Figure 14. Example flyer for recruiting Stockholm Swedish talkers for recording of the SwehVd database.

SUPPLEMENTARY INFORMATION

§1 ADDITIONAL INFORMATION ABOUT THE SWEHVD DATABASEE

Participant recruitment

Participants were recruited through word-of-mouth, flyers at Stockholm University Campus, and online channels (accindi.se). Figure 14 is an example of flyers distributed at Stockholm University Campus. The flyer gives information on criteria for participation, recording procedure, reimbursement and contact information to experimenter (first author).

§1.1 Word list

Word list with all target and filler words, recorded by all talkers in the SwehVd database.

Table 5. Words recorded by the female talkers of Stockholm Swedish for the SwehVd database

Target words	Vowel IPA
hid	[iː]
hidd	[ɪ]
hyd	[yː]
hydd	[ʏ]
hed	[eː]
hedd	[ɛ]
häd	[ɛː]
hädd	[ɛ]
härd	[æː]
härr	[æ]
höd	[øː]
hödd	[ø]
hörd	[œː]
hörr	[œ]
hud	[ʉː]
hudd	[ɵ]
hod	[uː]
hodd	[ʊ]
håd	[oː]
hådd	[ɔ]
had	[ɑː]
hadd	[a]

Filler words	
titt	tand
damm	dipp
tå	buss
bål	ding
dill	porr
tugga	mitt
mat	dopp
norr	tal
must	namn
pil	pall
dina	bar
biff	till
Tina	mål
borr	Nina
dal	då
Pål	nick
nunna	ditt
mil	dugga
ting	mall
ball	bil
piff	par
tipp	morr
puss	nav
topp	nå

§2 ADDITIONAL INFORMATION FOR STUDY 1

Evaluation of implementations of Syrdal & Gopal's (1986) second dimension

For the second dimension, distinguishing between front and back vowels, Syrdal and Gopal (1986) evaluates two different bark-difference measures: F2-F1 and F3-F2. Previous studies had concluded that F2-F1 distinguishes between all Swedish vowels (Fant, 1983), however, in Syrdal and Gopal (1986)'s evaluation of American English, the F3-F2 dimension provided a better fit. Given that there seems to be language specific effects concerning Syrdal and Gopal (1986)'s second dimension (e.g., Adank, 2003), here we compare the two difference measures for the vowels in the SwehVd database.

Figure 15 displays the separability index for the two implementations. The first version uses the F2-F1 bark-difference metric for the second dimension, whereas the second version (labelled *SyrdalGopal2 (Bark)*) implements the second dimension as suggested by Syrdal and Gopal (1986), F3-F2. As evident from Figure 15, the first implementation performs better at separating categories in the SwehVd data, which replicates Fant (1983).

We also evaluated the two Syrdal & Gopal implementations in terms of model predictions for perception. Figure 16 displays the categorization accuracy for models trained on normalized data under the two implementations of the Syrdal & Gopal account. Mirroring the results from the separability index, the first implementation using F2-F1 for the second dimension, outperforms the implementation using F3-F2 bark-difference measure. These results taken together indicate that

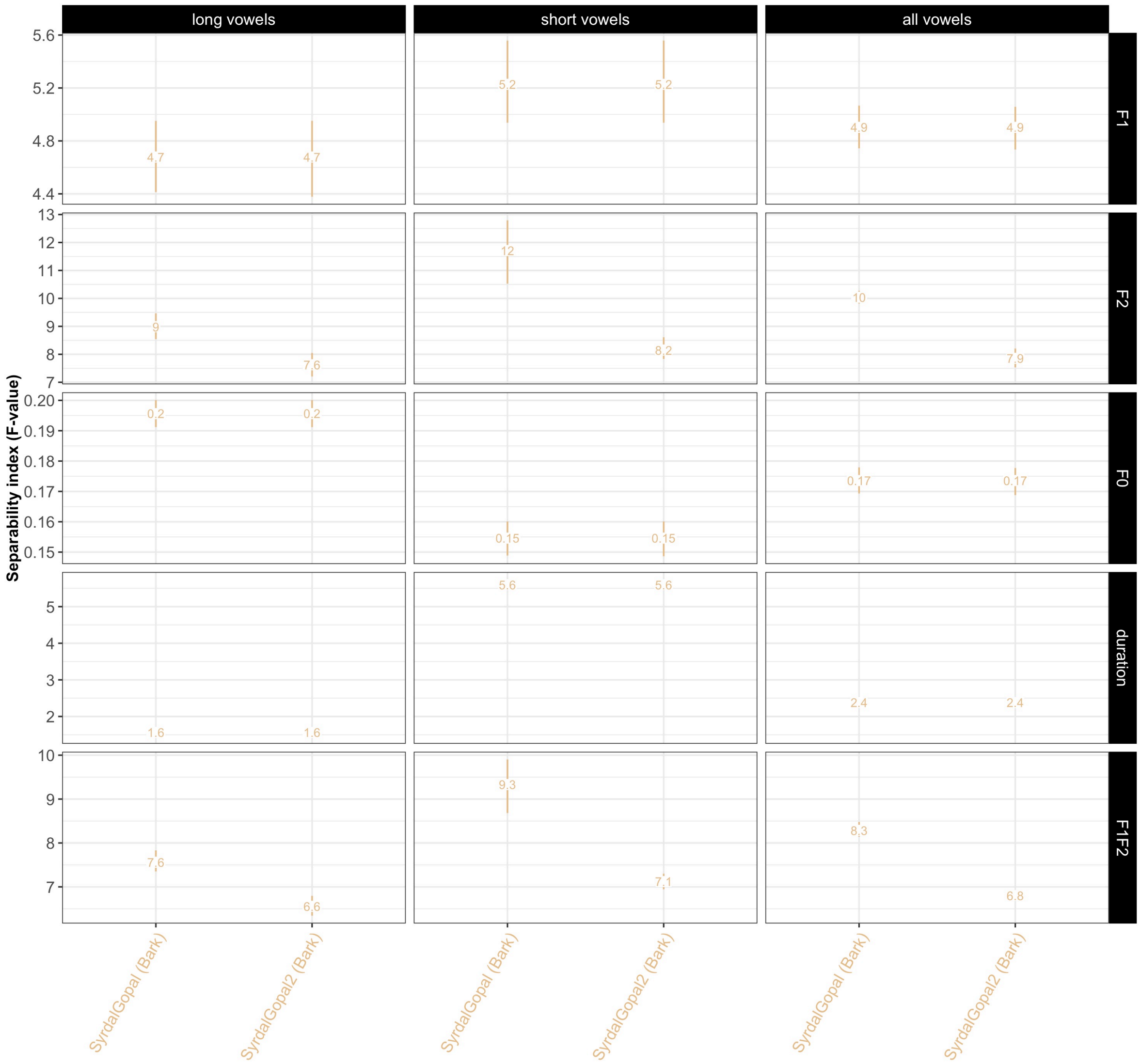

Figure 15. Separability indices of the two versions of the Syrdal & Gopal (1986) account for long vowels, short vowels, and long and short vowels together, shown for four of the five cues considered in this study and the combined F1-F2. Labels indicate mean across the five test folds. Intervals show average bootstrapped 95% confidence intervals across the test folds. Note that the ranges of the y-axes varies across plots.

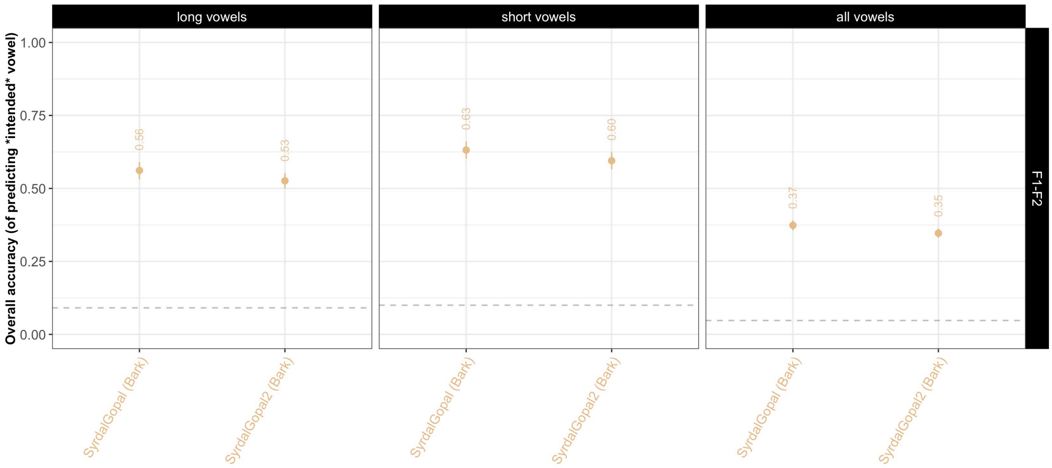

Figure 16. Predicted recognition accuracy of ideal observer under two versions of the Syrdal & Gopal (1986) account for long vowels, short vowels, and long and short vowels together, shown for the F1-F2 cue combination. Labels indicate mean across the five test folds. Intervals show average bootstrapped 95% confidence intervals across the test folds. The dashed horizontal line indicates chance (different across columns because of the different number of long and short vowels).

the F2-F1 implementation is more suitable for the materials used here, we therefore decided to use the first implementation throughout this paper.

Correlation matrices for all normalization accounts

Here we include correlation matrices for the SwehVd vowel data, transformed into the 15 different normalization spaces.

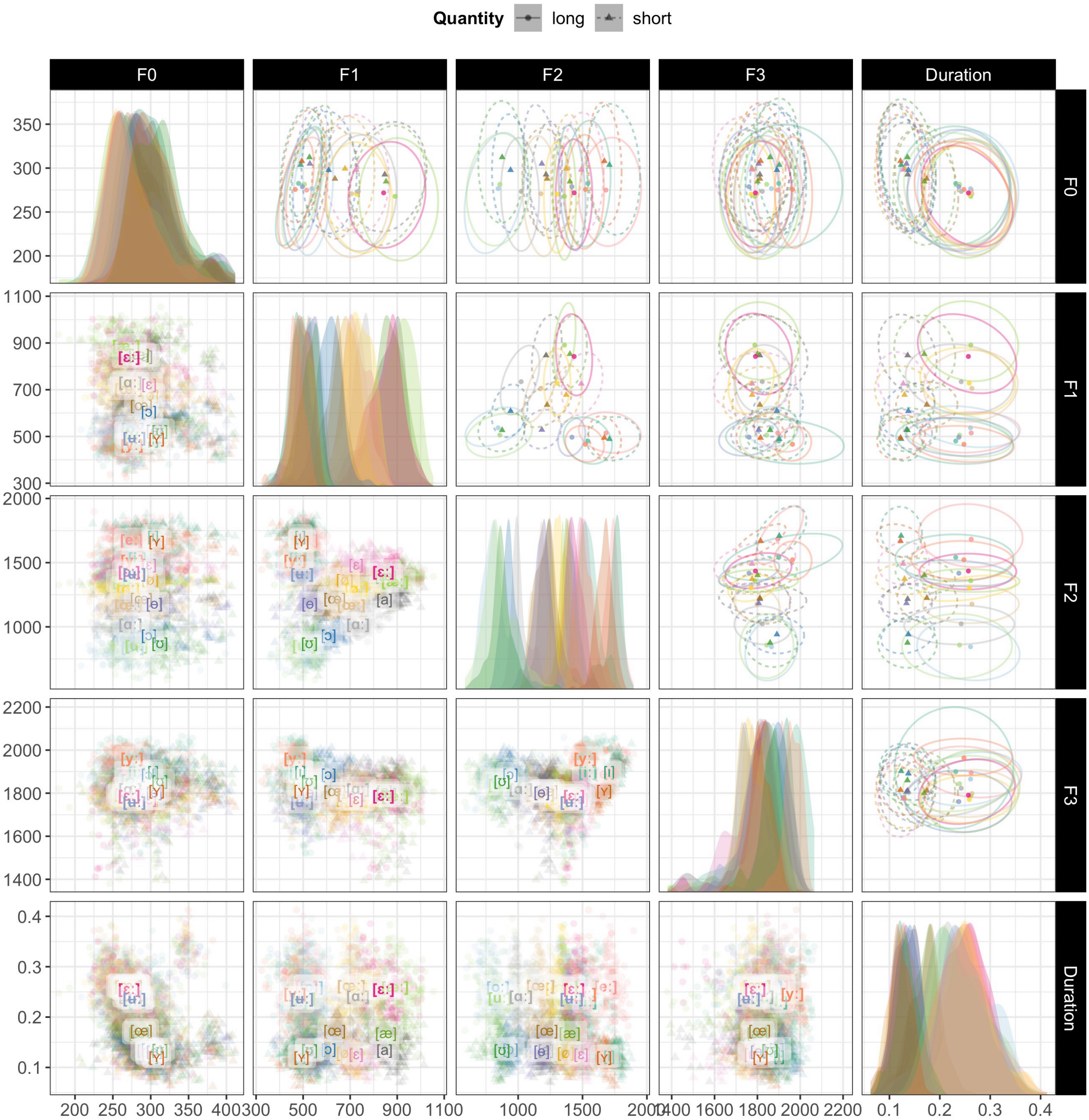

Figure 17. The SwehVd vowel data in Mel space. Points show repetitions of each of the 21 Central Swedish vowels by 16 female native talkers in the database in F0-F3 and vowel duration cue space. Vowel labels indicate category means across talkers. Long vowels are boldfaced. Ellipses show bivariate Gaussian 95% confidence interval of category means.

§3 ADDITIONAL INFORMATION FOR STUDY 2

Per-vowel categorization accuracy of models trained on long and short vowels separately

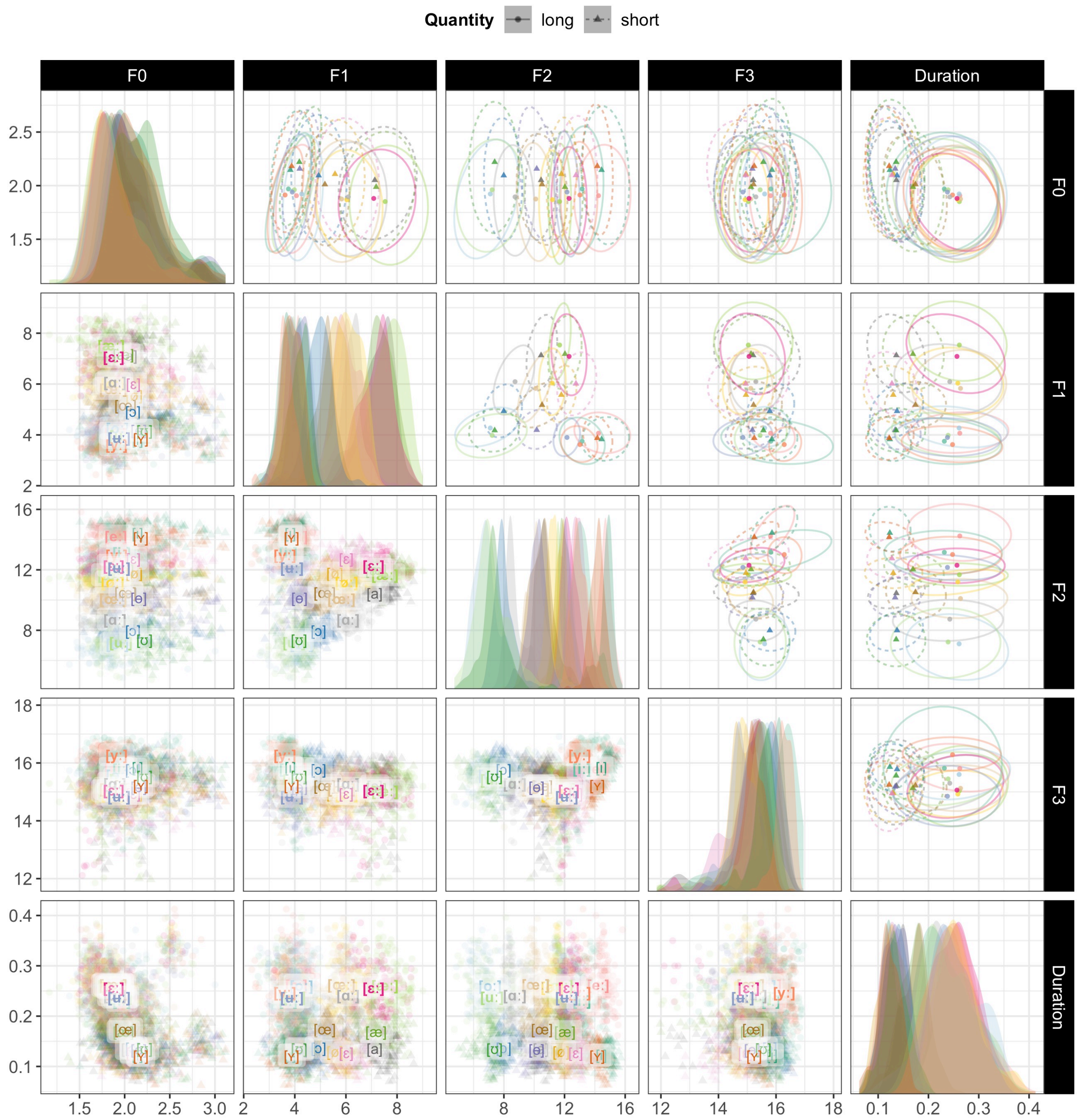

Figure 18. The SwehVd vowel data in Bark space. Points show repetitions of each of the 21 Central Swedish vowels by 16 female native talkers in the database in F0-F3 and vowel duration cue space. Vowel labels indicate category means across talkers. Long vowels are boldfaced. Ellipses show bivariate Gaussian 95% confidence interval of category means.

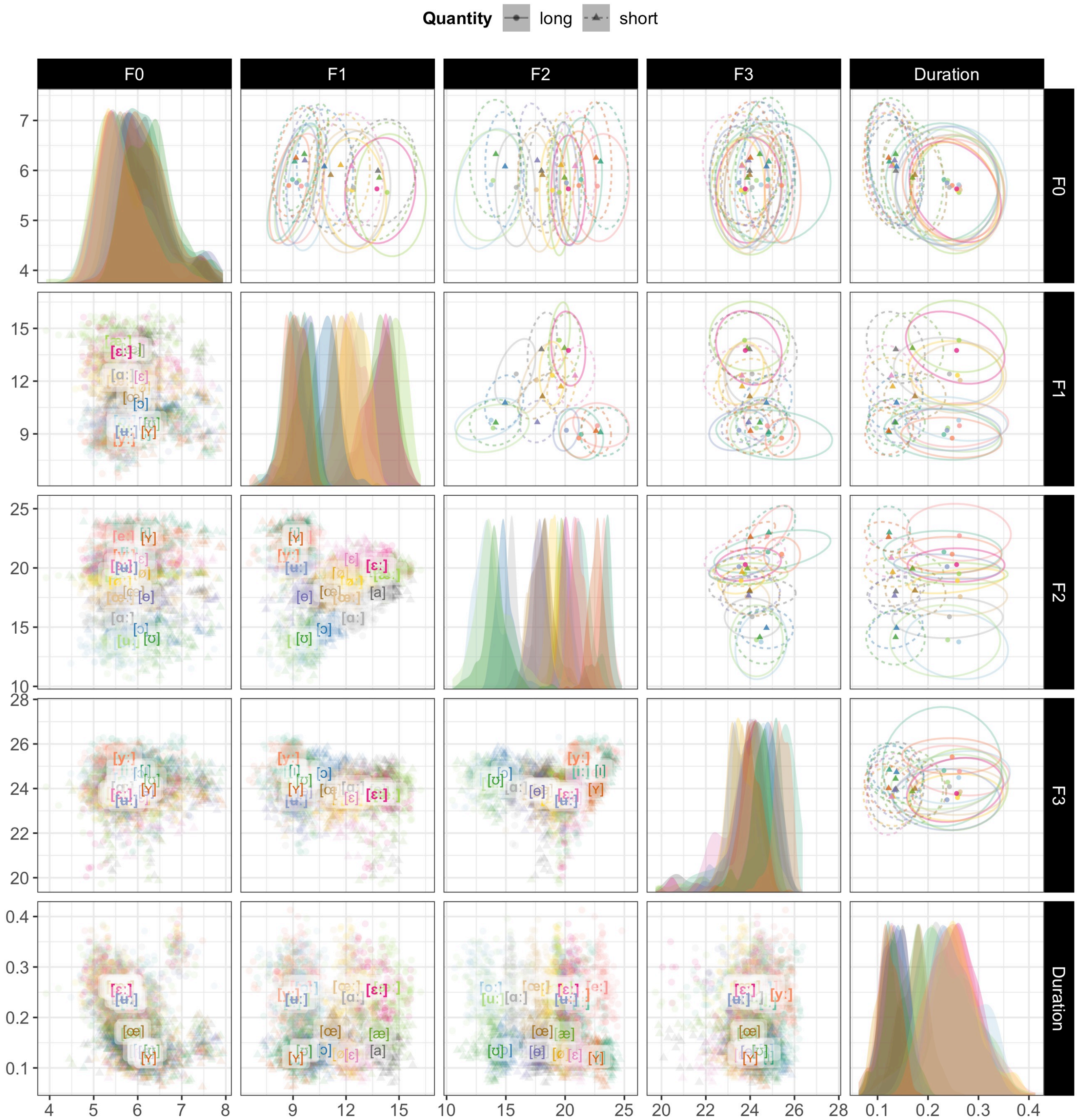

Figure 19. The SwehVd vowel data in ERB space. Points show repetitions of each of the 21 Central Swedish vowels by 16 female native talkers in the database in F0-F3 and vowel duration cue space. Vowel labels indicate category means across talkers. Long vowels are boldfaced. Ellipses show bivariate Gaussian 95% confidence interval of category means.

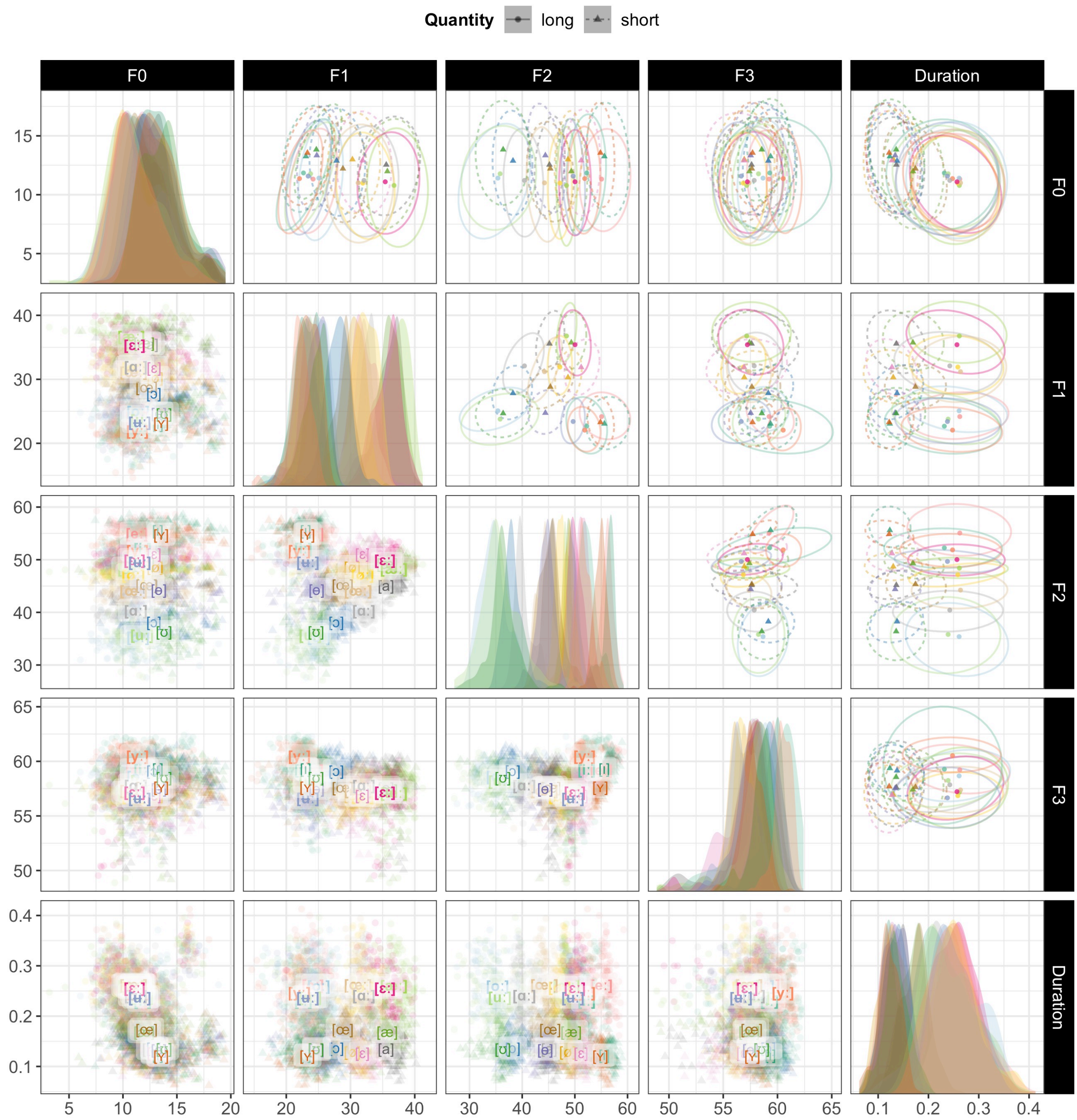

Figure 20. The SwehVd vowel data in semitones space. Points show repetitions of each of the 21 Central Swedish vowels by 16 female native talkers in the database in F0-F3 and vowel duration cue space. Vowel labels indicate category means across talkers. Long vowels are boldfaced. Ellipses show bivariate Gaussian 95% confidence interval of category means.

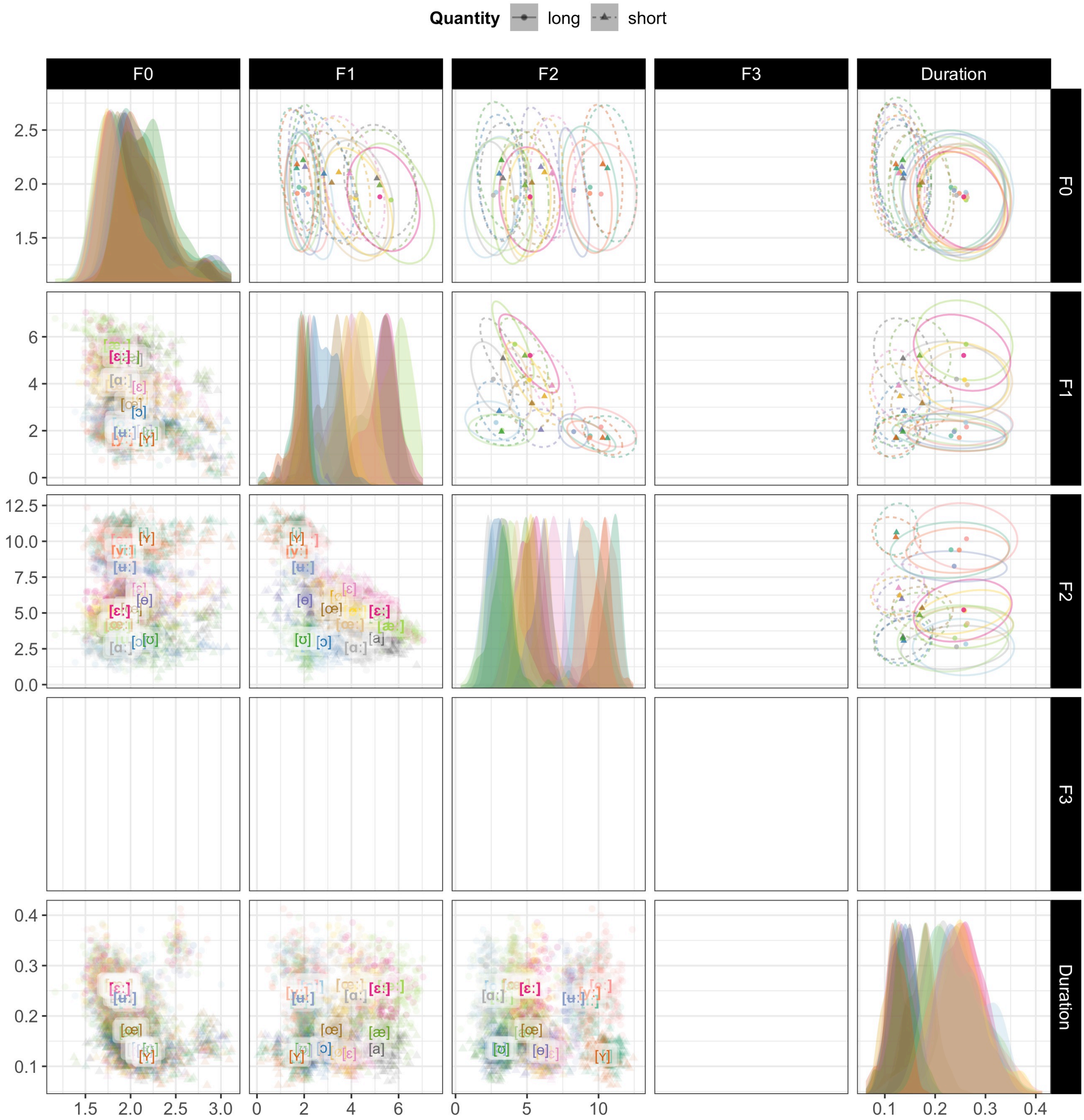

Figure 21. The SwehVd vowel data in SyrdalGopal (Bark) space. Points show repetitions of each of the 21 Central Swedish vowels by 16 female native talkers in the database in F0-F3 and vowel duration cue space. Vowel labels indicate category means across talkers. Long vowels are boldfaced. Ellipses show bivariate Gaussian 95% confidence interval of category means.

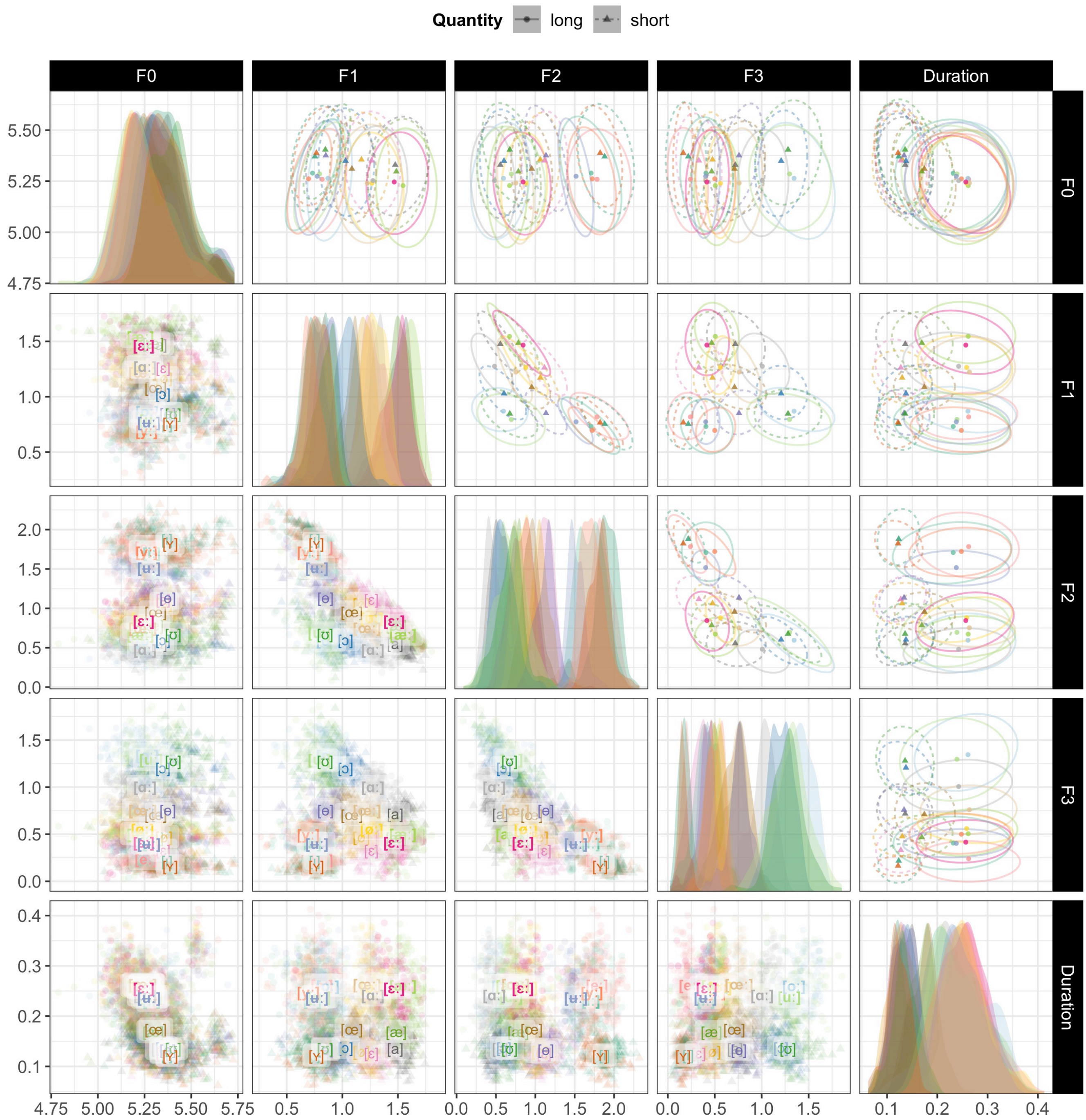

Figure 22. The SwehVd vowel data in Miller (log) space. Points show repetitions of each of the 21 Central Swedish vowels by 16 female native talkers in the database in F0-F3 and vowel duration cue space. Vowel labels indicate category means across talkers. Long vowels are boldfaced. Ellipses show bivariate Gaussian 95% confidence interval of category means.

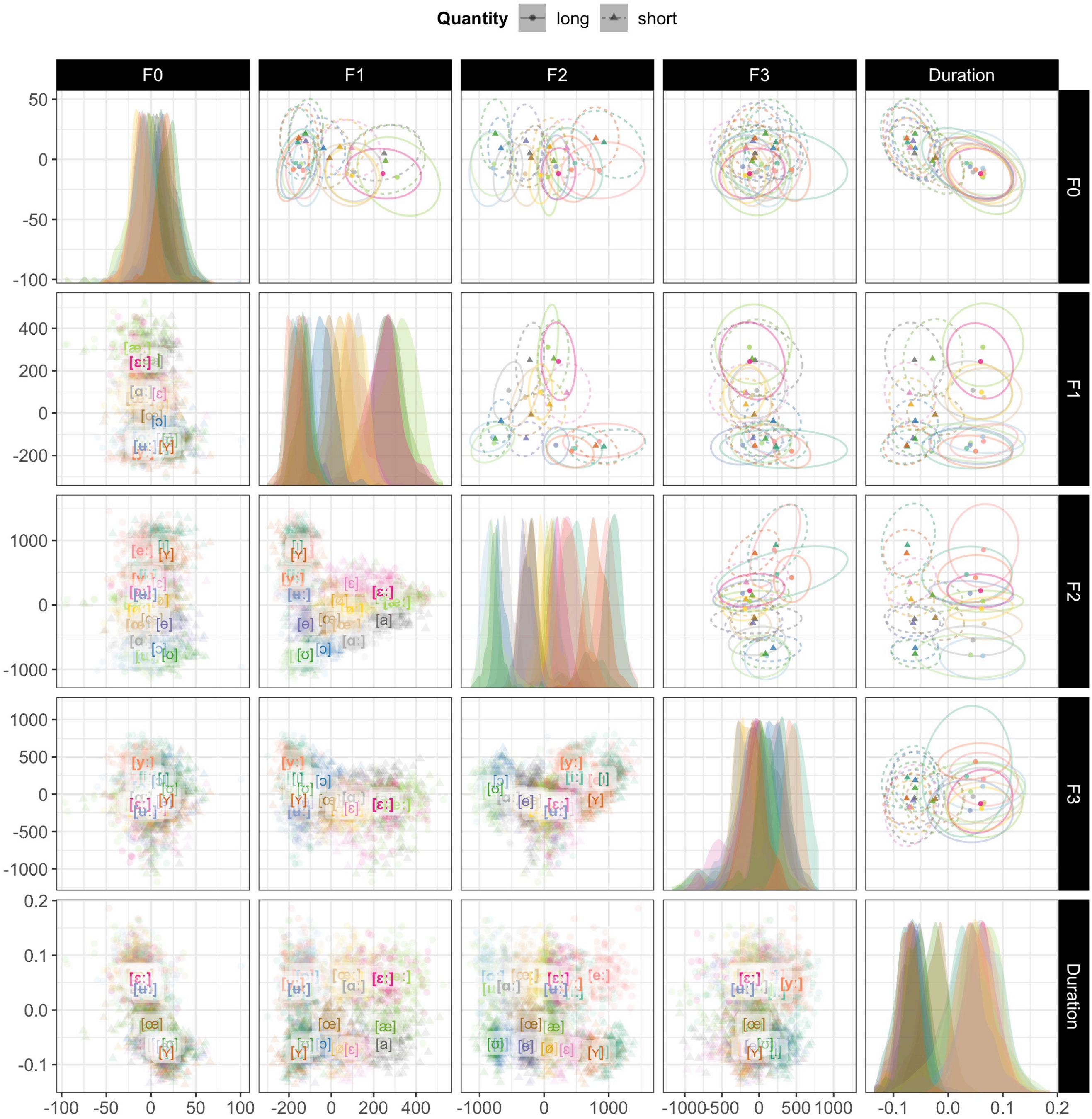

Figure 23. The SwehVd vowel data in C-CuRE Hz space. Points show repetitions of each of the 21 Central Swedish vowels by 16 female native talkers in the database in F0-F3 and vowel duration cue space. Vowel labels indicate category means across talkers. Long vowels are boldfaced. Ellipses show bivariate Gaussian 95% confidence interval of category means.

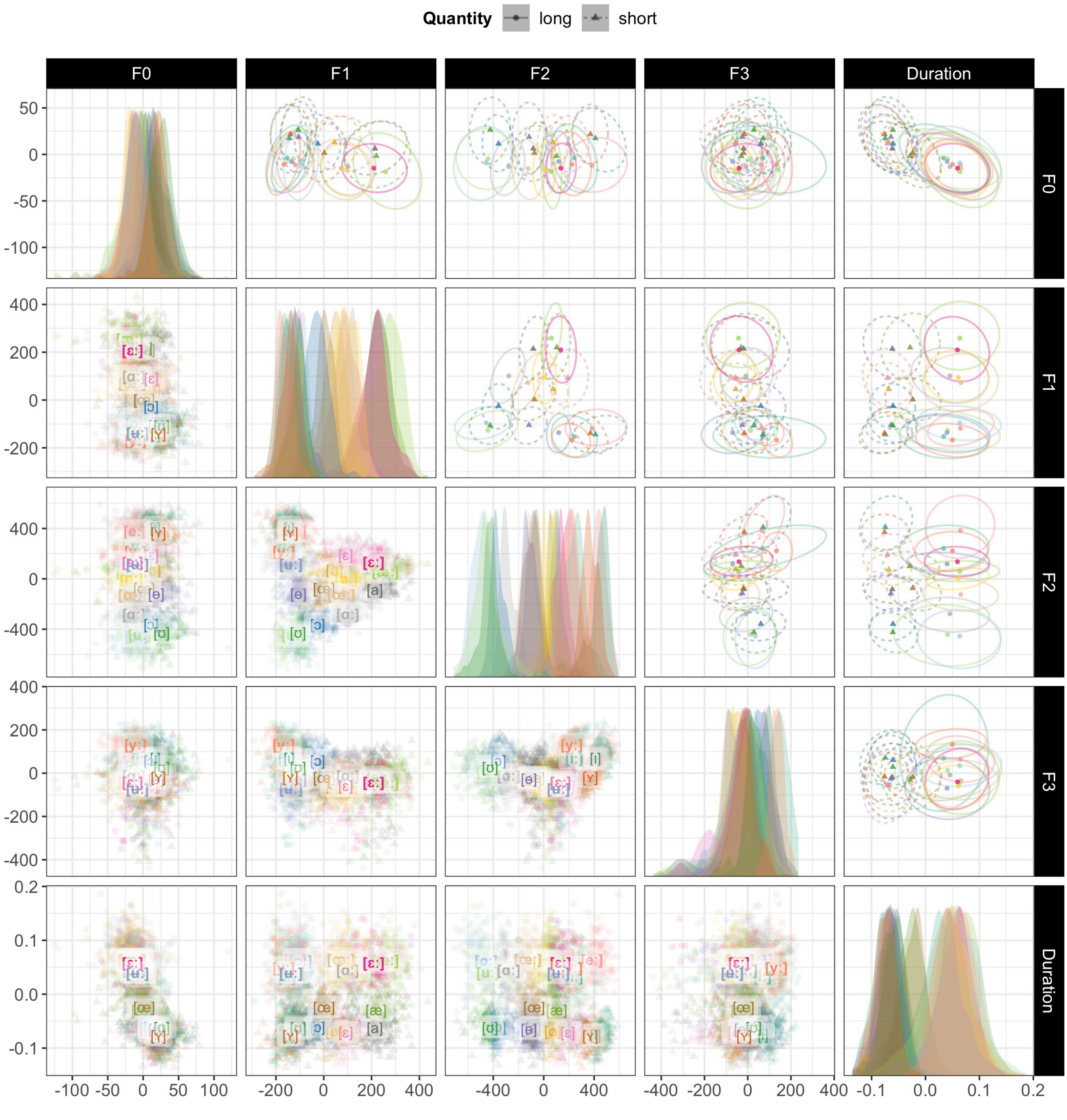

Figure 24. The SwehVd vowel data in C-CuRE Mel space. Points show repetitions of each of the 21 Central Swedish vowels by 16 female native talkers in the database in F0-F3 and vowel duration cue space. Vowel labels indicate category means across talkers. Long vowels are boldfaced. Ellipses show bivariate Gaussian 95% confidence interval of category means.

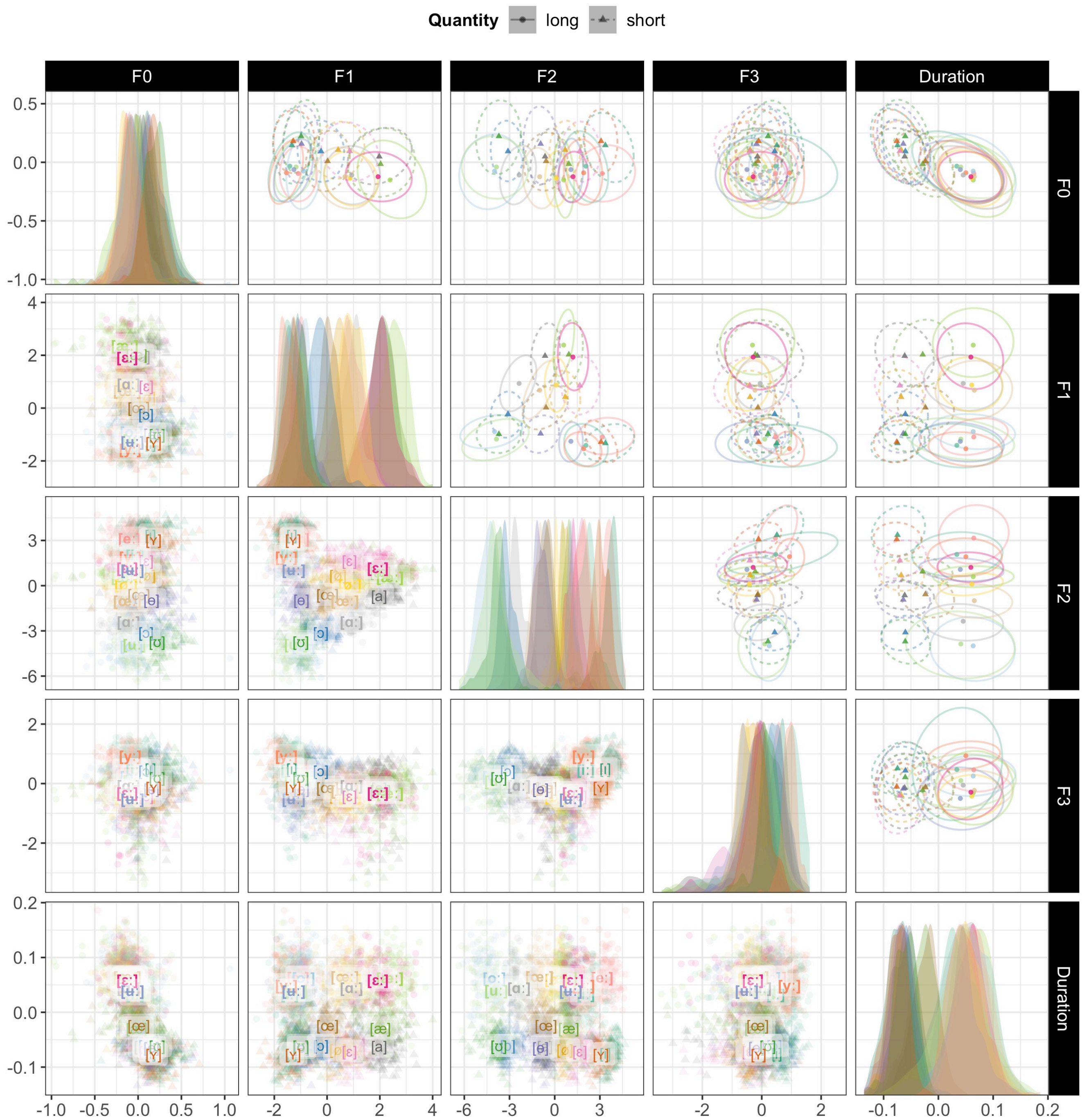

Figure 25. The SwehVd vowel data in C-CuRE Bark space. Points show repetitions of each of the 21 Central Swedish vowels by 16 female native talkers in the database in F0-F3 and vowel duration cue space. Vowel labels indicate category means across talkers. Long vowels are boldfaced. Ellipses show bivariate Gaussian 95% confidence interval of category means.

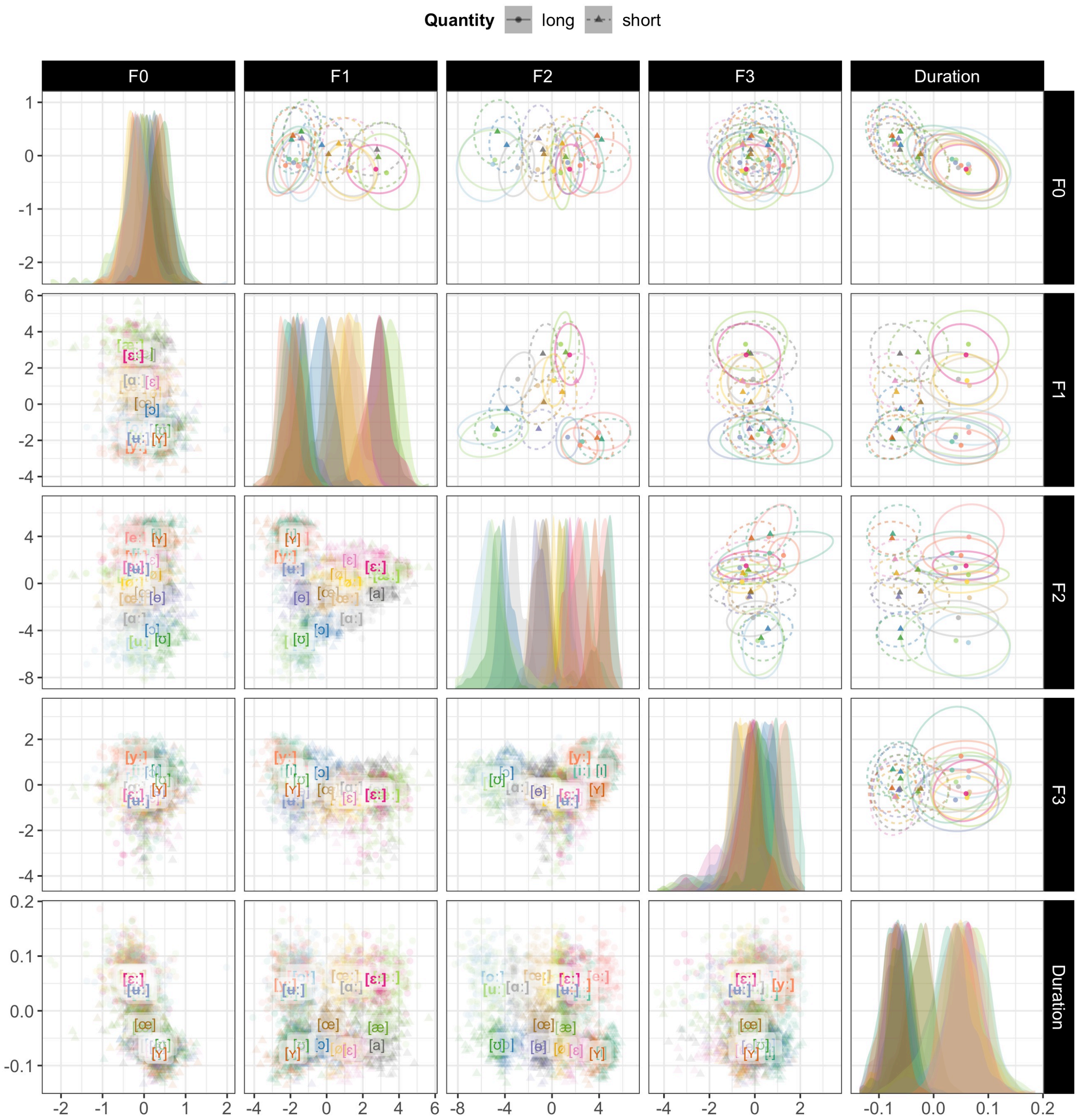

Figure 26. The SwehVd vowel data in C-CuRE ERB space. Points show repetitions of each of the 21 Central Swedish vowels by 16 female native talkers in the database in F0-F3 and vowel duration cue space. Vowel labels indicate category means across talkers. Long vowels are boldfaced. Ellipses show bivariate Gaussian 95% confidence interval of category means.

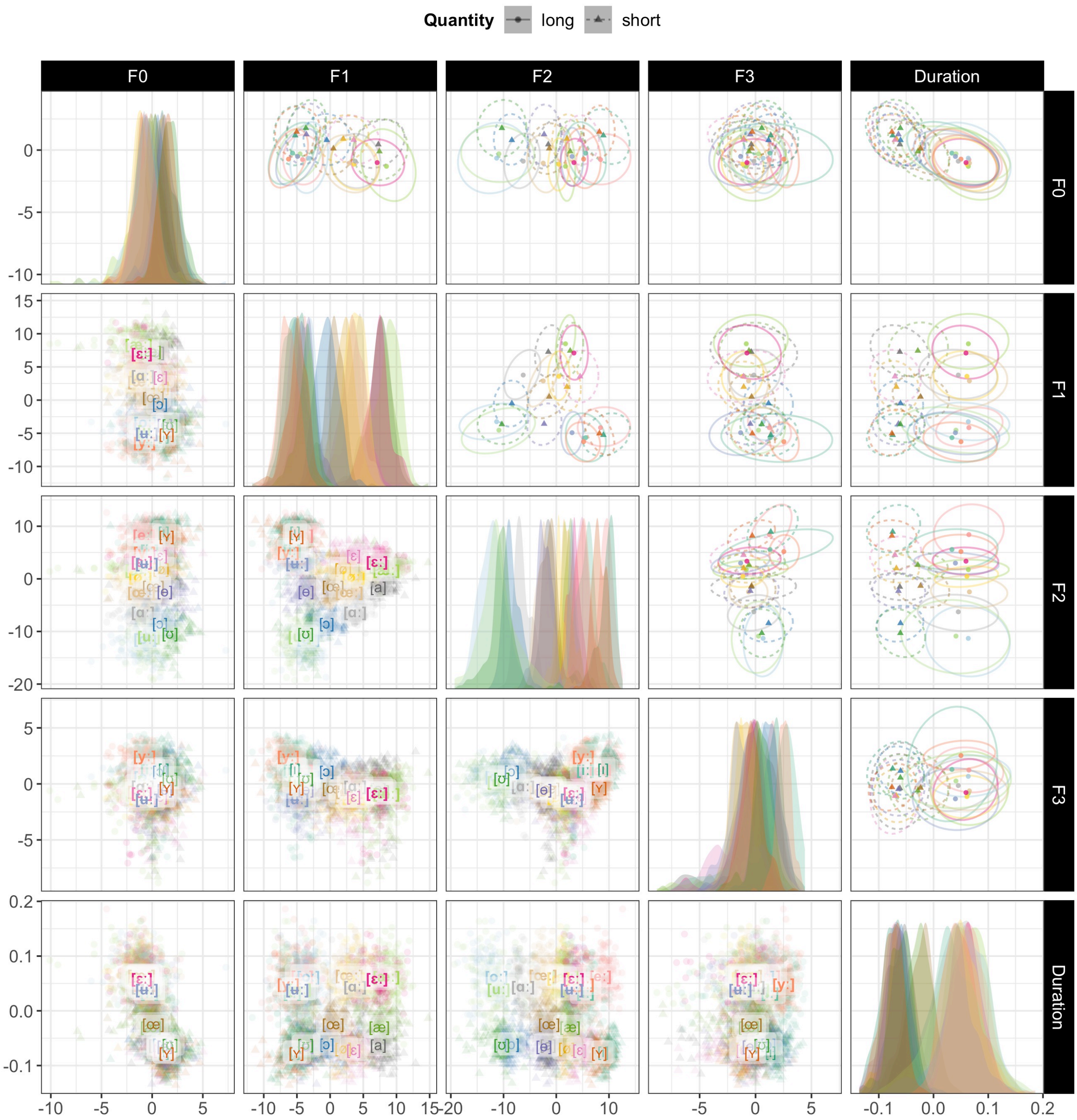

Figure 27. The SwehVd vowel data in C-CuRE semitones space. Points show repetitions of each of the 21 Central Swedish vowels by 16 female native talkers in the database in F0-F3 and vowel duration cue space. Vowel labels indicate category means across talkers. Long vowels are boldfaced. Ellipses show bivariate Gaussian 95% confidence interval of category means.

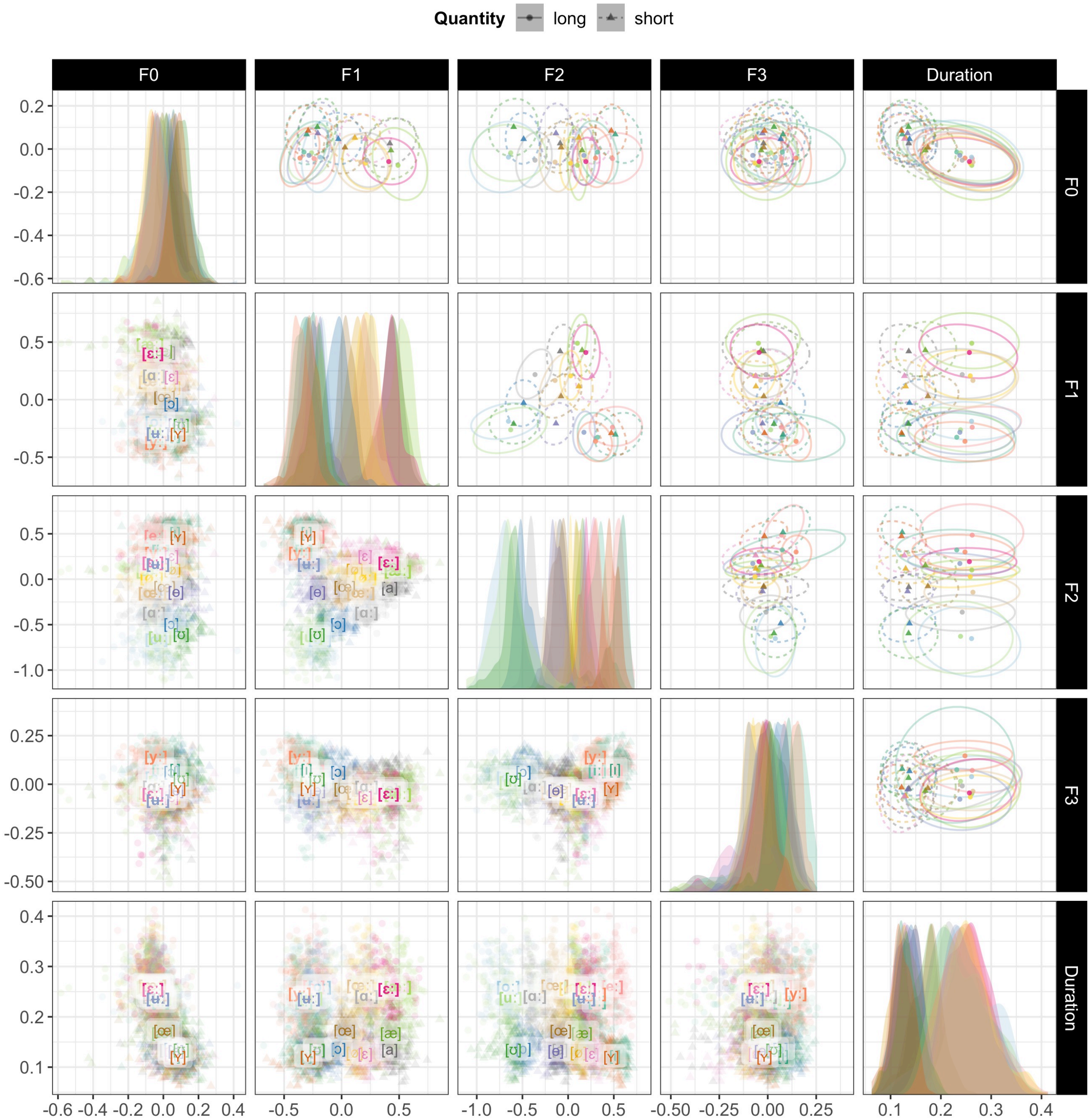

Figure 28. The SwehVd vowel data in Nearey1 (log) space. Points show repetitions of each of the 21 Central Swedish vowels by 16 female native talkers in the database in F0-F3 and vowel duration cue space. Vowel labels indicate category means across talkers. Long vowels are boldfaced. Ellipses show bivariate Gaussian 95% confidence interval of category means.

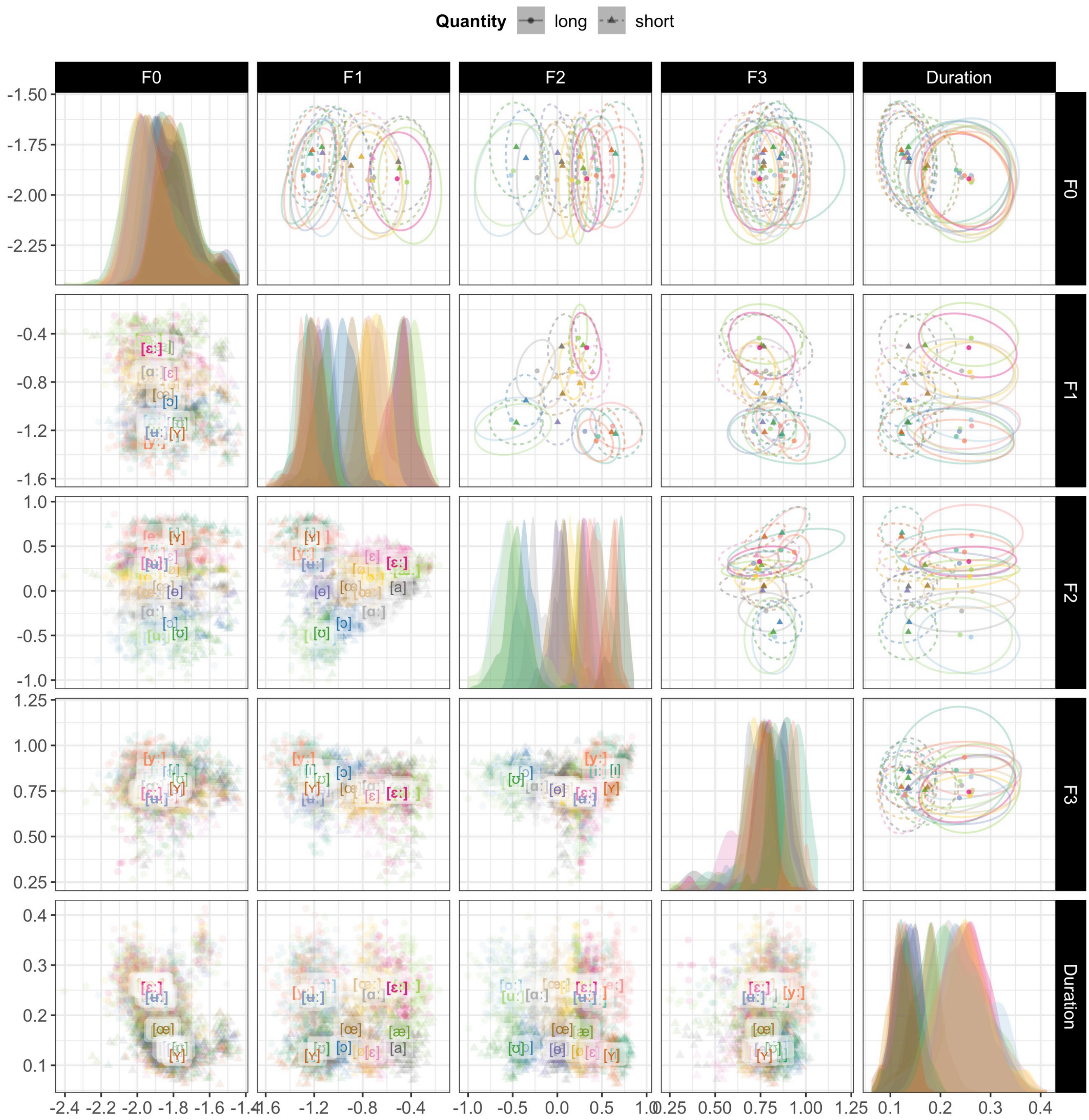

Figure 29. The SwehVd vowel data in Nearey2 (log) space. Points show repetitions of each of the 21 Central Swedish vowels by 16 female native talkers in the database in F0-F3 and vowel duration cue space. Vowel labels indicate category means across talkers. Long vowels are boldfaced. Ellipses show bivariate Gaussian 95% confidence interval of category means.

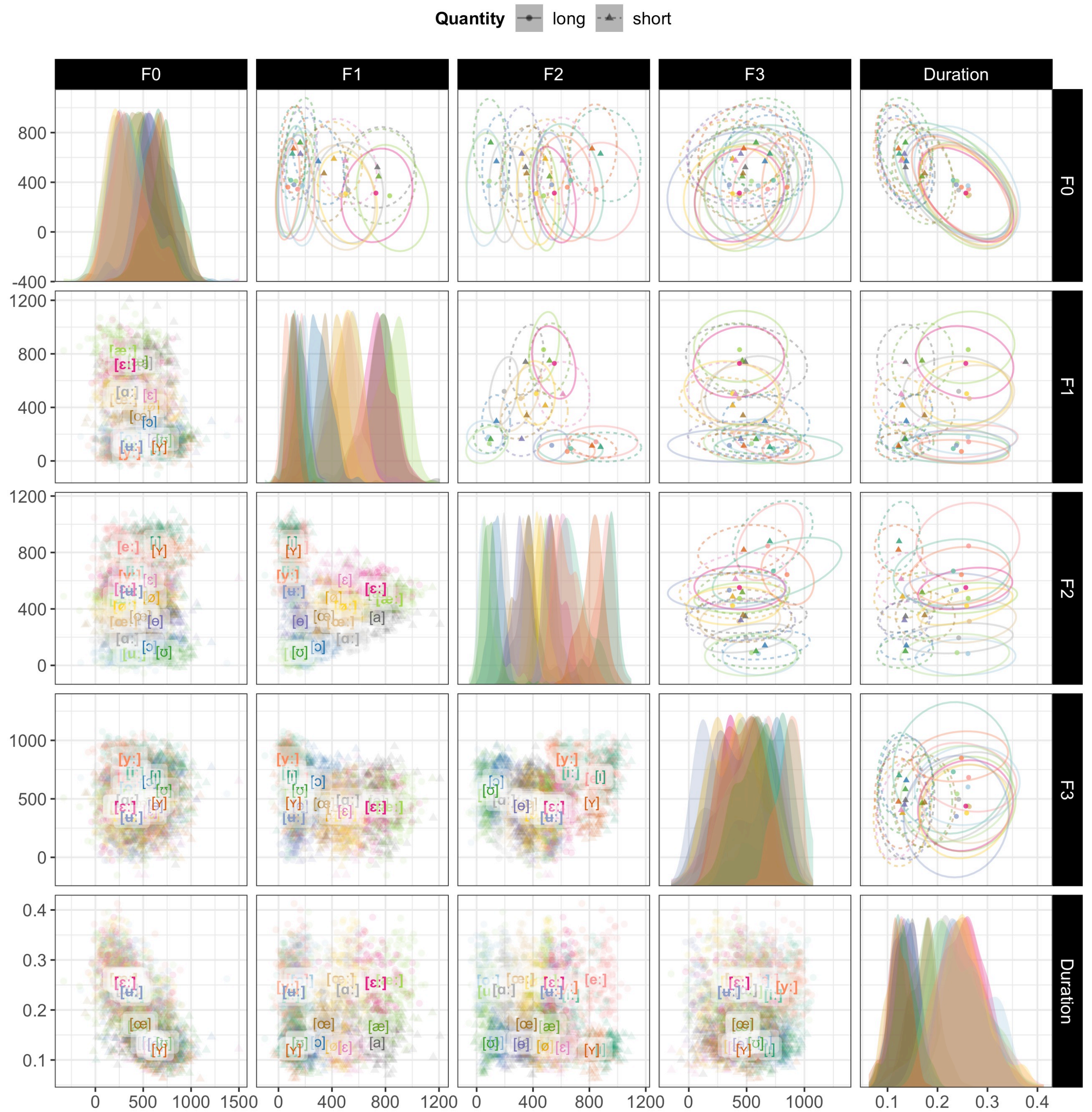

Figure 30. The SwehVd vowel data in Gerstman (Hz) space. Points show repetitions of each of the 21 Central Swedish vowels by 16 female native talkers in the database in F0-F3 and vowel duration cue space. Vowel labels indicate category means across talkers. Long vowels are boldfaced. Ellipses show bivariate Gaussian 95% confidence interval of category means.

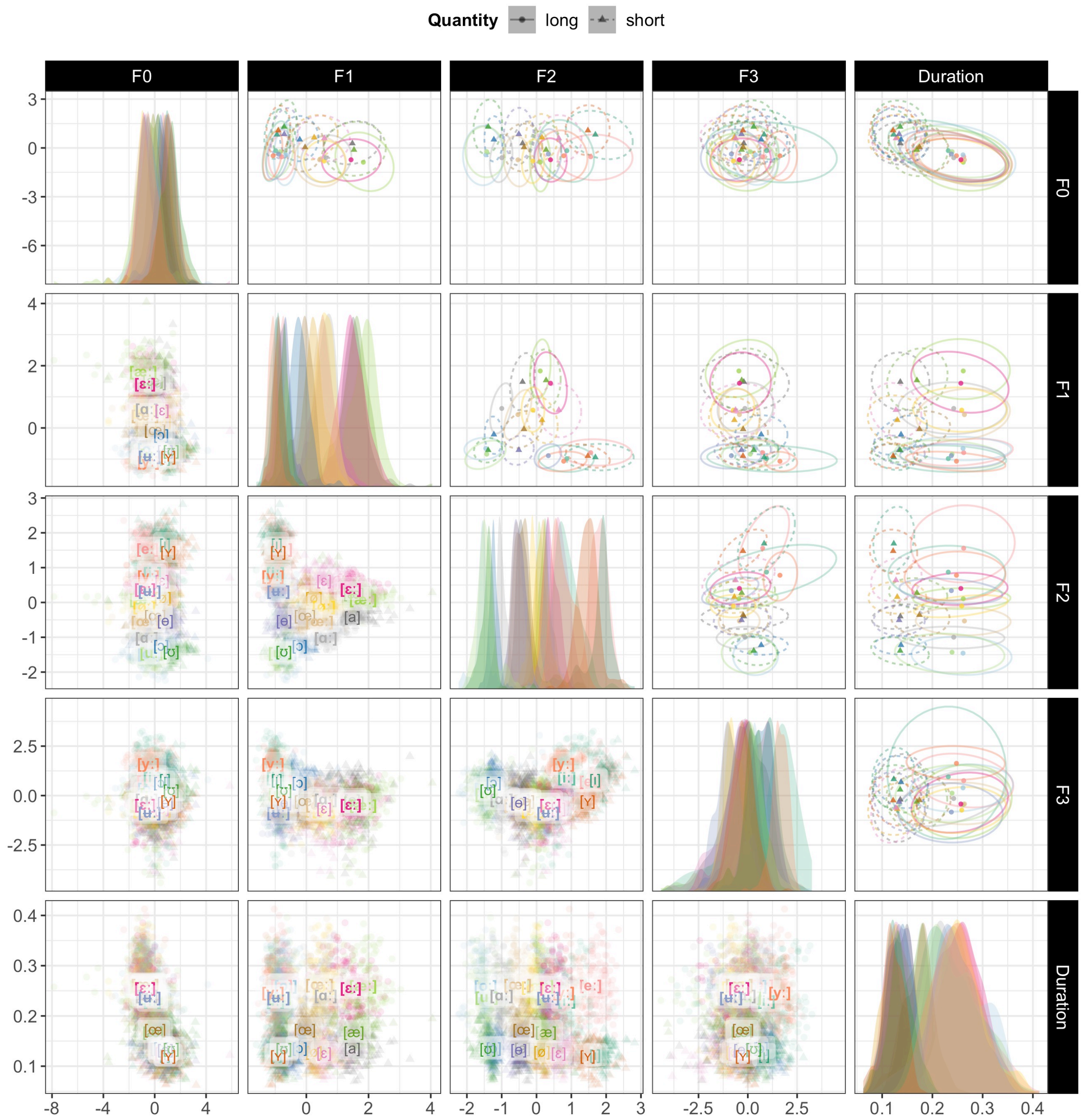

Figure 31. The SwehVd vowel data in Lobanov (Hz) space. Points show repetitions of each of the 21 Central Swedish vowels by 16 female native talkers in the database in F0-F3 and vowel duration cue space. Vowel labels indicate category means across talkers. Long vowels are boldfaced. Ellipses show bivariate Gaussian 95% confidence interval of category means.

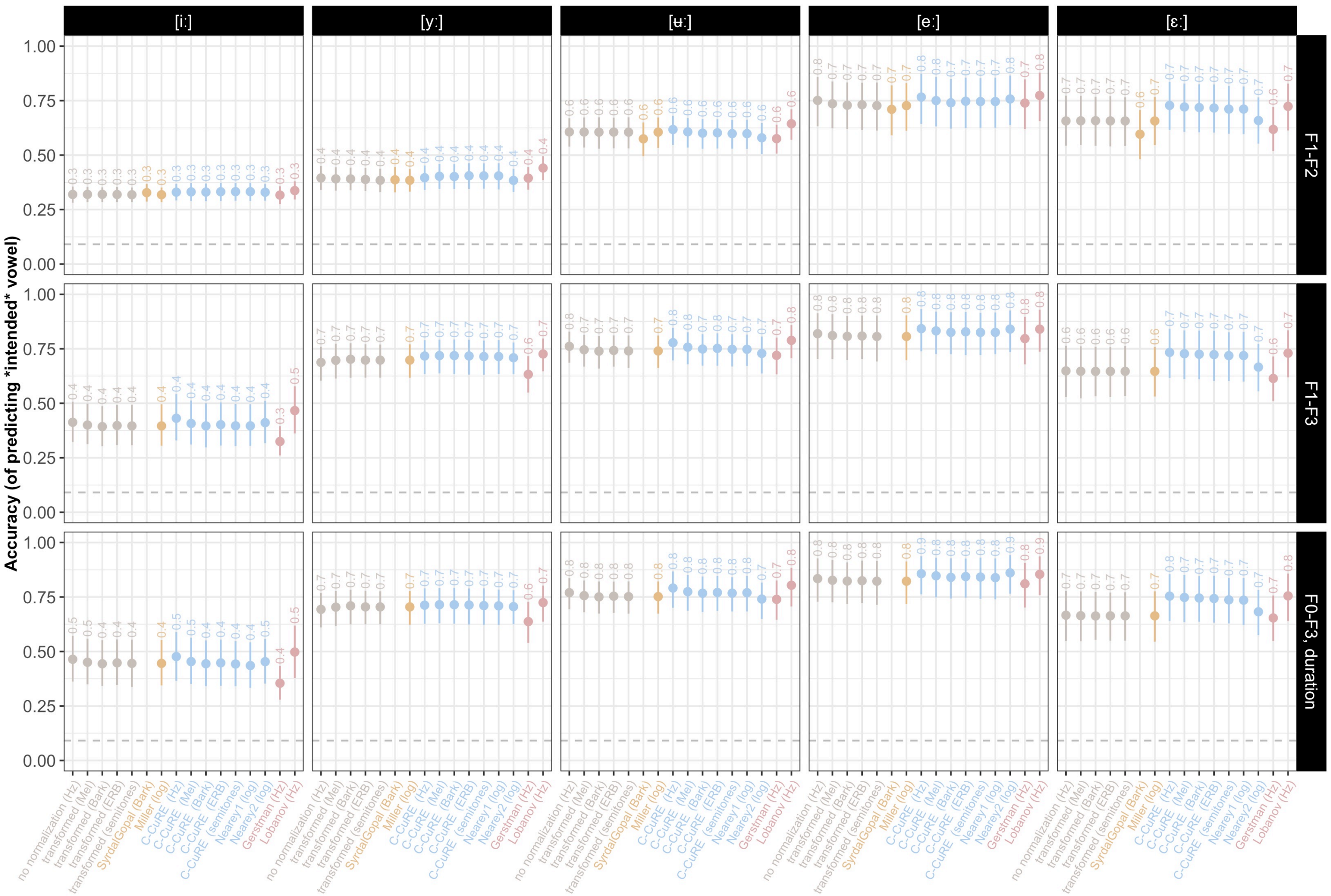

Figure 32. Per-vowel predicted categorization accuracy of the ideal observers trained on the **long** vowels, under different assumptions about the relevant cues. Point ranges indicate the average mean accuracy and average 95% bootstrapped CI across the five folds. Chance level is indicated by grey line.

Figure 33. (*Continued from last page*)Per-vowel predicted categorization accuracy of the ideal observers trained on the **long** vowels, under different assumptions about the relevant cues. Point ranges indicate the average mean accuracy and average 95% bootstrapped CI across the five folds. Chance level is indicated by grey line.

Confusion and difference matrices of ideal observers

To further explore effects of neighbouring categories, and which categories are more easily confused by the models and with what, we plot confusion matrices of the worst and best performing models trained on the long, short or all Central Swedish vowels, under the different assumptions about the relevant cues. Next to the confusion matrices, we plot difference matrices to facilitate comparison.

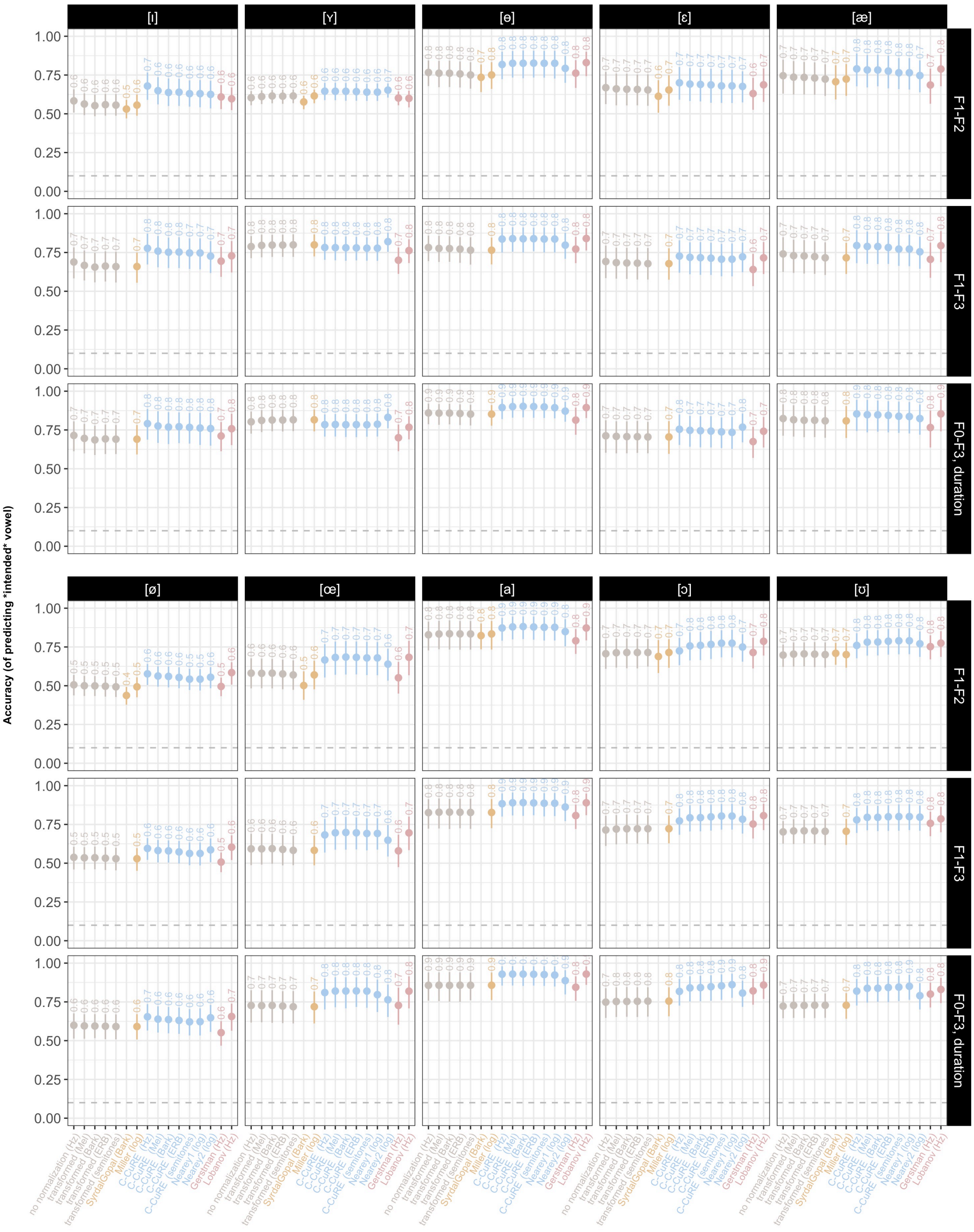

Figure 34. Per-vowel predicted categorization accuracy of the ideal observers trained on the **short** vowels, under different assumptions about the relevant cues. Point ranges indicate the average mean accuracy and average 95% bootstrapped CI across the five folds. Chance level is indicated by grey line.

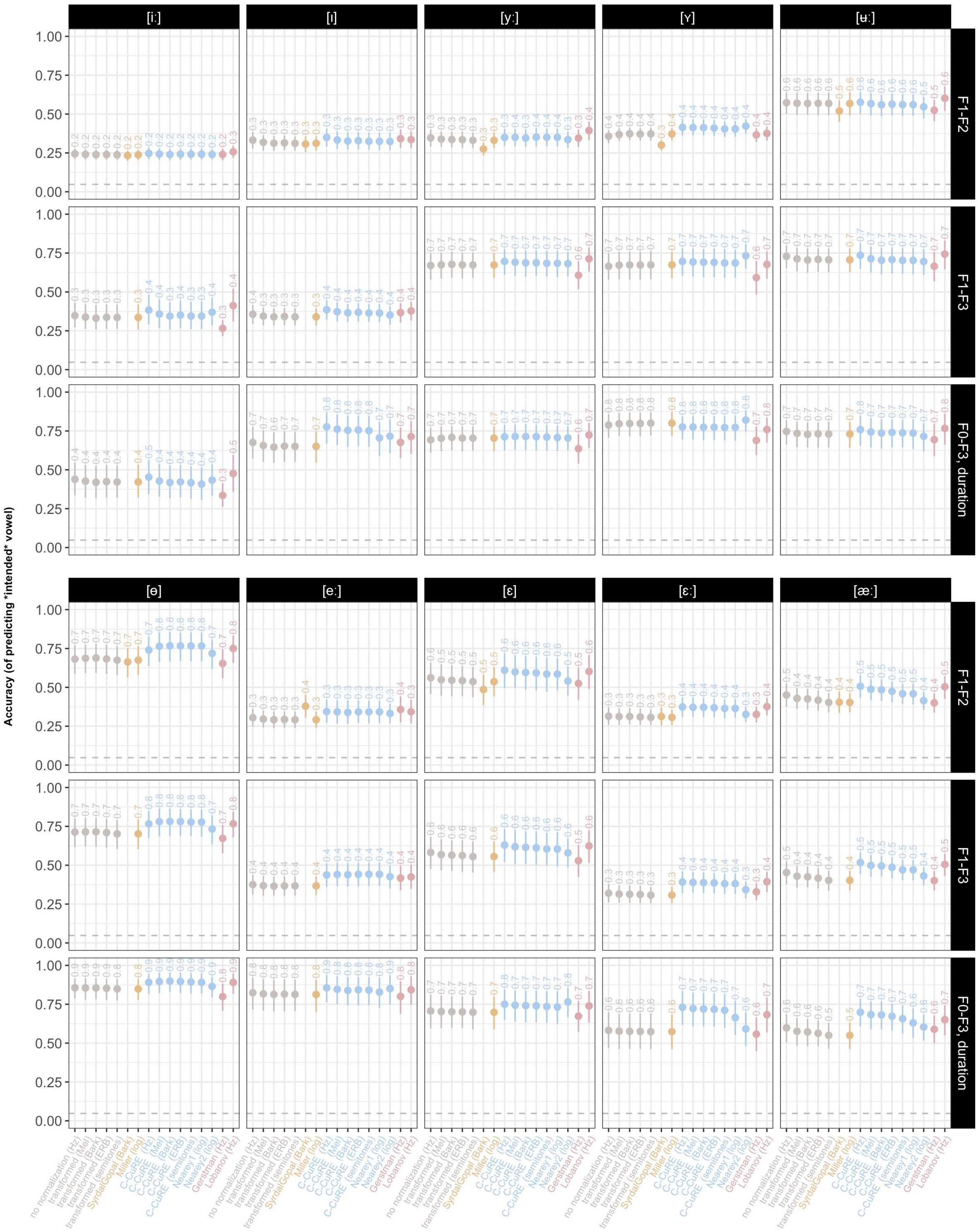

Figure 35. Per-vowel predicted categorization accuracy of the ideal observers trained on **all** vowels, under different assumptions about the relevant cues. Point ranges indicate the average mean accuracy and average 95% bootstrapped CI across the five folds. Chance level is indicated by grey line.

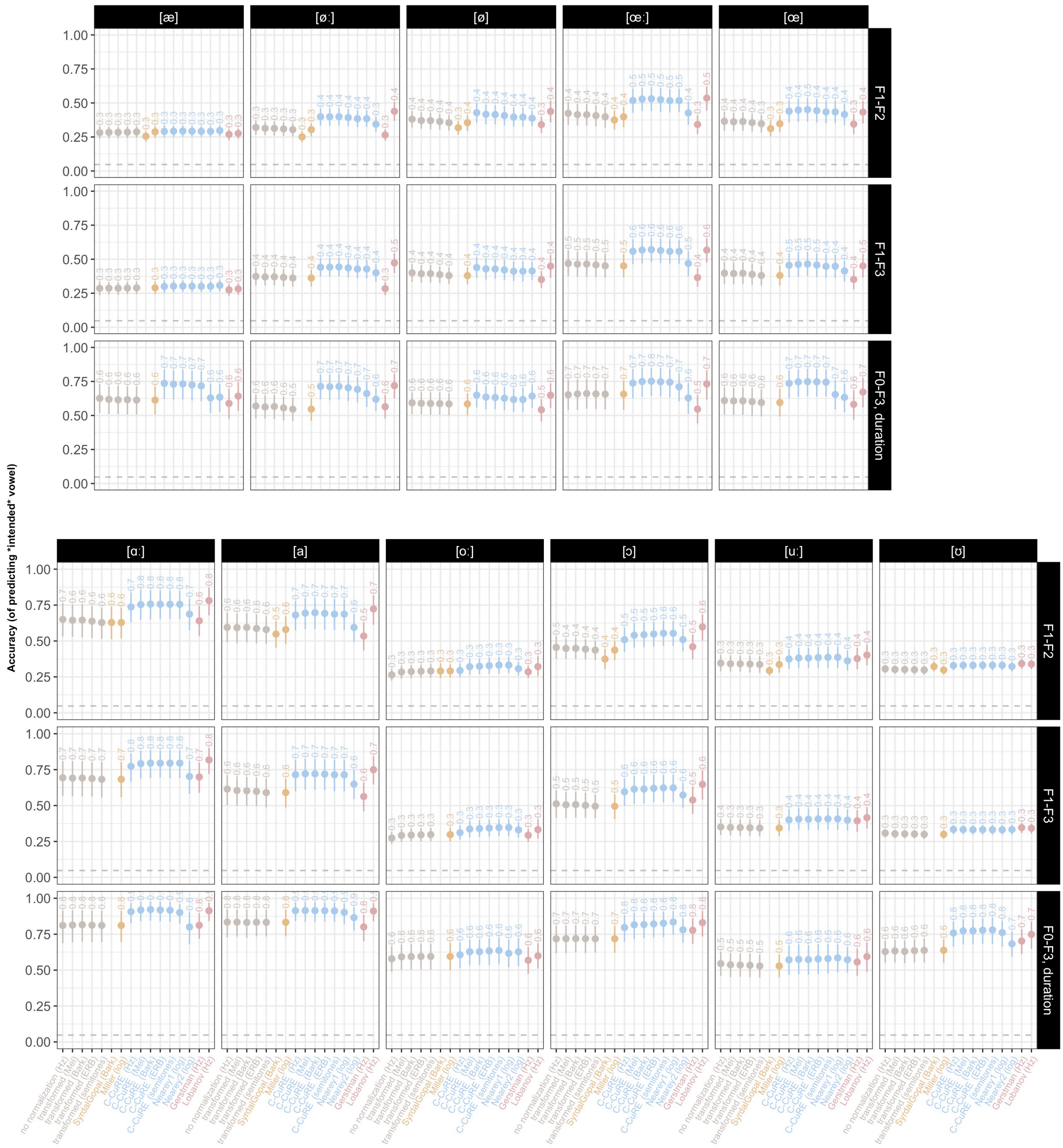

Figure 36. (*Continued from last page*)Per-vowel predicted categorization accuracy of the ideal observers trained on **all** vowels, under different assumptions about the relevant cues. Point ranges indicate the average mean accuracy and average 95% bootstrapped CI across the five folds. Chance level is indicated by grey line.

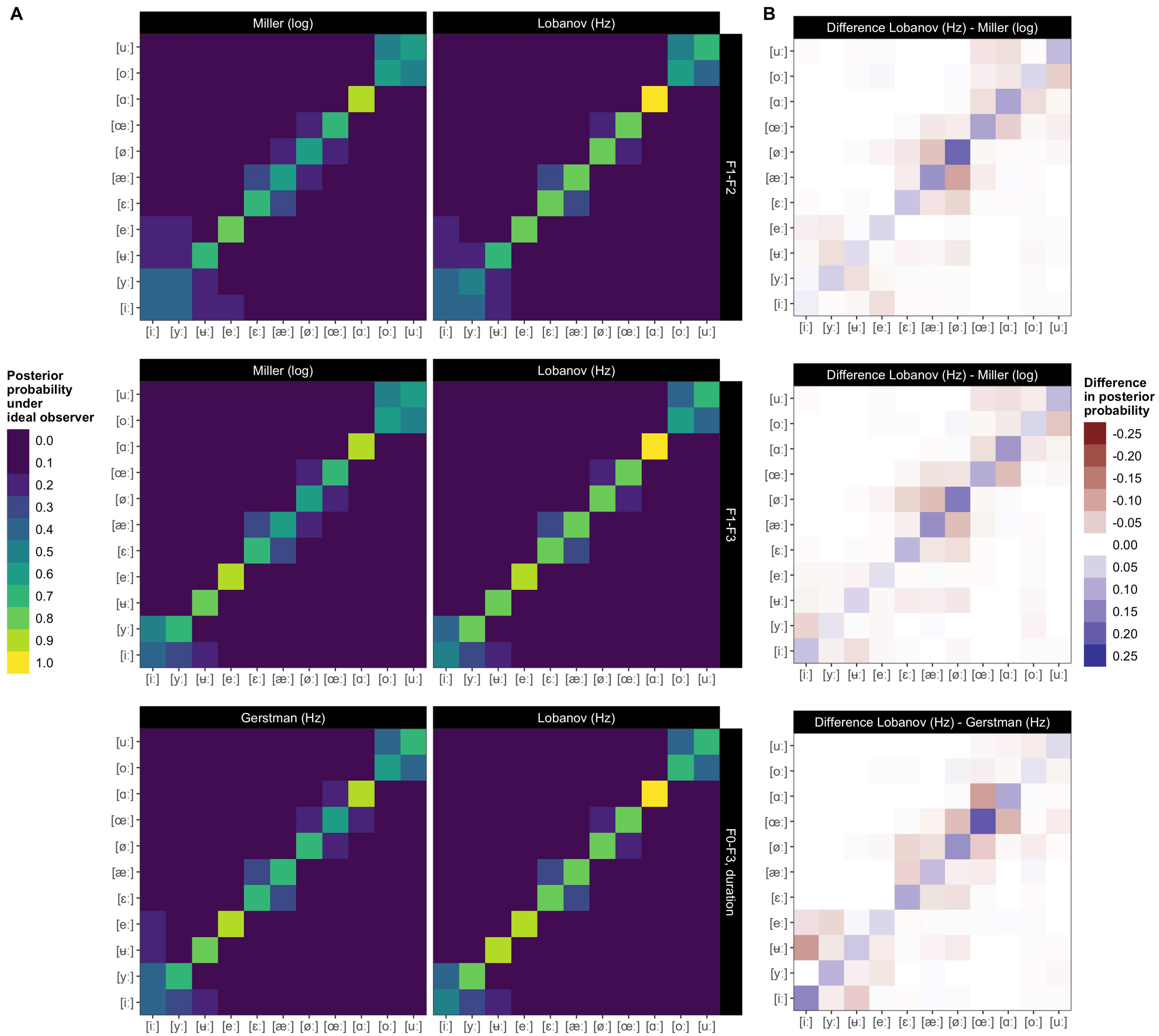

Figure 37. Illustration of the category-specific differences in predictions of the worst and best performing normalization models for each combination of cues (**rows**). The confusion matrices (**Panel A**) plot the predictions for the worst (**left**) and best (**right**) performing models in predicting the **long** vowels, under different assumptions about the relevant cues. Vowel intended by talker (x-axis) is plotted against vowel selected by ideal observer model (y-axis). Color fill indicates the posterior probability of the models predicting the intended vowel. The difference matrices (**Panel B**) illustrates the differences in predictions between the best and the worst performing models. Color fill indicates the difference in the posterior probability of the models predicting the intended vowel. More **purple** indicates an increase in posterior probability for the former over the latter model, more **red** indicates an advantage for the latter over the former.

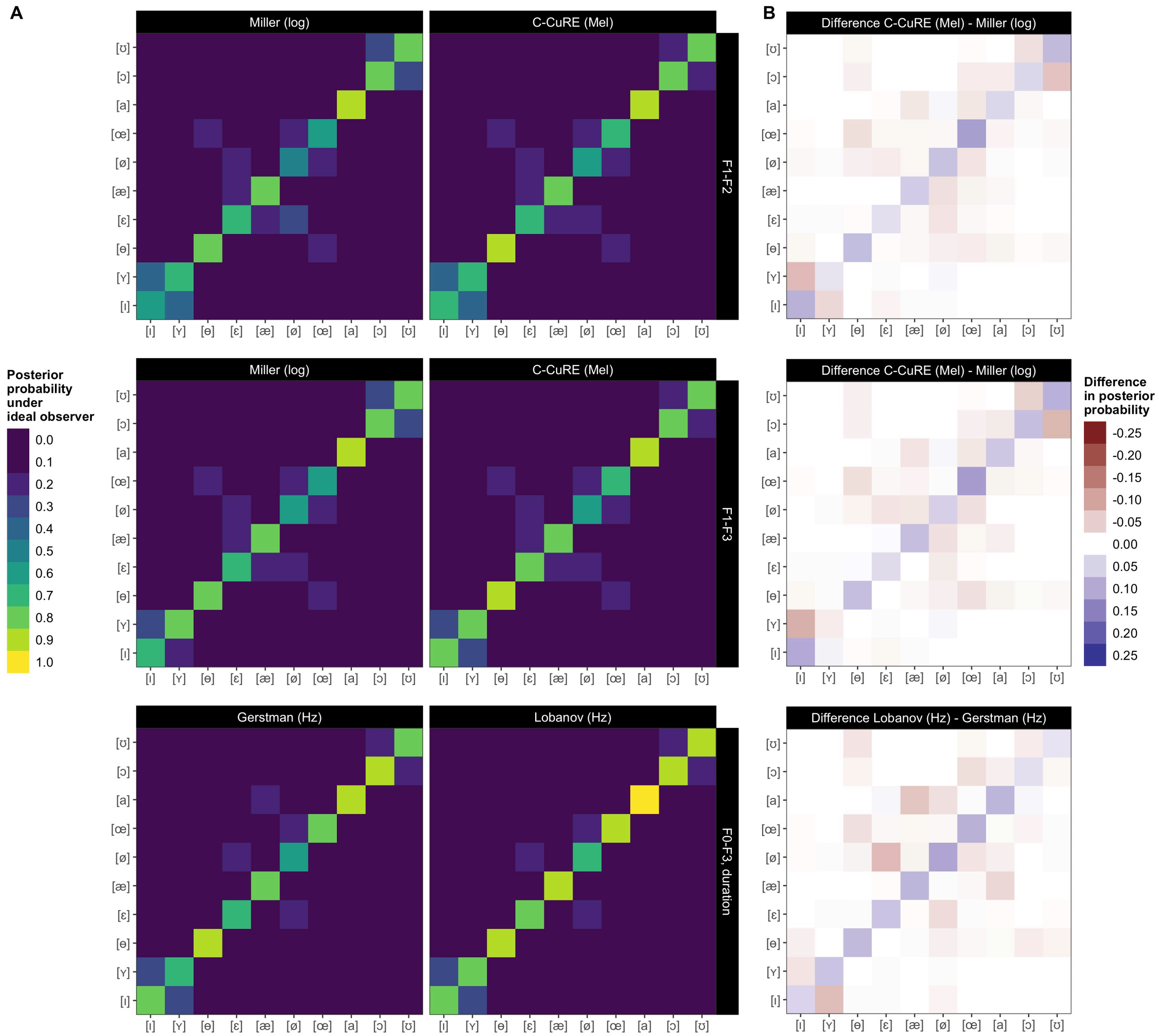

Figure 38. Illustration of the category-specific differences in predictions of the worst and best performing normalization models for each combination of cues (**rows**). The confusion matrices (**Panel A**) plot the predictions for the worst (**left**) and best (**right**) performing models in predicting the **short** vowels, under different assumptions about the relevant cues. Vowel intended by talker (x-axis) is plotted against vowel selected by ideal observer model (y-axis). Color fill indicates the posterior probability of the models predicting the intended vowel. The difference matrices (**Panel B**) illustrates the differences in predictions between the best and the worst performing models. Color fill indicates the difference in the posterior probability of the models predicting the intended vowel. More **purple** indicates an increase in posterior probability for the former over the latter model, more **red** indicates an advantage for the latter over the former.

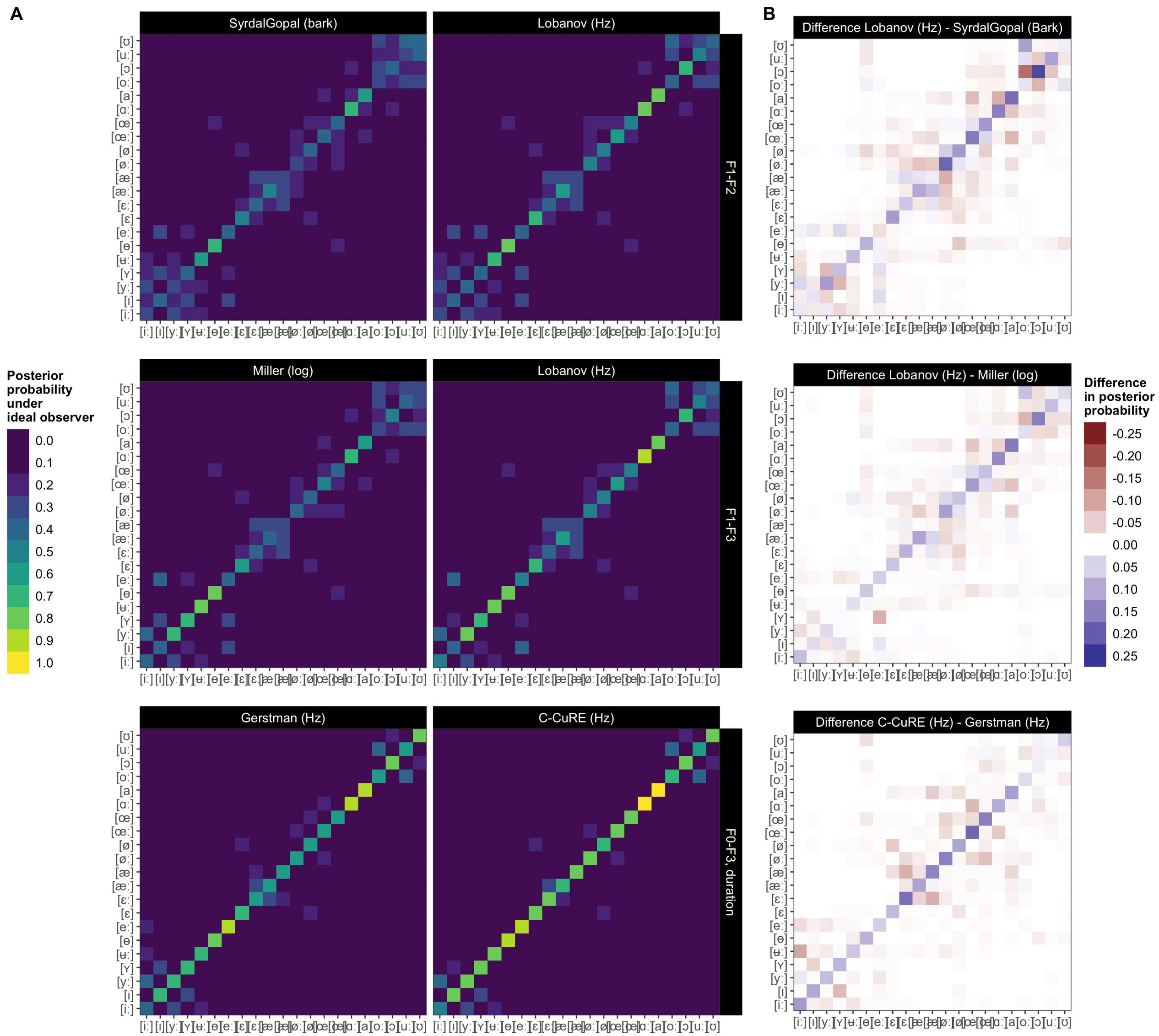

Figure 39. Illustration of the category-specific differences in predictions of the worst and best performing normalization models for each combination of cues (**rows**). The confusion matrices (**Panel A**) plot the predictions for the worst (**left**) and best (**right**) performing models in predicting the **all** vowels, under different assumptions about the relevant cues. Vowel intended by talker (x-axis) is plotted against vowel selected by ideal observer model (y-axis). Color fill indicates the posterior probability of the models predicting the intended vowel. The difference matrices (**Panel B**) illustrates the differences in predictions between the best and the worst performing models. Color fill indicates the difference in the posterior probability of the models predicting the intended vowel. More **purple** indicates an increase in posterior probability for the former over the latter model, more **red** indicates an advantage for the latter over the former.