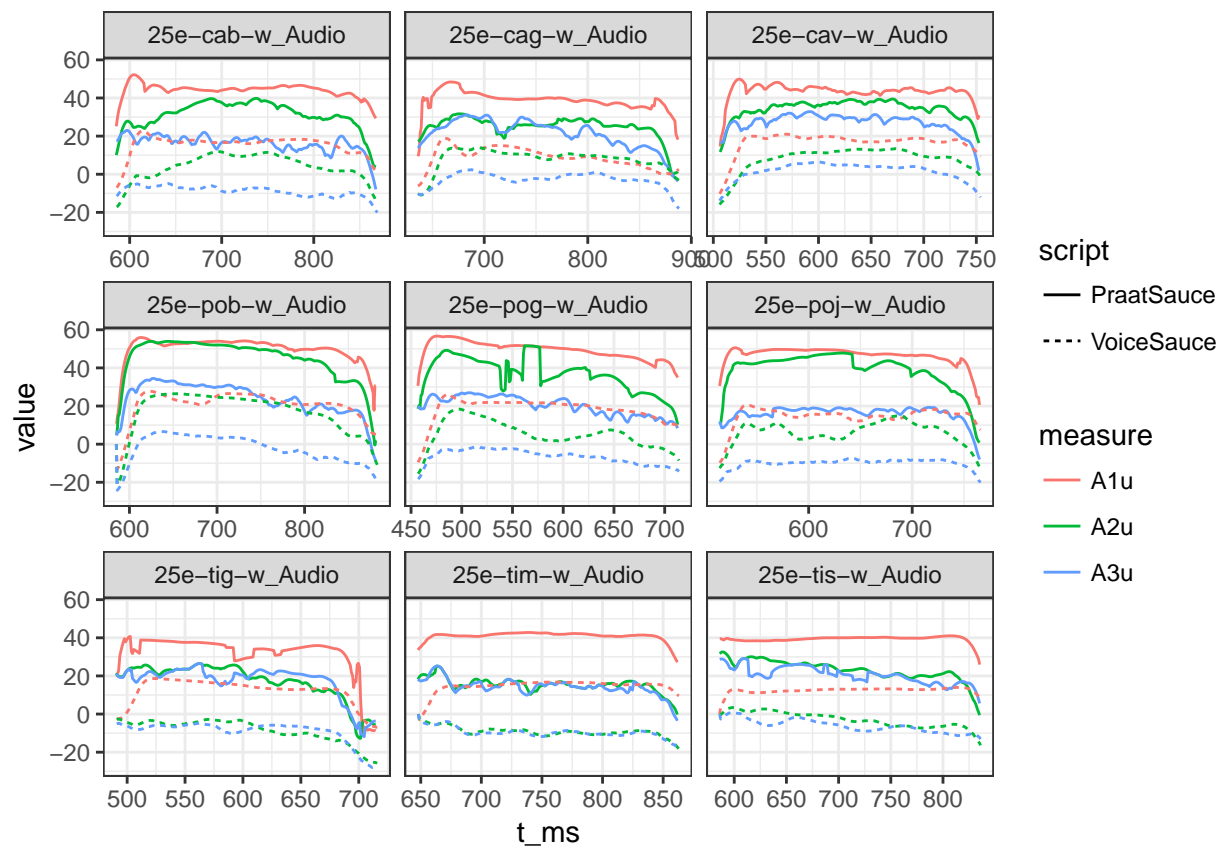
PraatSauce vs. VoiceSauce H1, H2, H4

Note that the choice of bandwidth estimator is irrelevant here.



The middle third of *cav* is a real problem for PraatSauce (at least with the chosen settings).

PraatSauce vs. VoiceSauce A1, A2, A3



The higher-order harmonics are not as much of a problem.

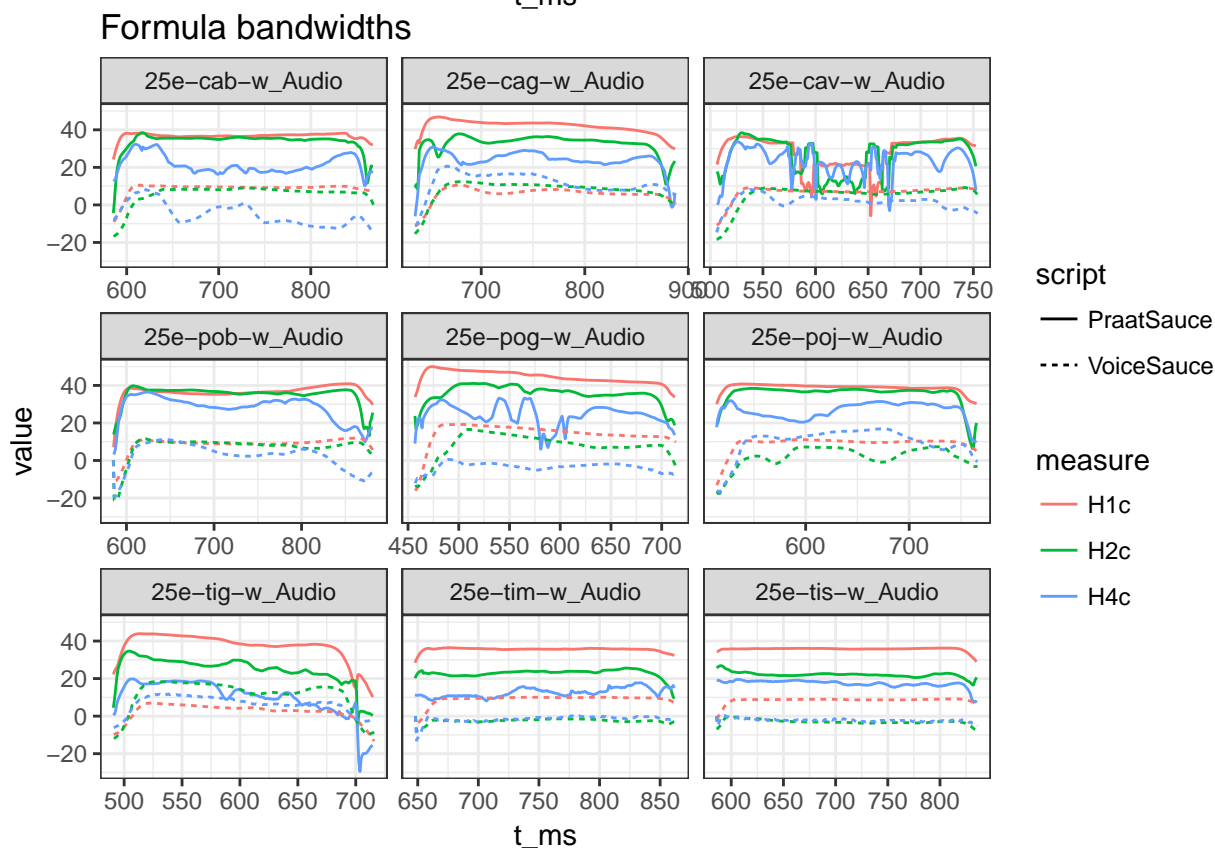
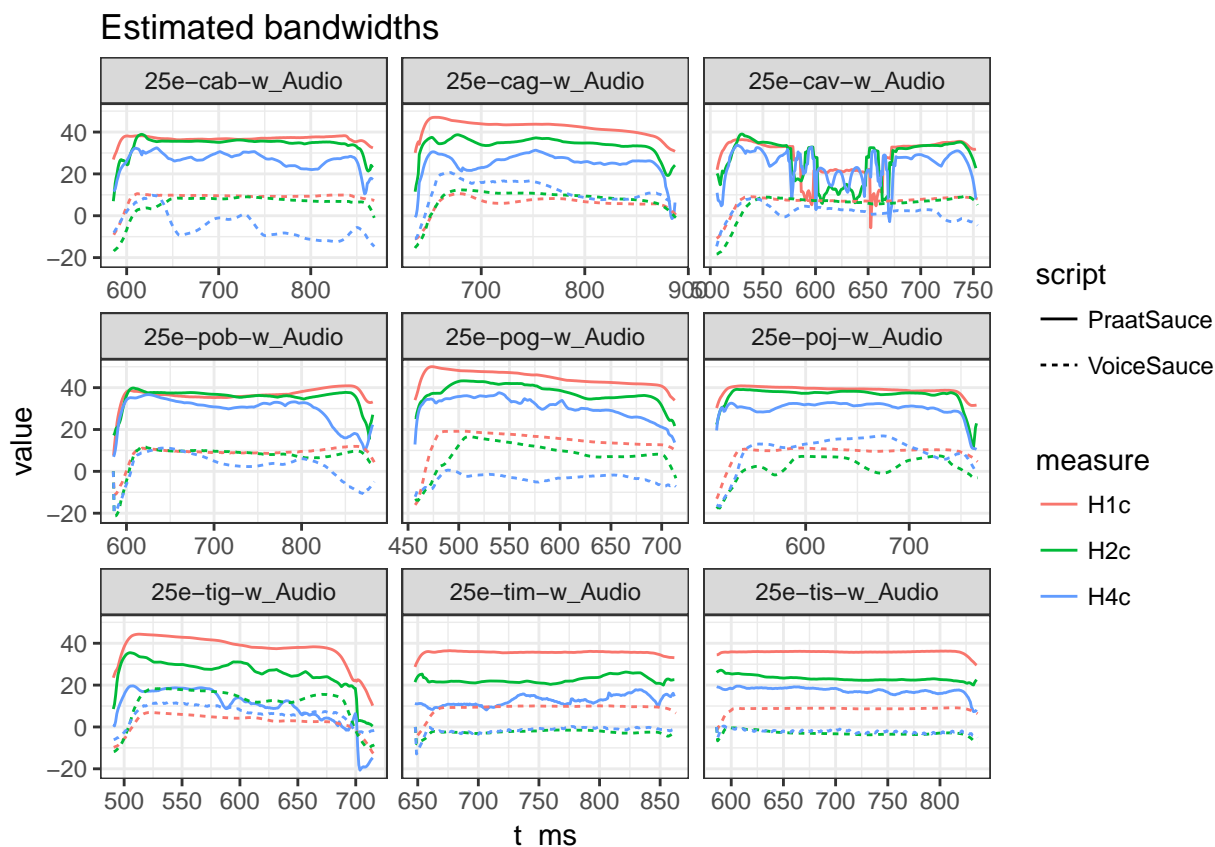
VoiceSauce estimates are consistently 20-25 dB lower than the PraatSauce estimates, and are sometimes negative, which seems... strange. This suggests to me they are being attenuated somewhere, though I have not been able to find the piece of code where this happens.

Corrected amplitudes

Here, choice of formant bandwidth estimator potentially matters.

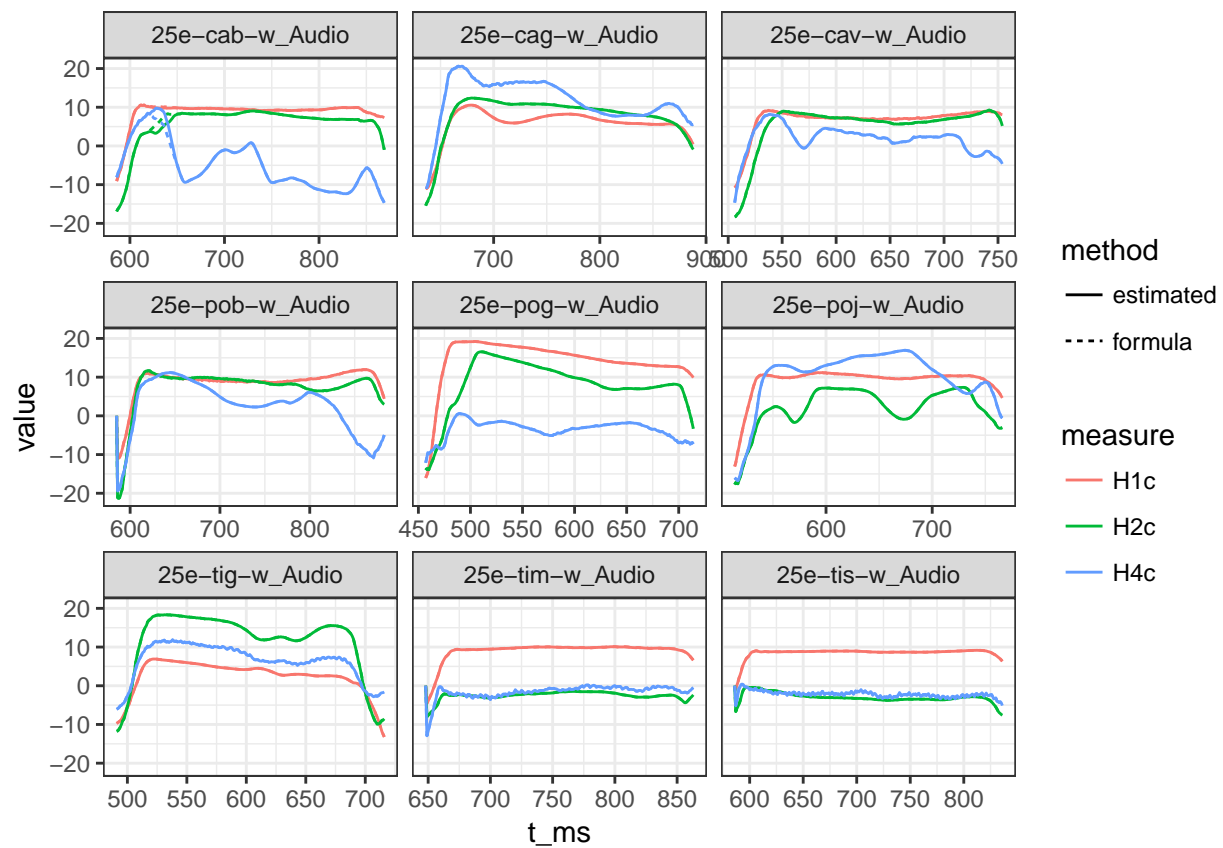
In these plots, PraatSauce is using Praat and VoiceSauce is using Snack estimates.

PraatSauce vs. VoiceSauce H1*, H2*, H4*



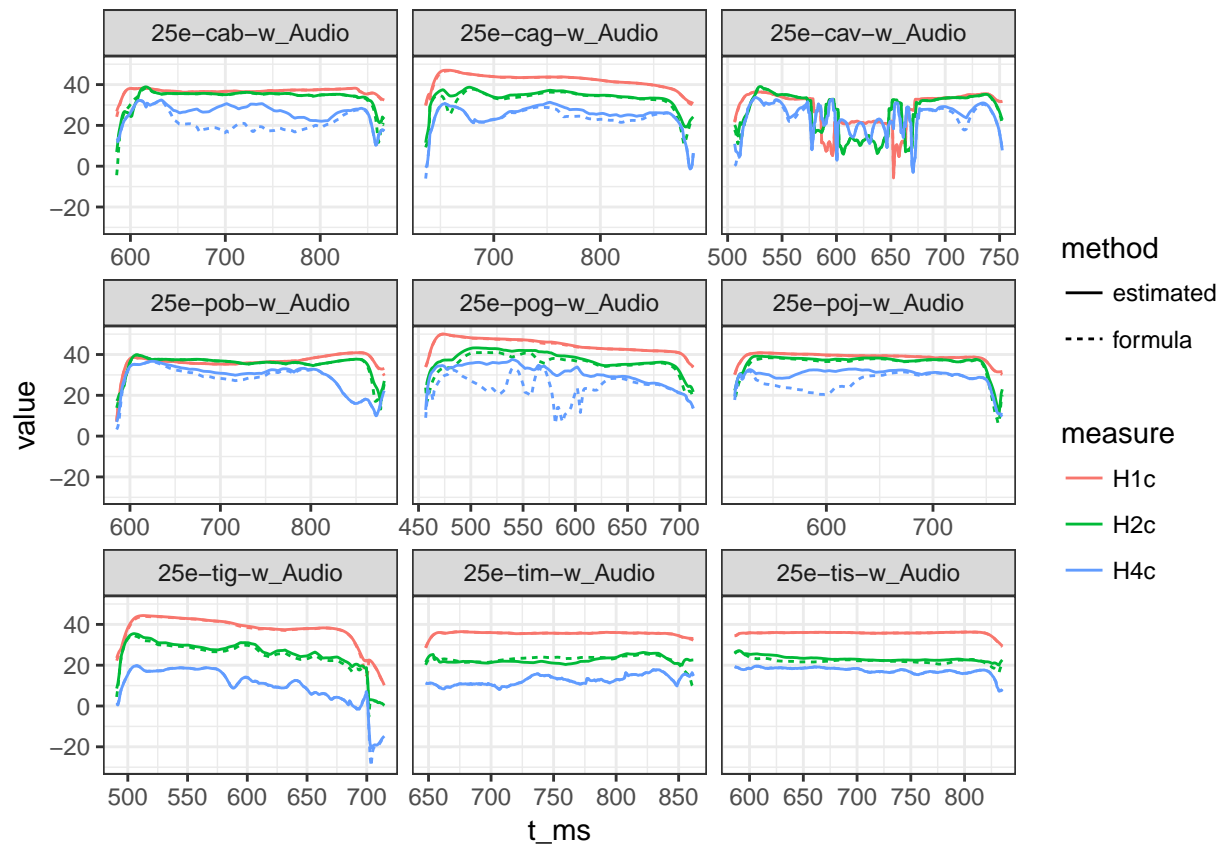
For VoiceSauce, using estimated bandwidths is virtually unnoticeable:

VoiceSauce estimated vs. formula bandwidths, H1*, H2*, H4*

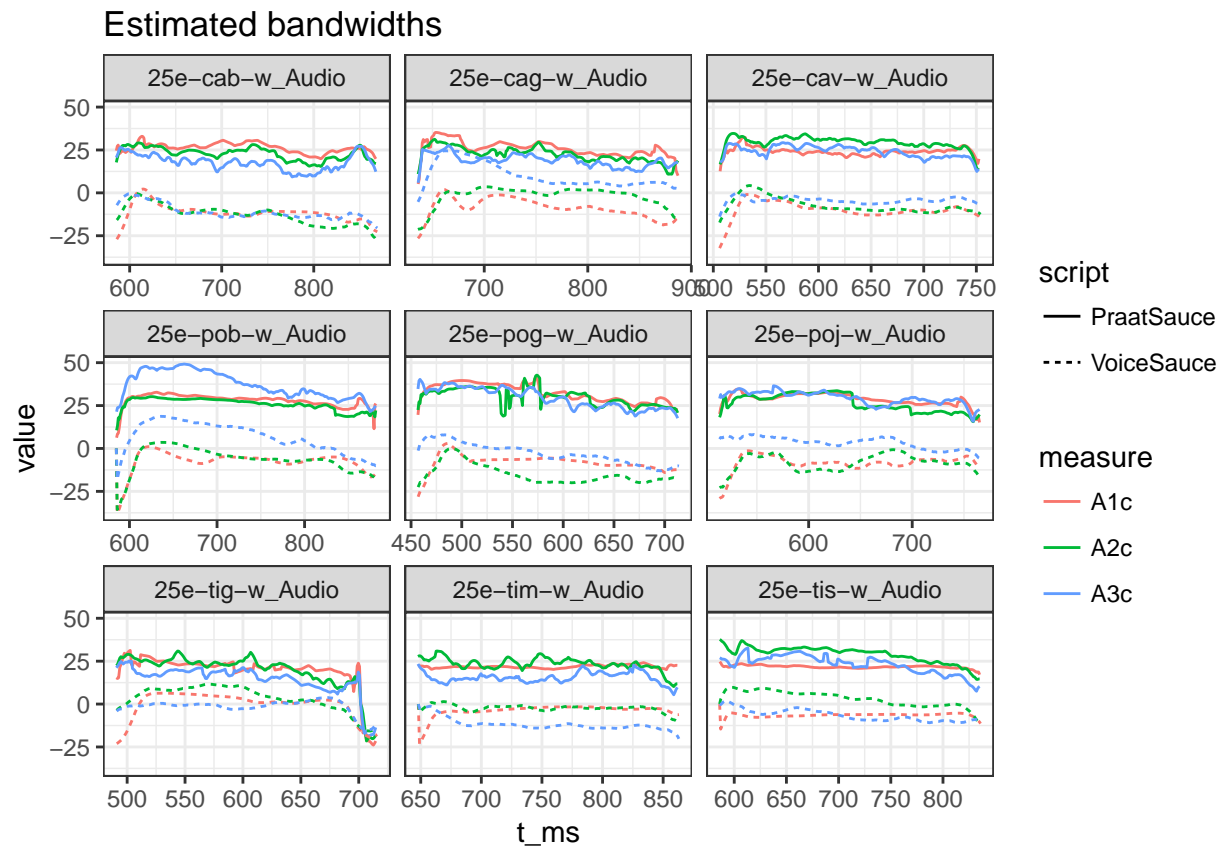


For PraatSauce, using the formula bandwidths makes only very minor differences:

PraatSauce estimated vs. formula bandwidths, H1*, H2*, H4*

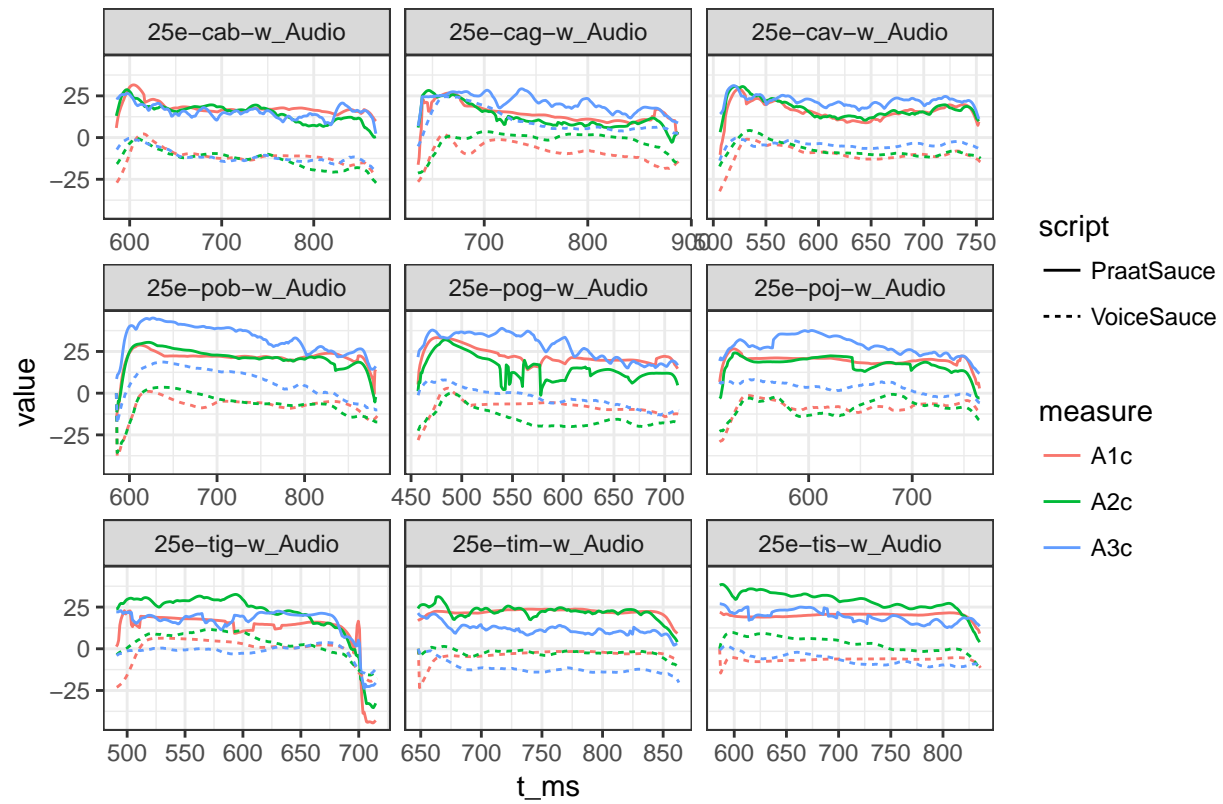


PraatSauce vs. VoiceSauce A1*, A2*, A3*

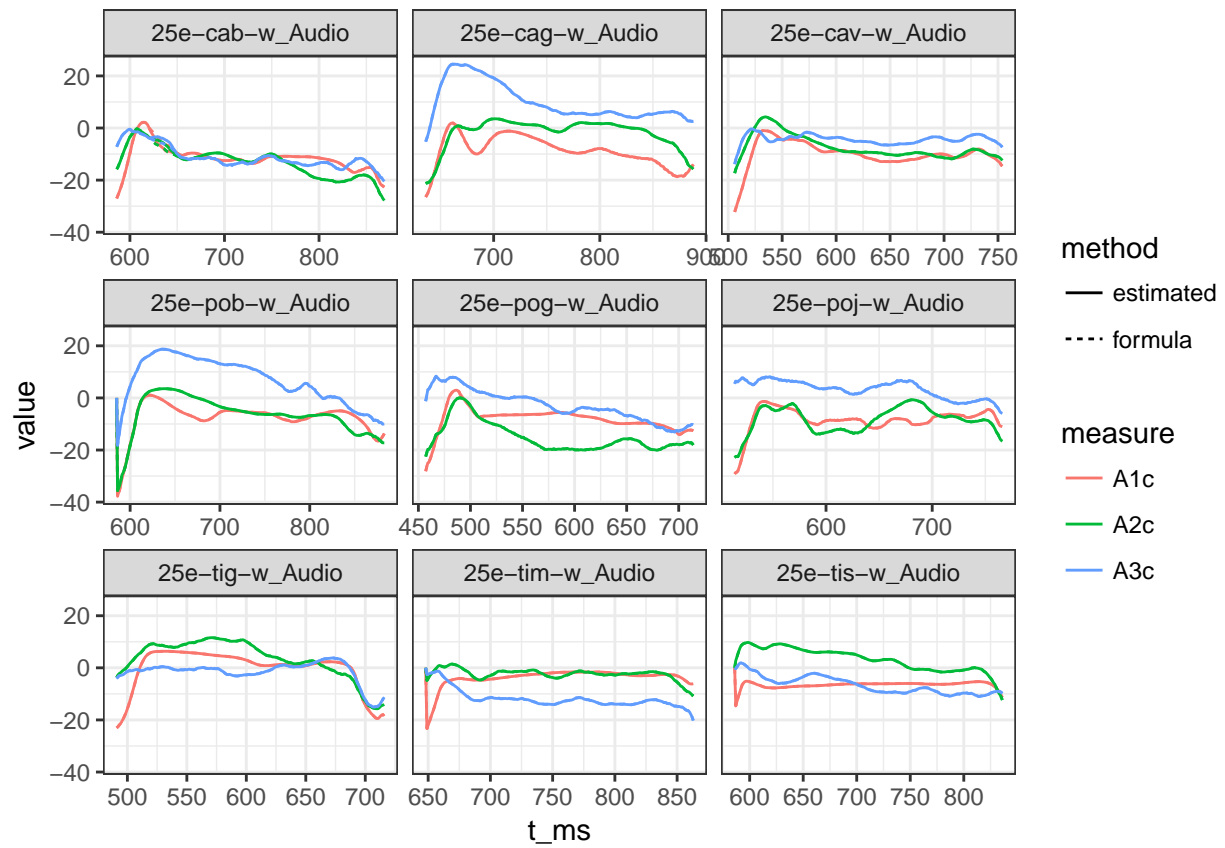


PraatSauce vs. VoiceSauce A1*, A2*, A3*

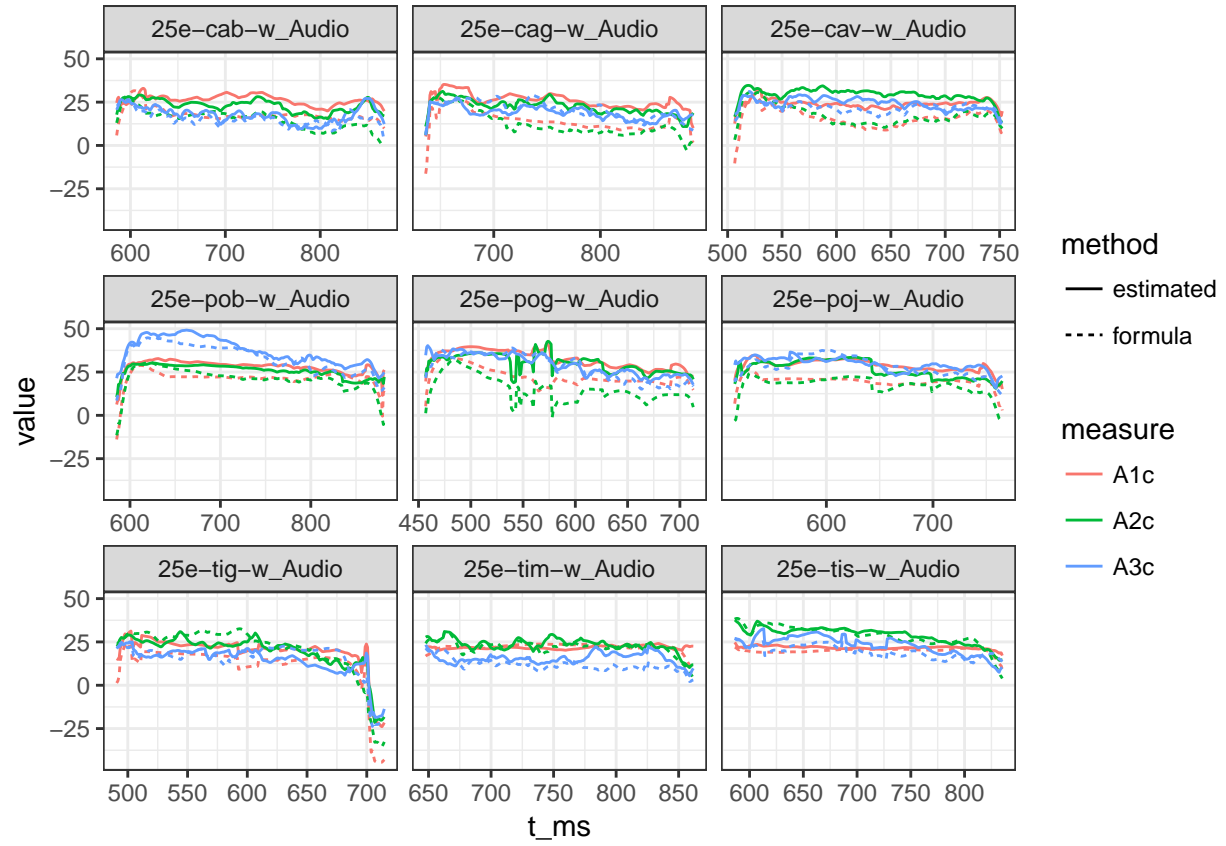
Formula bandwidths



VoiceSauce estimated vs. formula bandwidths, A1*, A2*, A3*

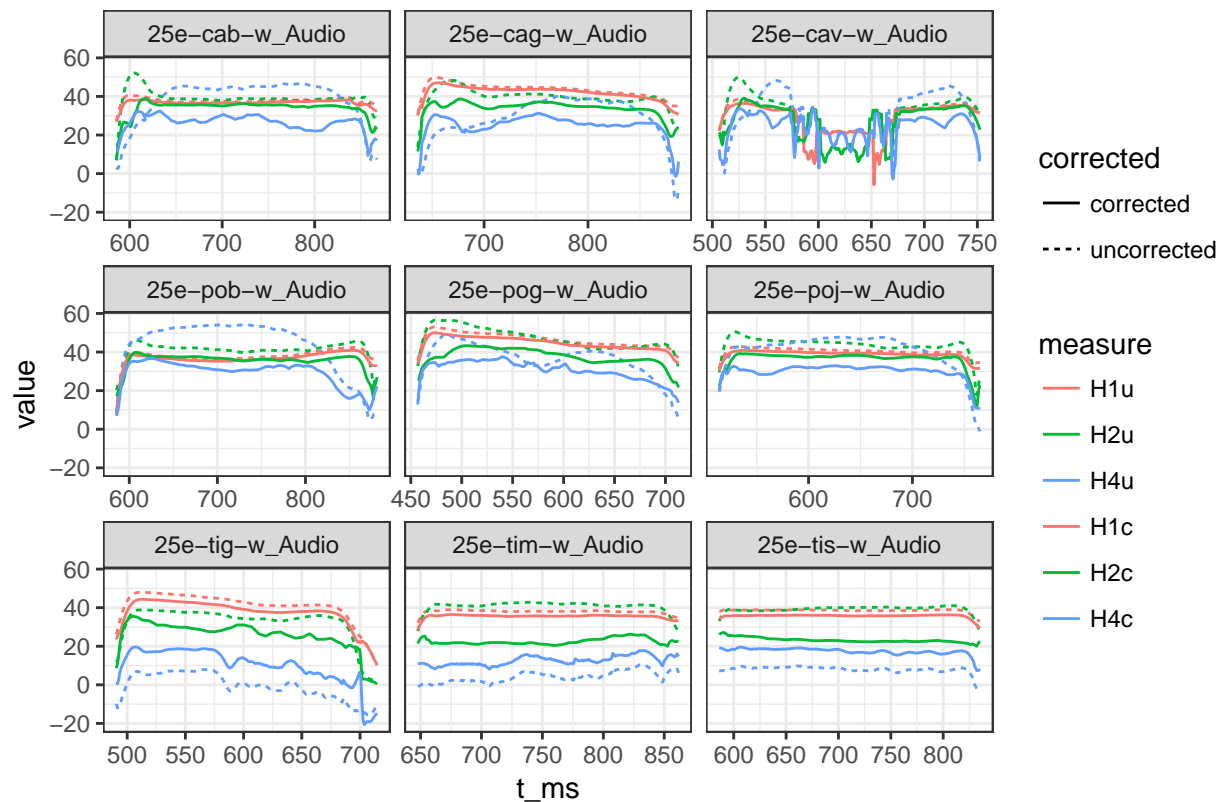


PraatSauce estimated vs. formula bandwidths, A1*, A2*, A3*

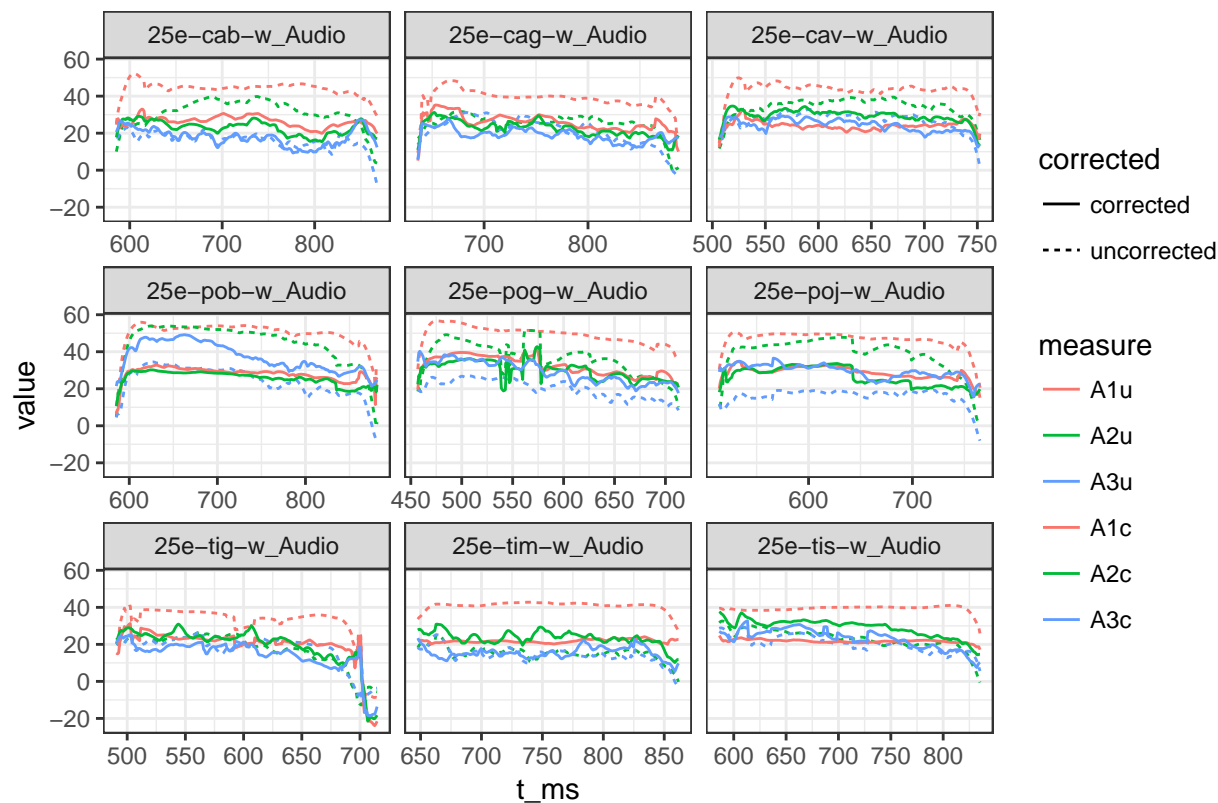


PraatSauce corrected vs. uncorrected

Estimated bandwidths

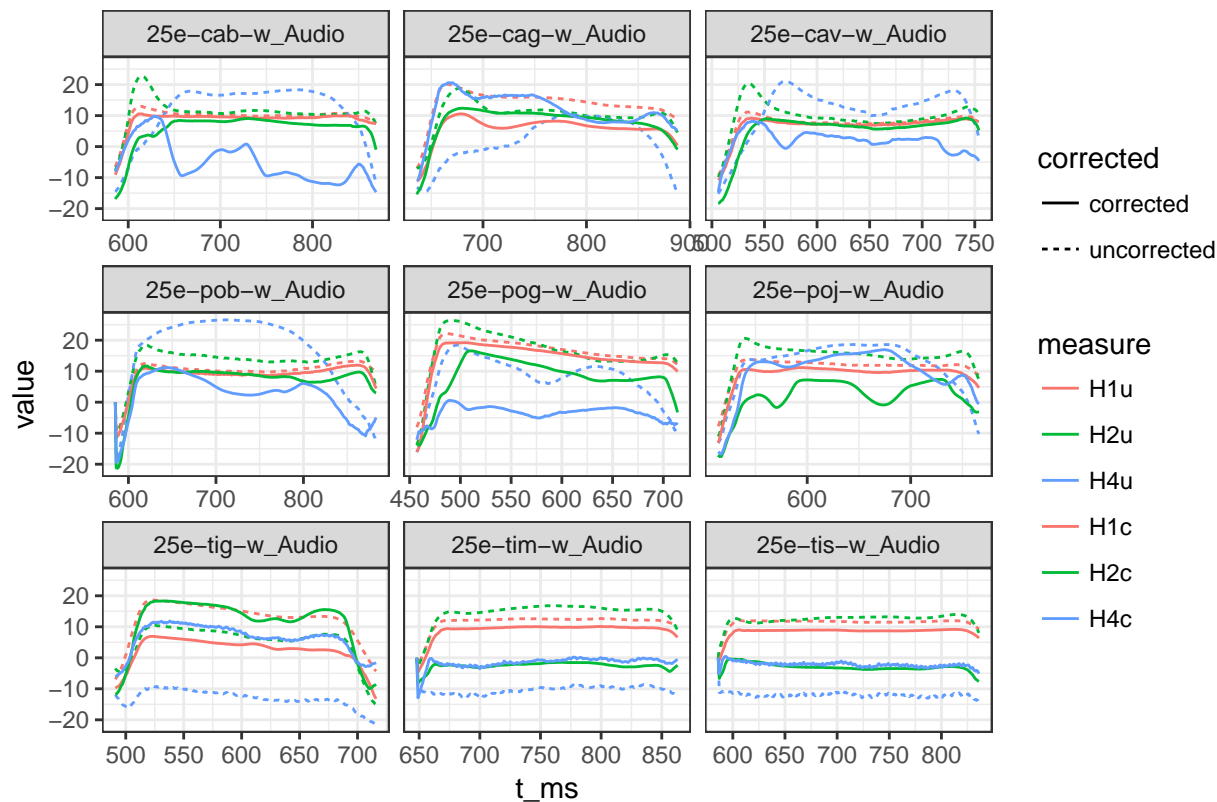


Formula bandwidths

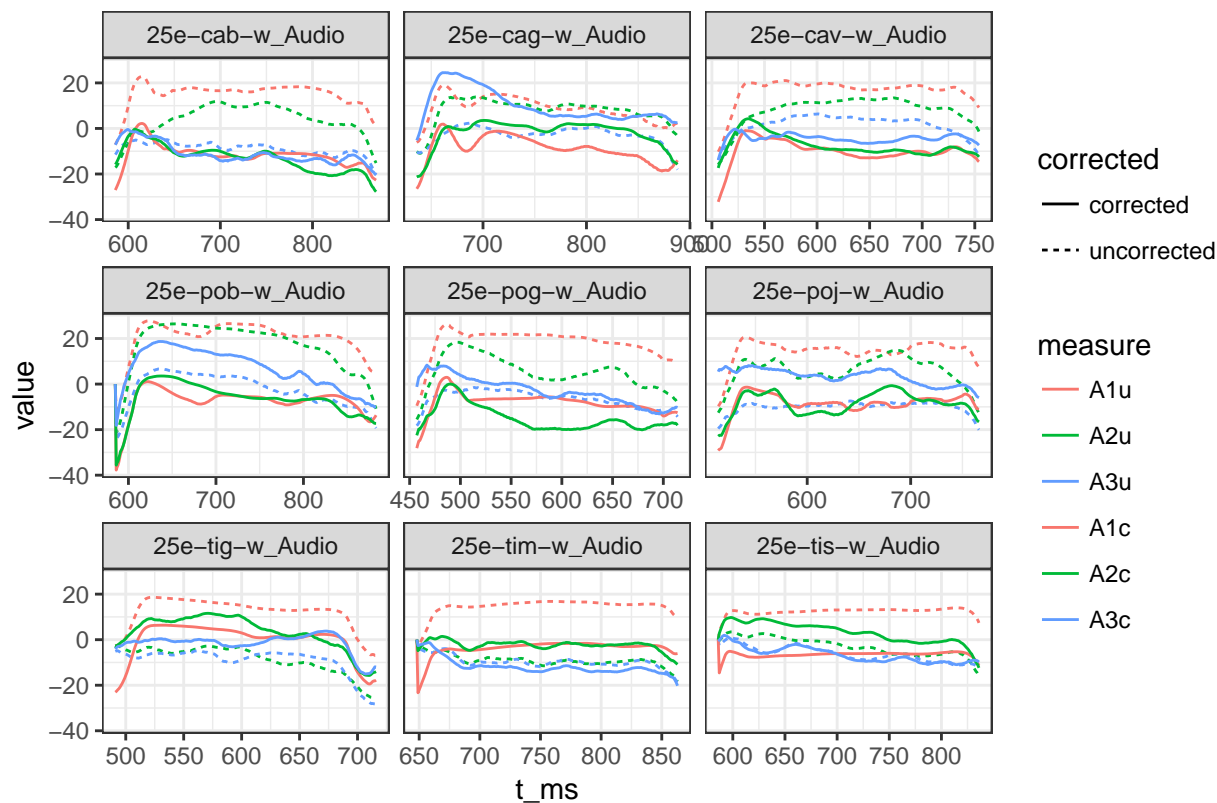


VoiceSauce corrected vs. uncorrected

Estimated bandwidths



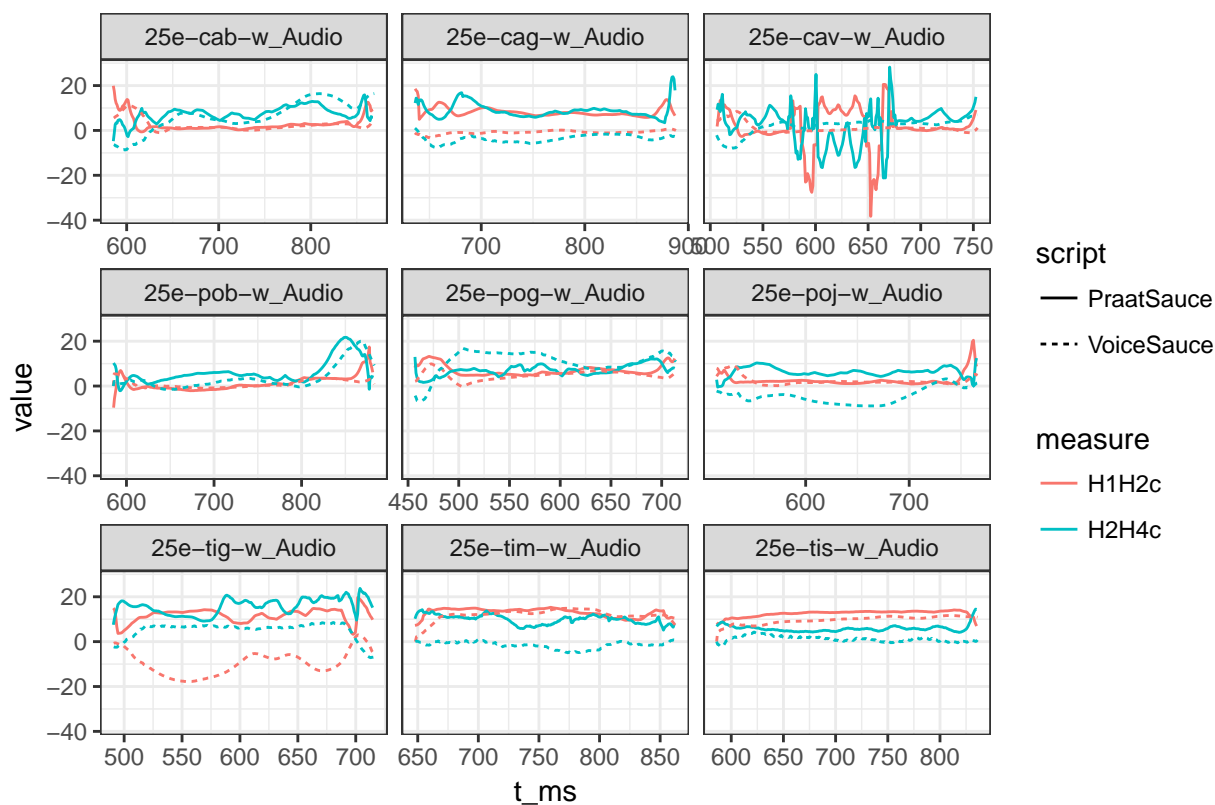
Formula bandwidths



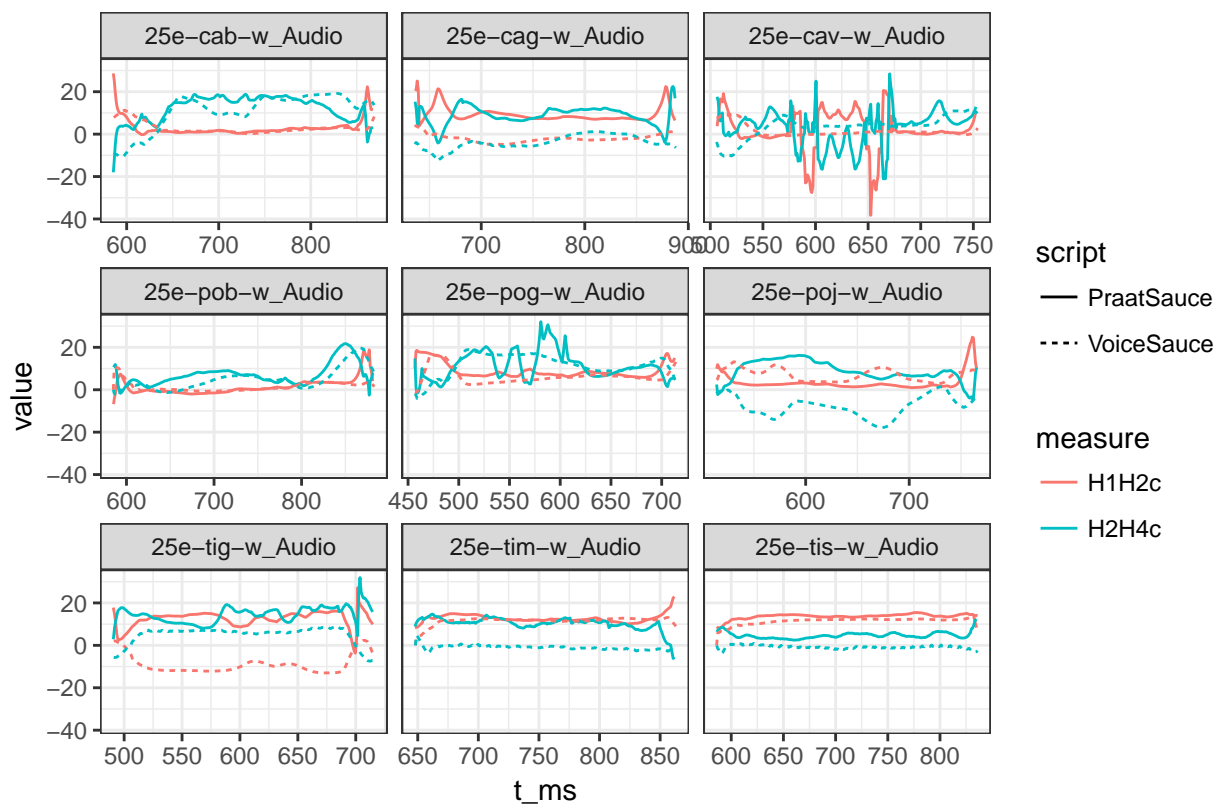
Corrected differences

PraatSauce vs. VoiceSauce $H1^*-H2^*$ & $H2^*-H4^*$

Estimated bandwidths

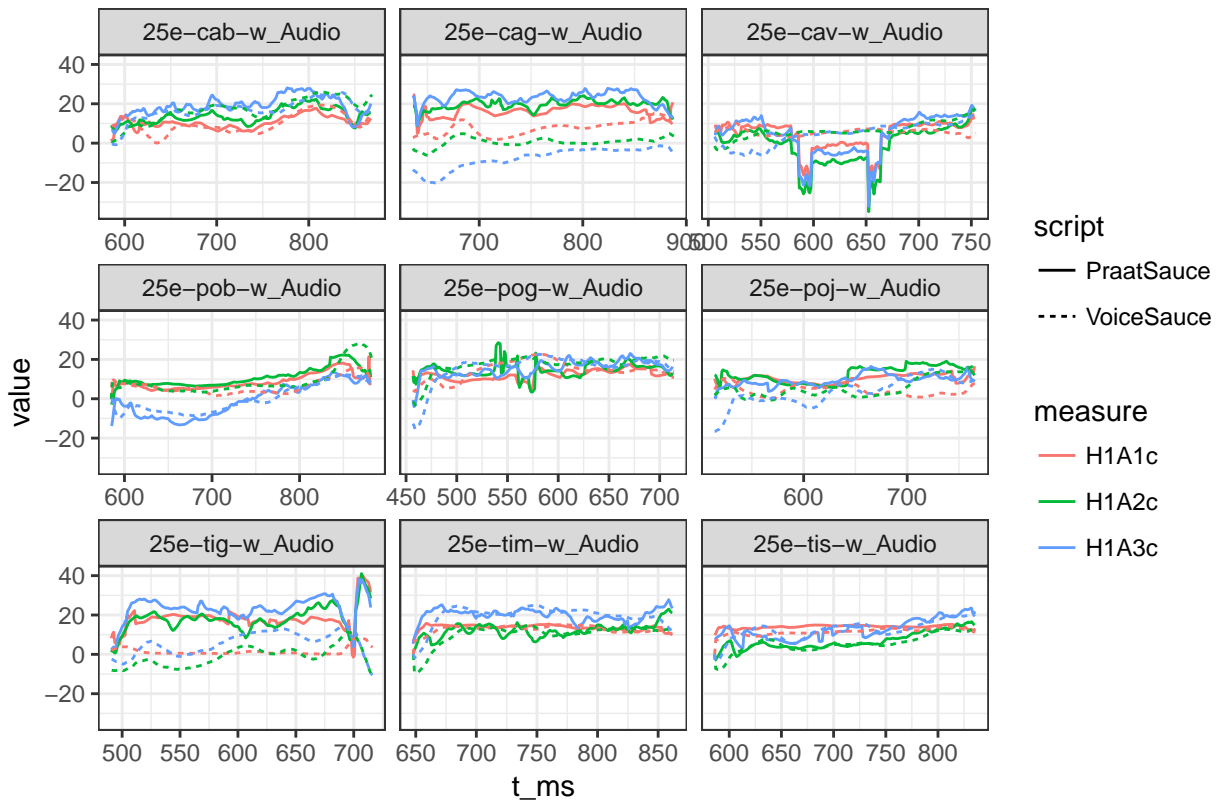


Formula bandwidths

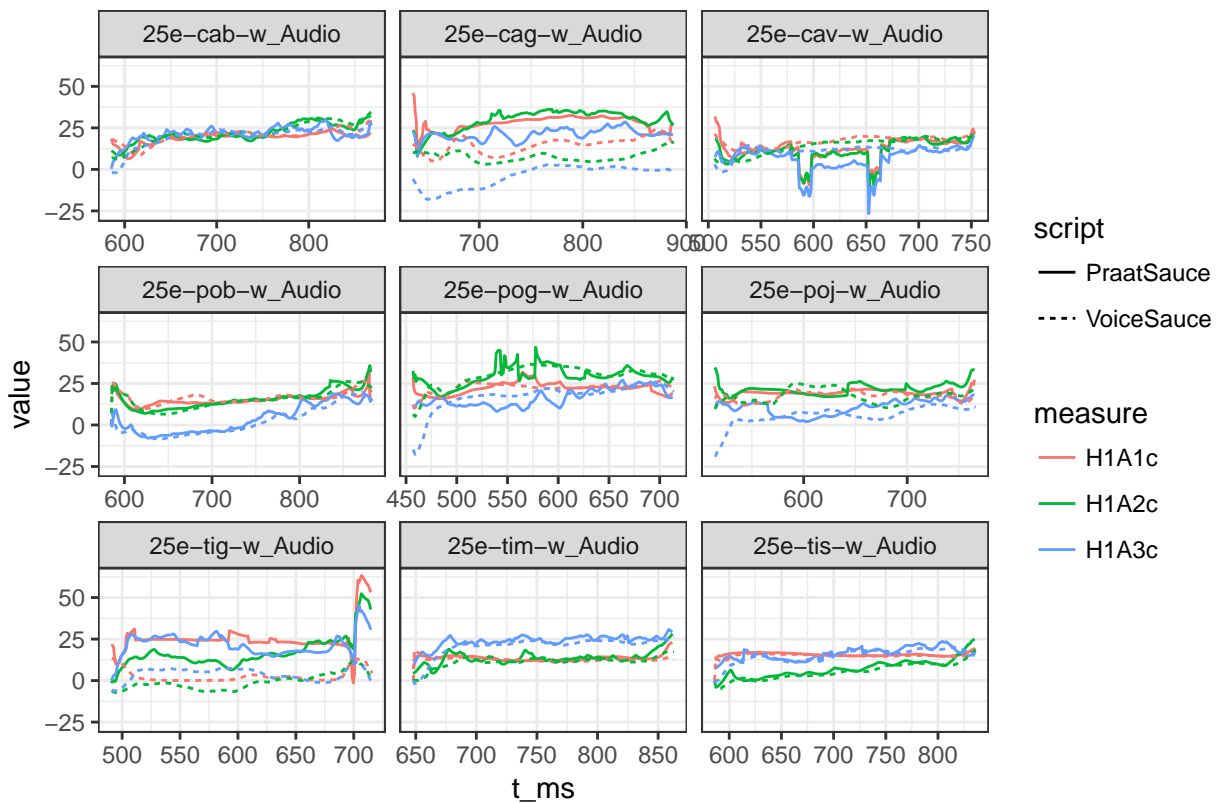


PraatSauce vs. VoiceSauce H1*-A1*, A2*, A3*

Estimated bandwidths



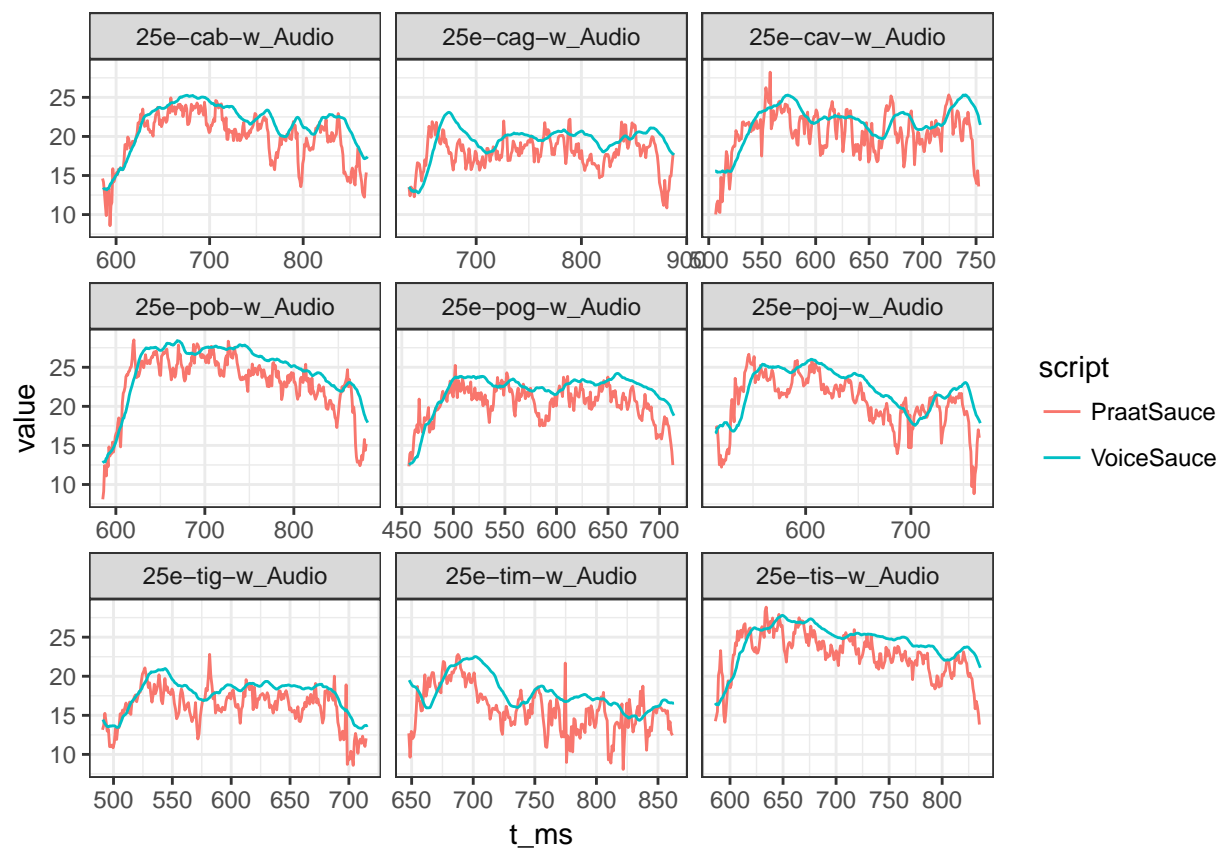
Formula bandwidths



PraatSauce seems to have higher difference estimates for (some of) the *-g* items.

The issue with the middle third of *cav* might be regarded as positive if this token is really being produced with nonmodal voice.

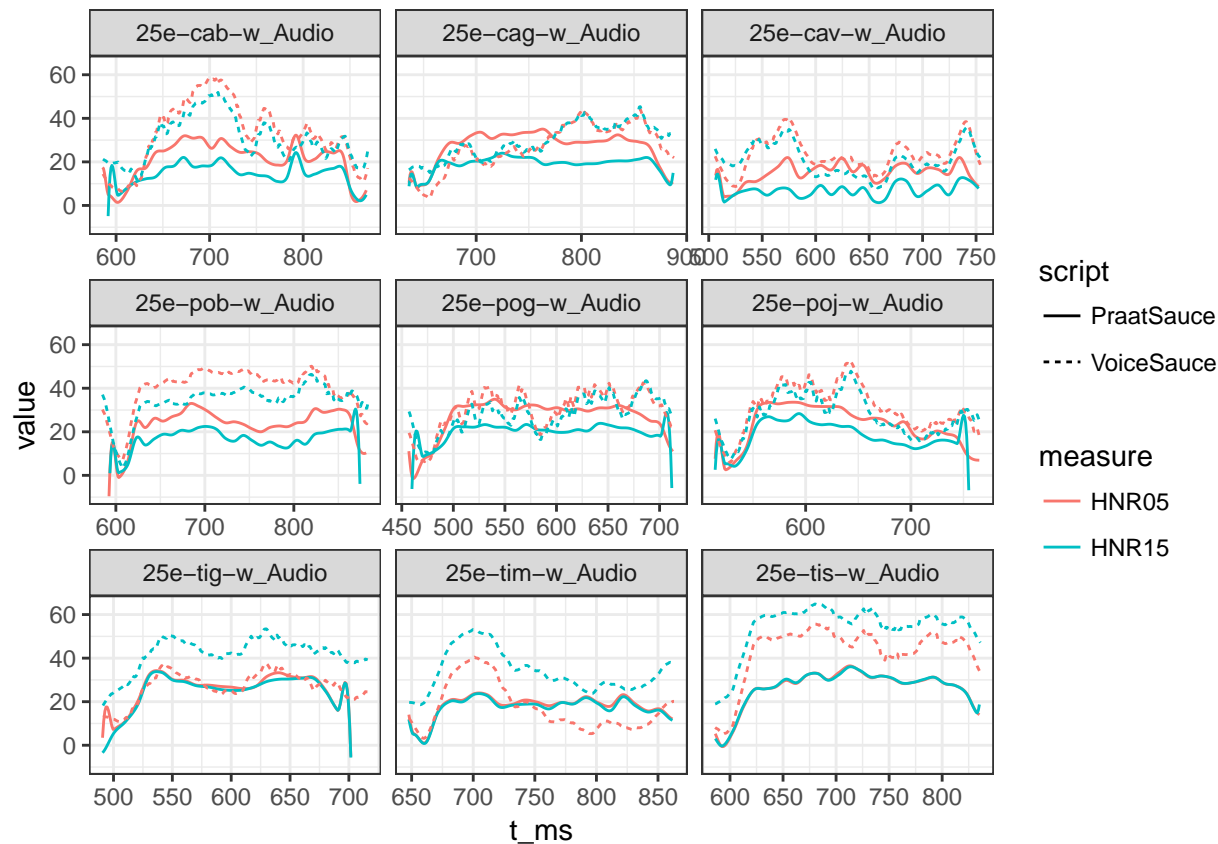
Cepstral peak prominence



Praat(Sauce) estimates are comparable if smoothed.

Harmonic to noise ratios

Here just showing HNR05 and HNR15 for clarity.

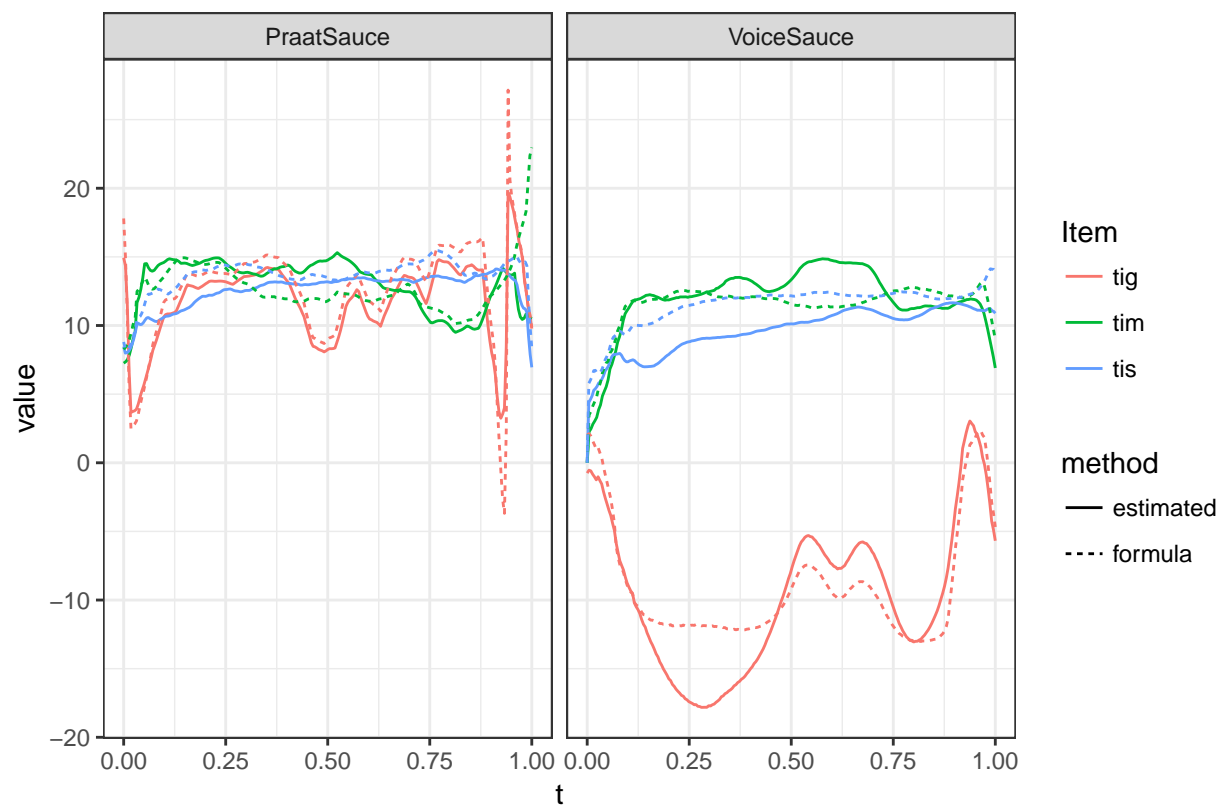


Again, the Praat estimates differ in amplitude, but maintain roughly the same trajectories. However, the PraatSauce implementation is much less sophisticated than that of VoiceSauce, and relies entirely on Praat's `To Harmonicity...` function.

Distinguishing voice qualities

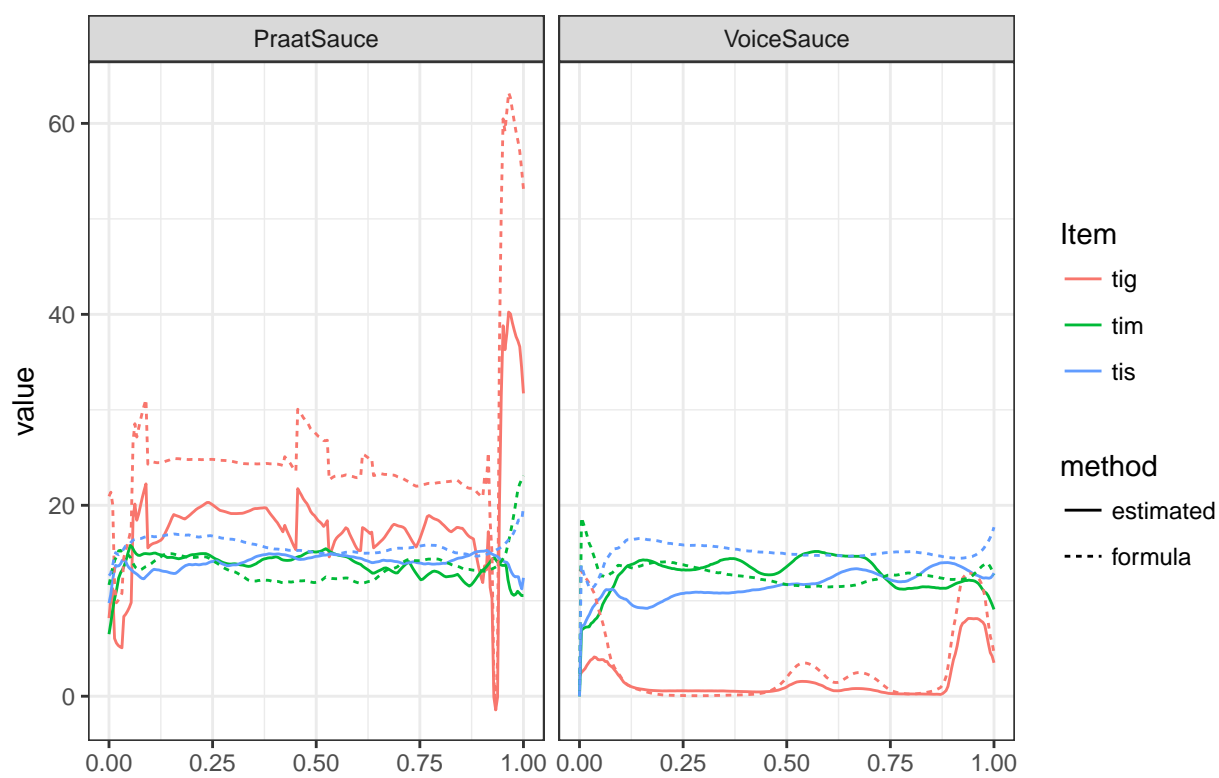
High vowels

VoiceSauce vs. PraatSauce H1*-H2*

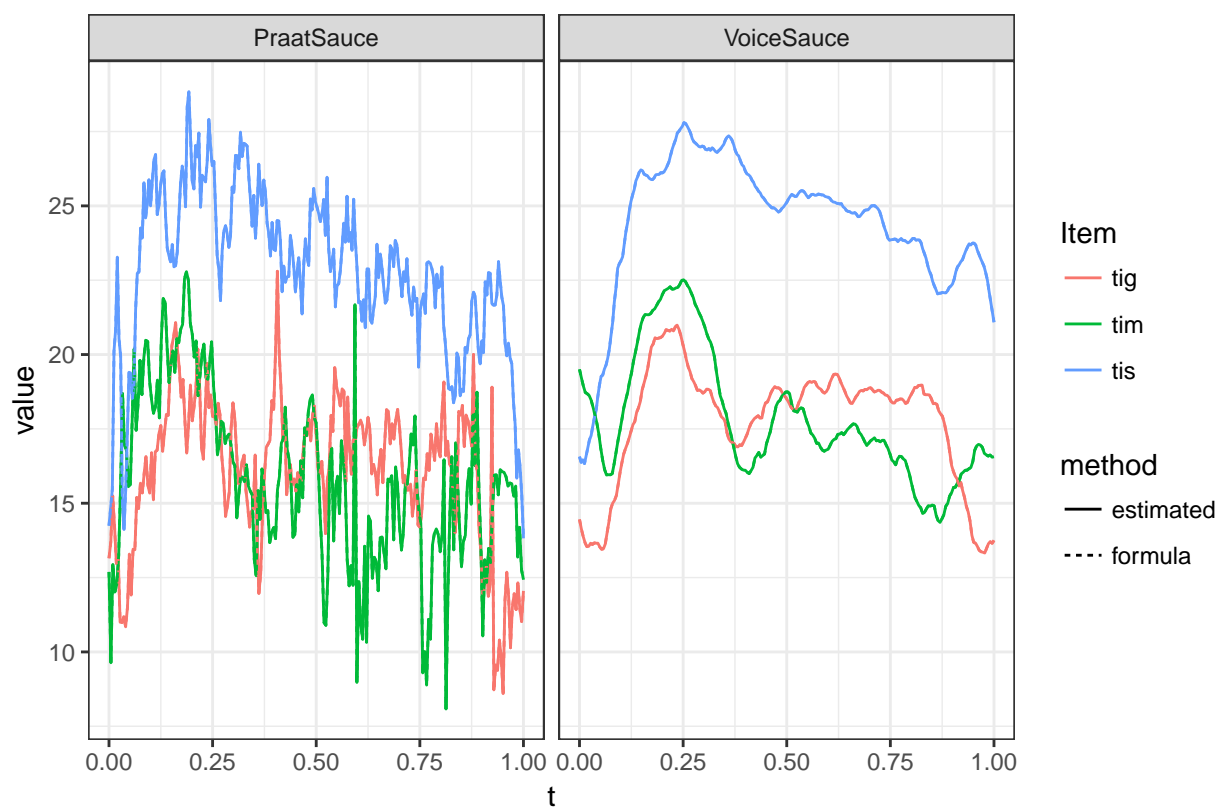


PraatSauce does not really find the expected positive spectral slope for **tig**. It is clearly distinguished by VoiceSauce, but the slope is unexpectedly negative, rather than positive.

VoiceSauce vs. PraatSauce H1*-A1*

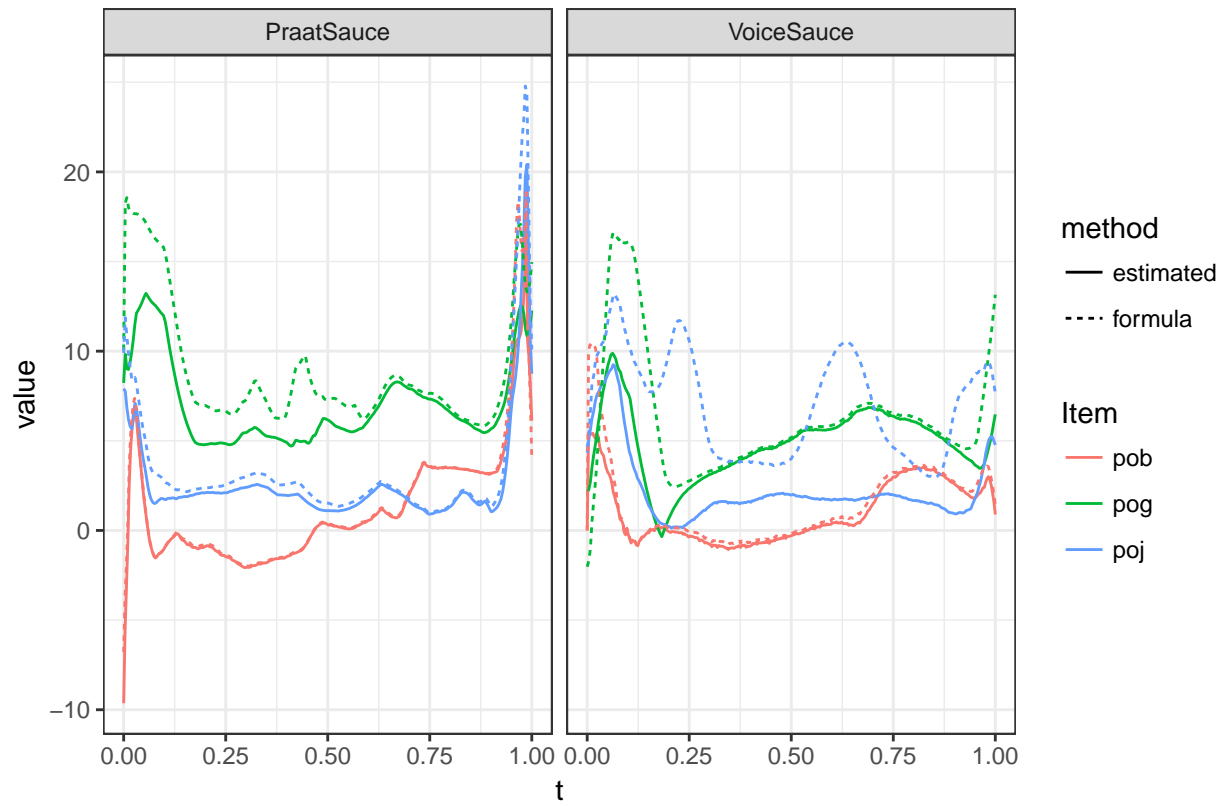


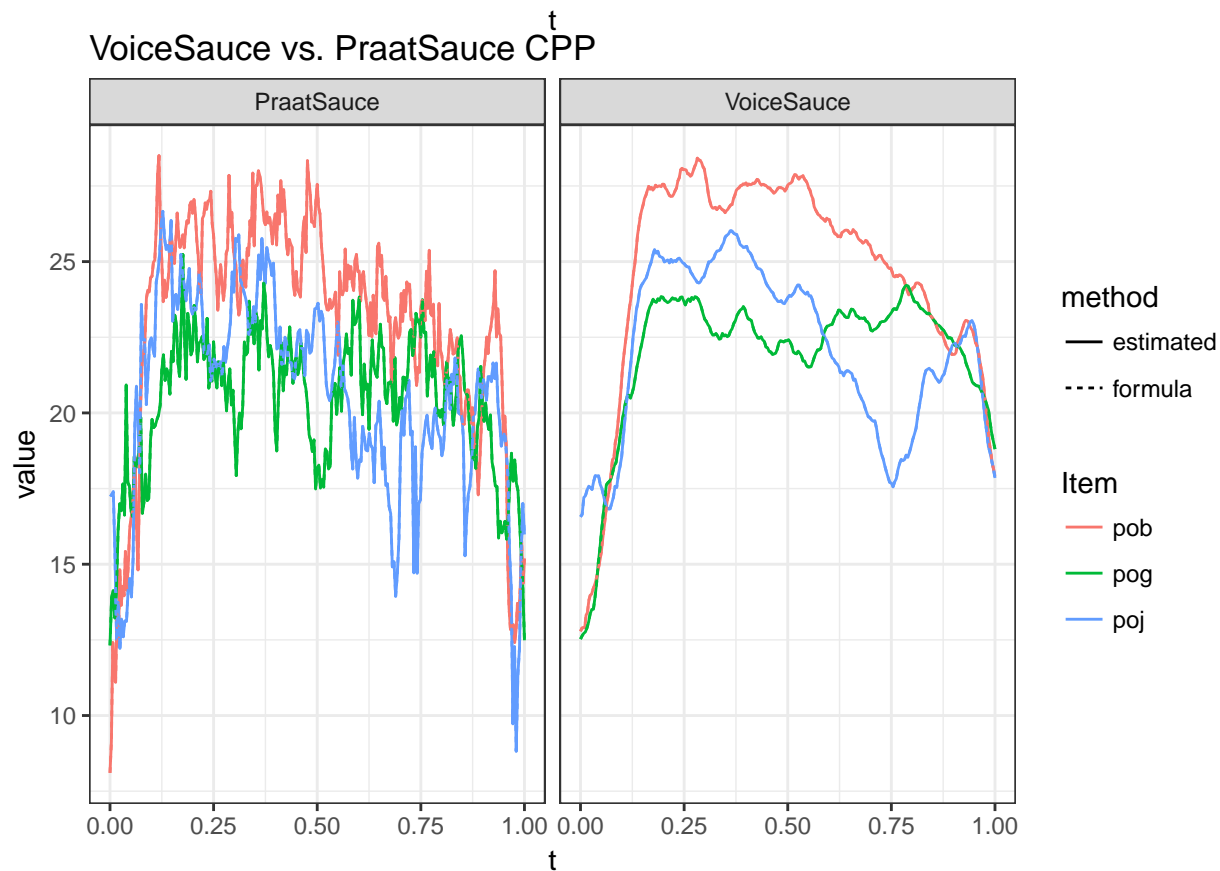
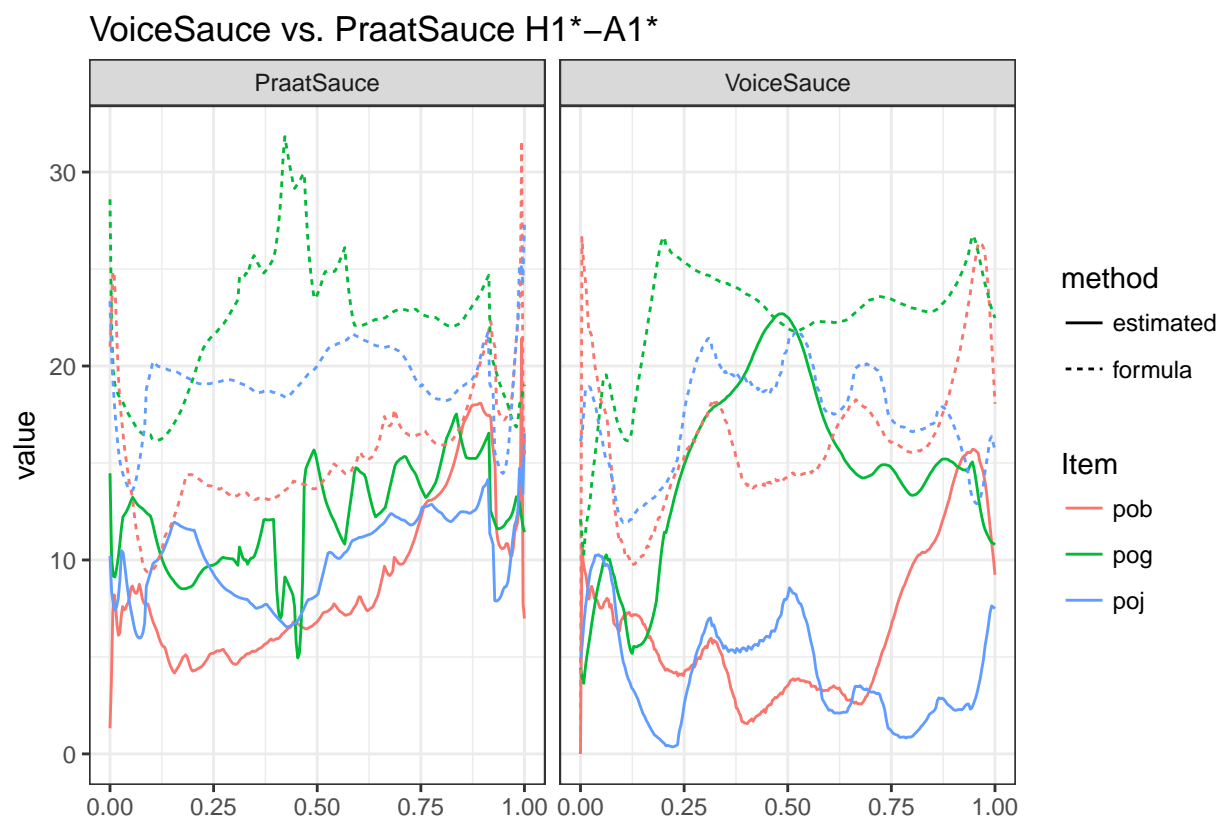
VoiceSauce vs. PraatSauce CPP



Mid vowels

VoiceSauce vs. PraatSauce H1*–H2*

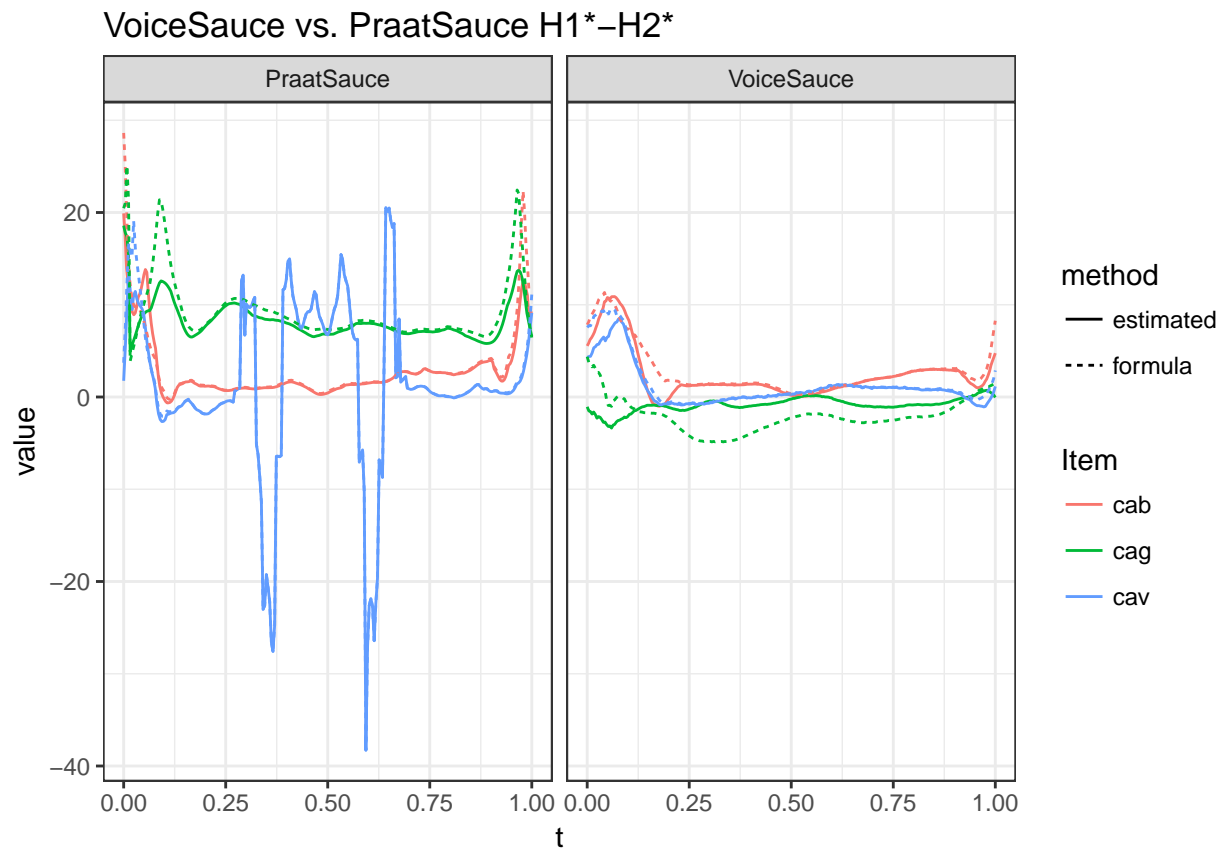




For the mid vowel, the breathy toned pog is more clearly differentiated, though again VoiceSauce makes a

clearer distinction.

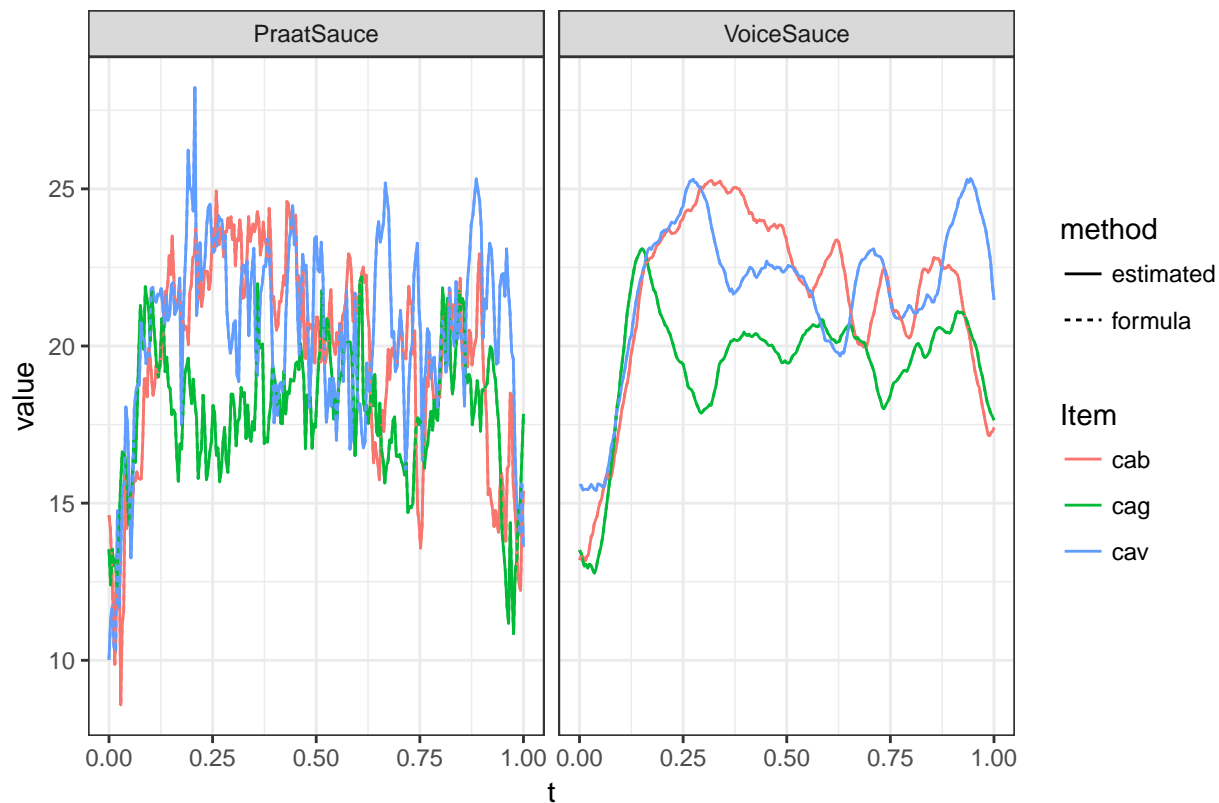
Low vowels



VoiceSauce vs. PraatSauce H1*-A1*



VoiceSauce vs. PraatSauce CPP



For the low vowel tokens, VoiceSauce doesn't seem to distinguish **cag** from other syllables except for via CPP.