

The perception of German wh-phras-final intonation: A contour clustering evaluation

Heiko Seeliger, Anne Lützeler, Constantijn Kaland
(University of Cologne)



- Seeliger & Kaland (2022):
 - Cluster analysis of F0 contours from a production study (Repp & Seeliger, *subm.*)
 - Target sentences: German *wh*-questions and *wh*-exclamatives
 - Two clusters of interest: Medium-high **plateaus**, some with very **late falls**
 - Form-wise, somewhere between continuation rise (**H-%**), calling contour (**L+H* !H-%**), and a non-standard **H-L%**
- This study:
 - Contour **similarity rating** by listeners, in order to confirm results

- Production study (read speech) on German *wh*-exclamatives and (embedded) *wh*-questions:

Exclamative: **Wo die schon überall Sirenen getestet hat!**

'The many places where she has already tested sirens!'

Question: (Weißt du zufällig,) **wo die schon überall Sirenen getestet hat?**

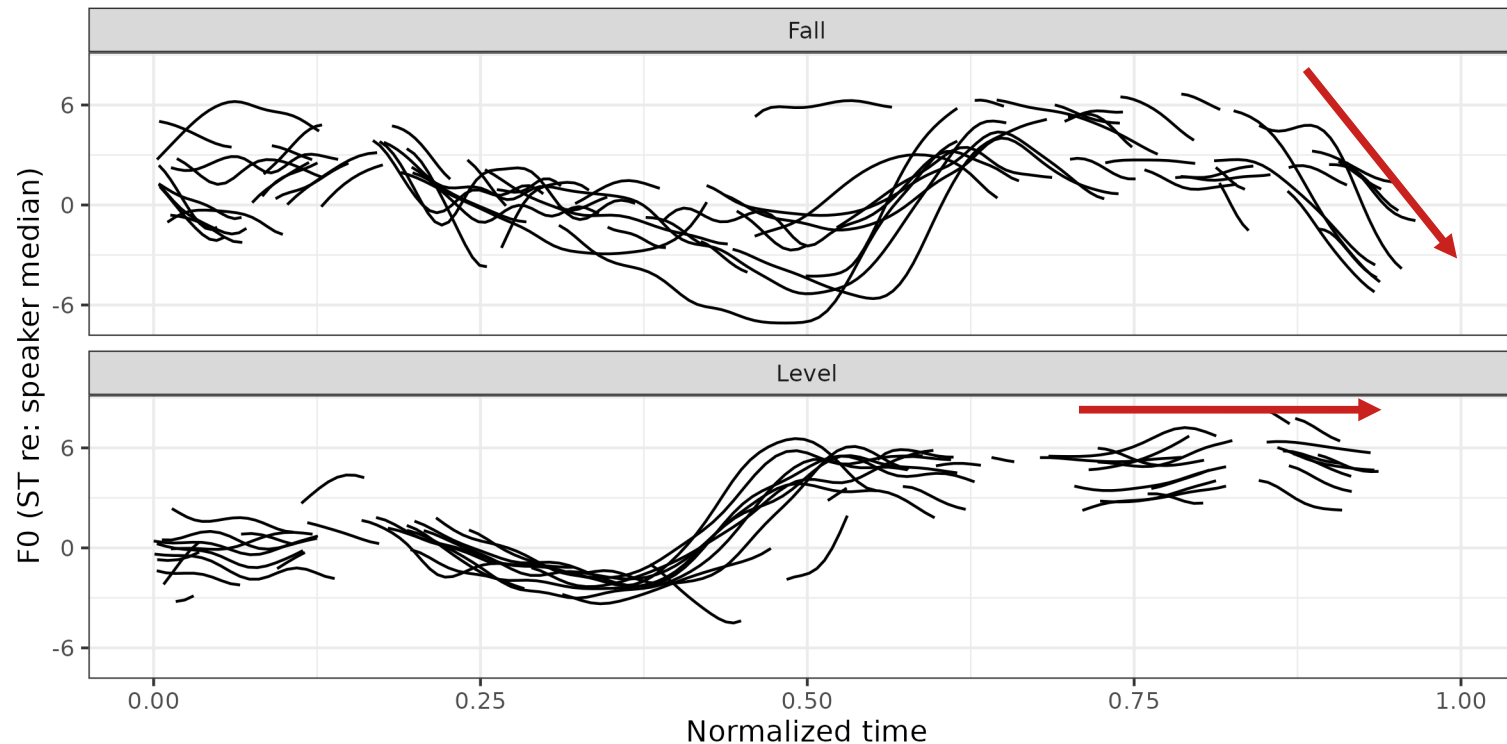
'(Do you happen to know) where she has already teste

- The app we used for cluster analysis is available at: <https://constantijnkaland.github.io/contourclustering/>
- We ran two separate cluster analyses:
 - **Whole utterance**, with 20 evenly spaced F0 points
 - **Final two syllables**, with 10 evenly spaced F0 points
- Combined analysis classifies whole contour, but gives extra weight to utterance-final pitch.

The contours

- Similar overall shape: utterance-medial jumps to plateaus
- Differences based on final two syllables: **falls** vs. **level**

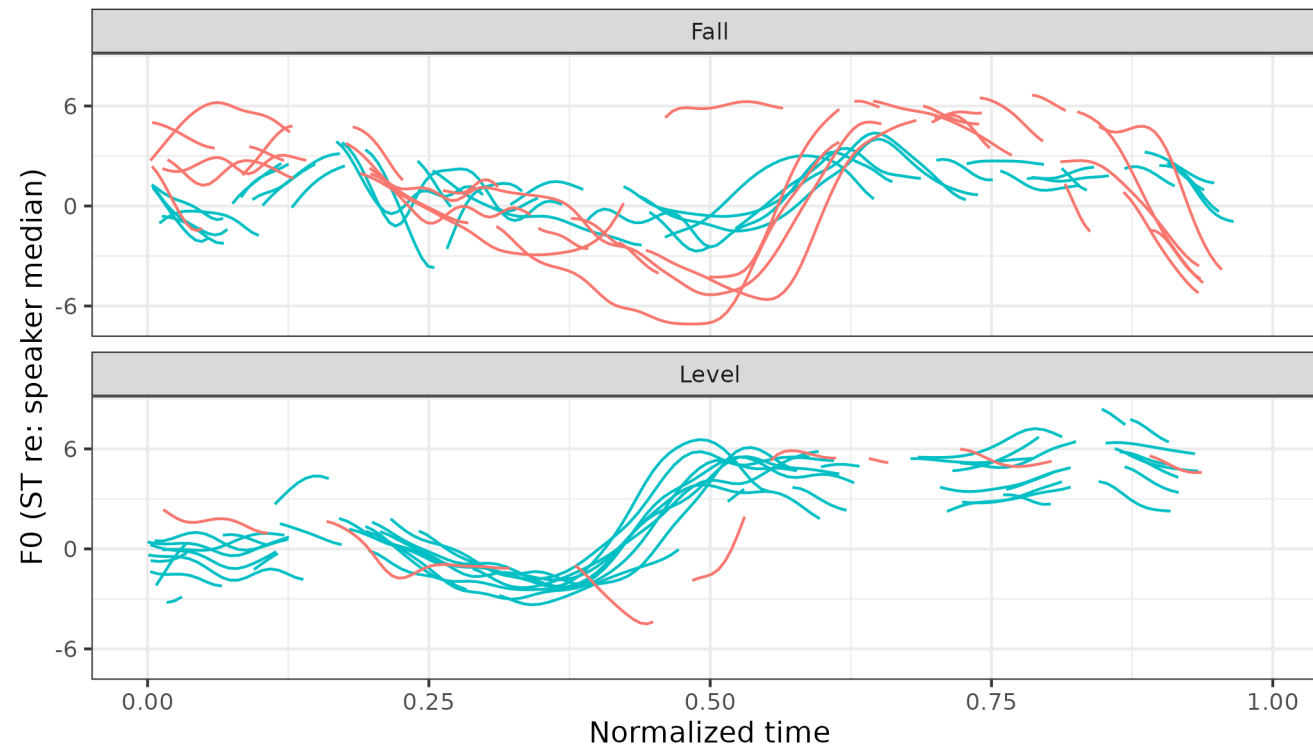
We selected the **10 most typical contours** from each cluster for the perception study (only original questions).



"Most typical": least differences to average contour within each cluster

The contours: Gender differences

- **Fall cluster:** clear late falls only from female speakers
- **Level cluster:** almost exclusively male speakers

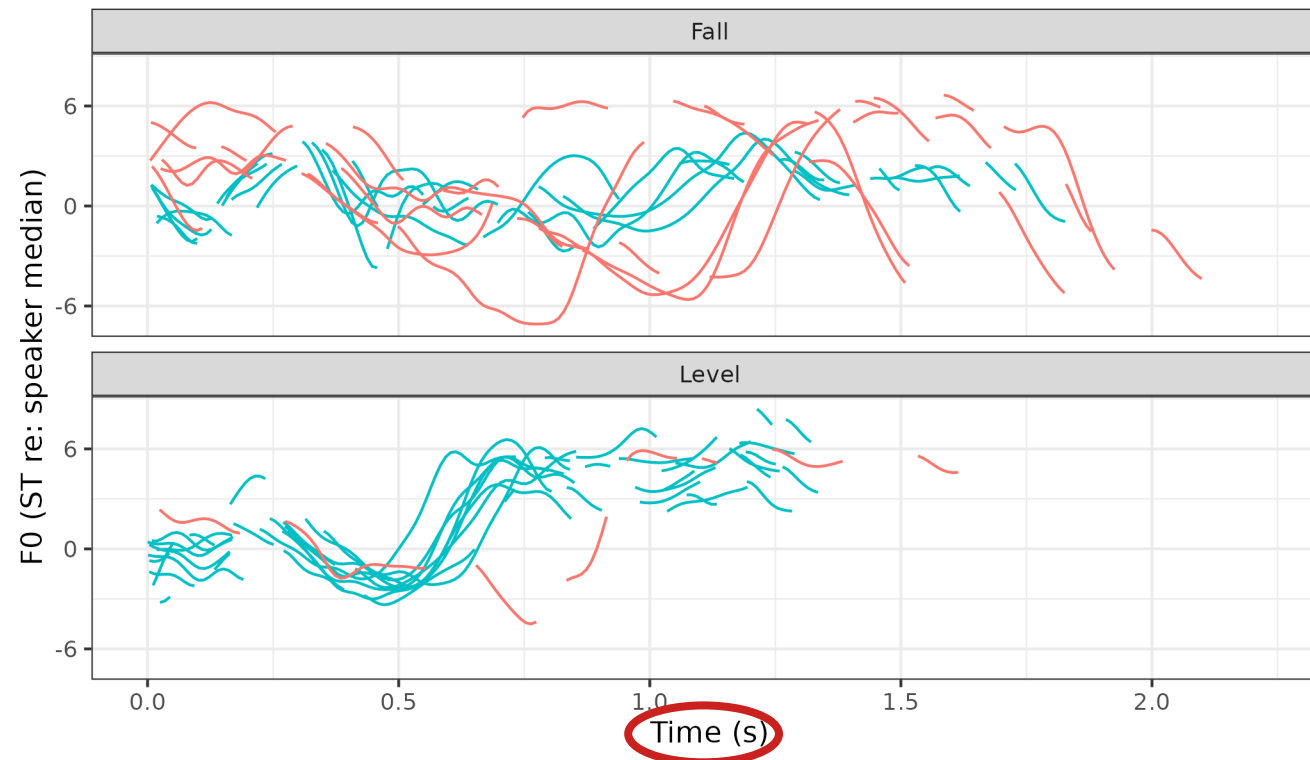


Fall cluster: selection of 5 most typical strongly falling and 5 most typical plateau-ish contours

Gender
— female
— male



The contours: Duration differences

- **Fall cluster:** much longer utterances than in Level cluster
- **Level cluster:** Alignment of rises in raw time much tighter

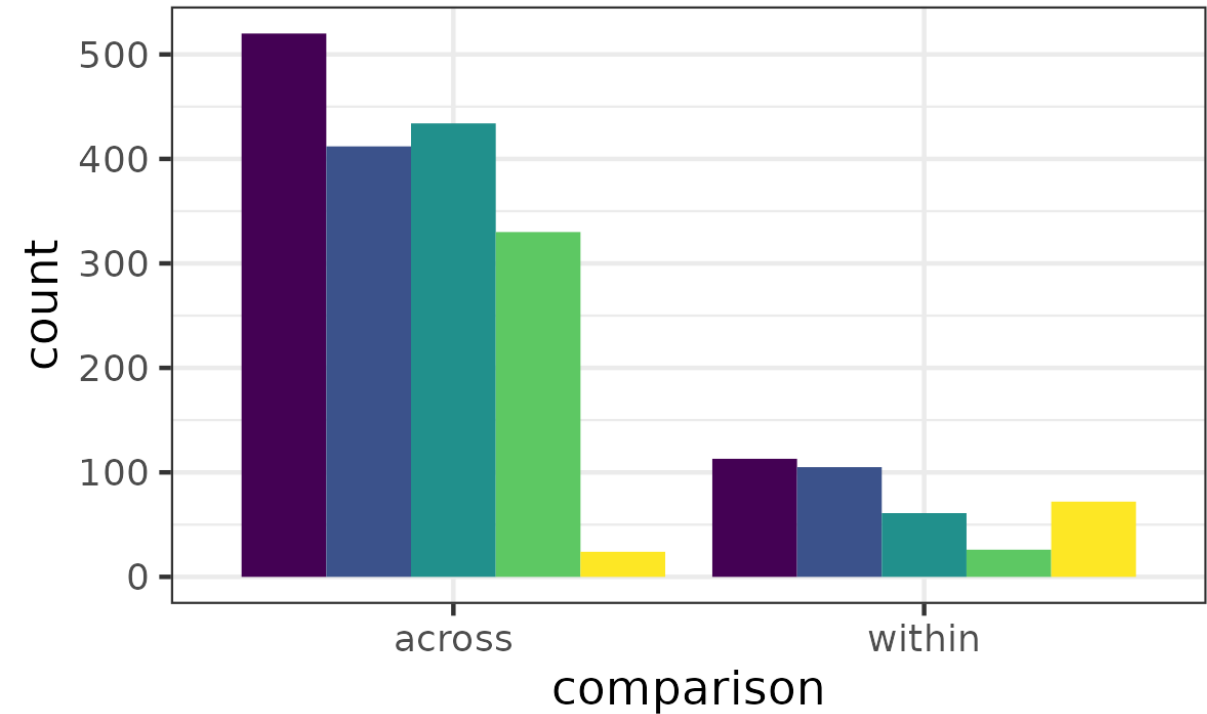
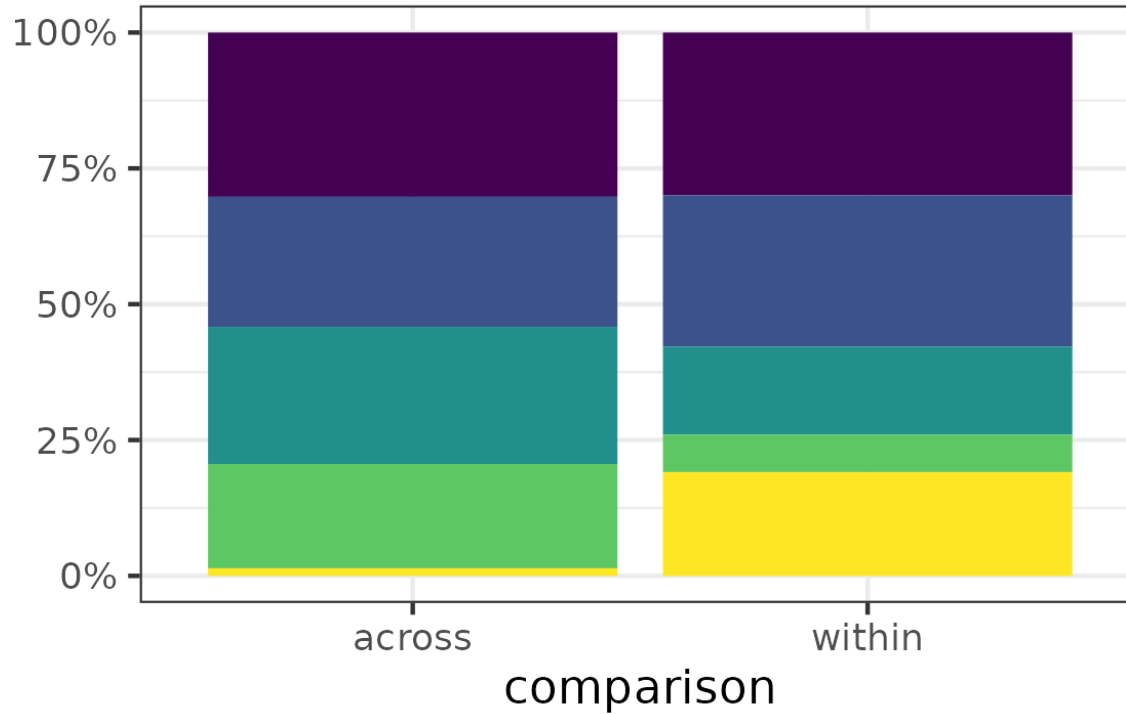


Note: The cluster analysis used normalized duration and speaker-normalized F0.

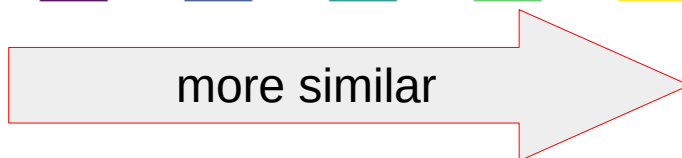
Design of the study

- Run on OpenSesame (Mathôt et al., 2012) with a custom Python script
 - We originally intended only between-cluster comparisons, but human error resulted in two within-cluster control conditions.
- Delexicalized contours – only F0 and duration available as cues
 - Praat: Sound > To Pitch > Interpolate > Smooth > To Sound (hum)  
- 21 listeners/raters, 10 x 10 comparisons, randomised order
 - 2097 data points (within Level: 189; within Fall: 188; between clusters: 1720)
 - Between-cluster condition: always in the order Level-Fall
- 5 point rating scale: 1 (unidentical) to 5 (identical)

Overall results: overview

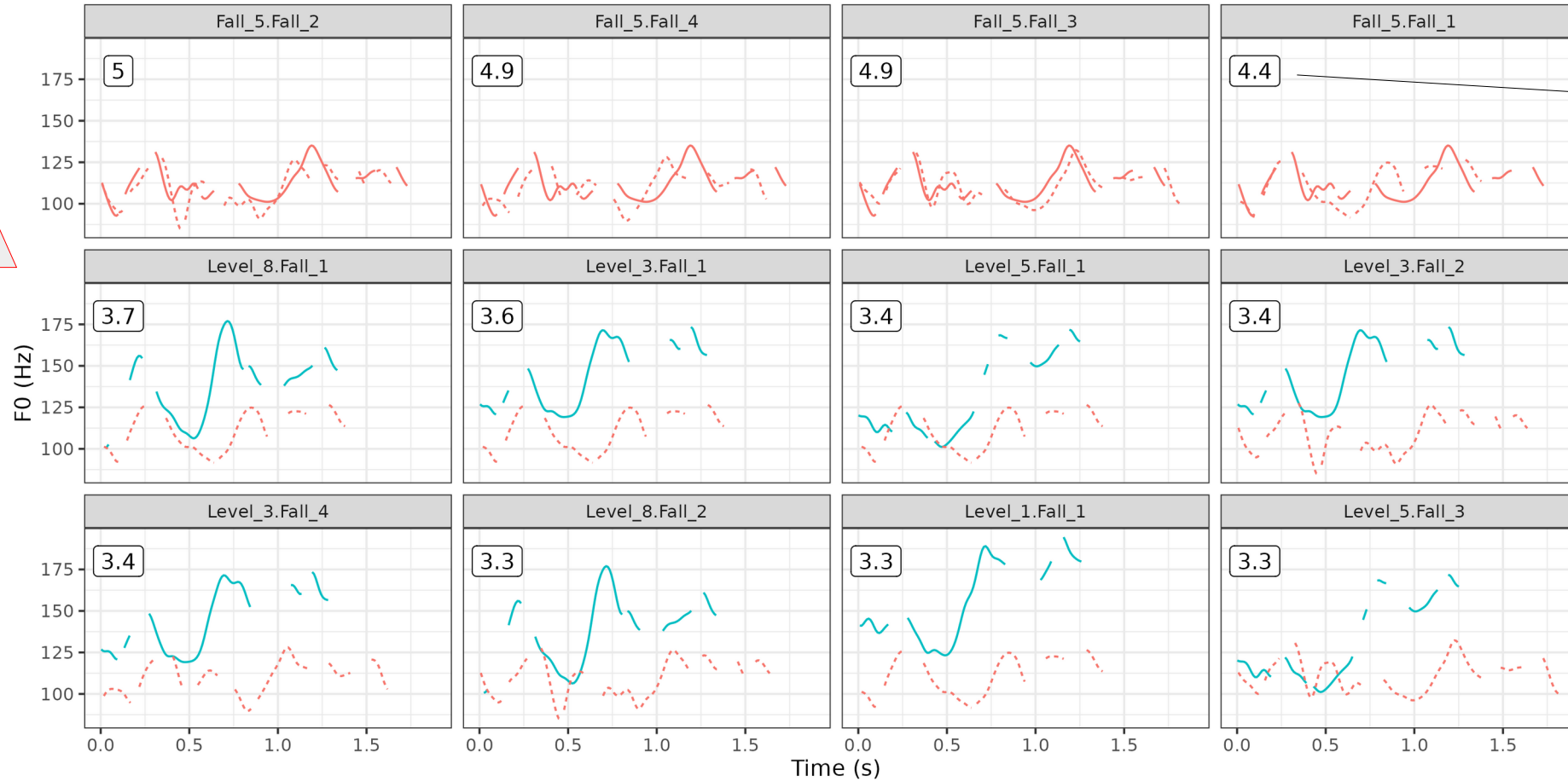


Rating 1 2 3 4 5



Overall results: 12 most similar contours (raw time and F0)

more similar

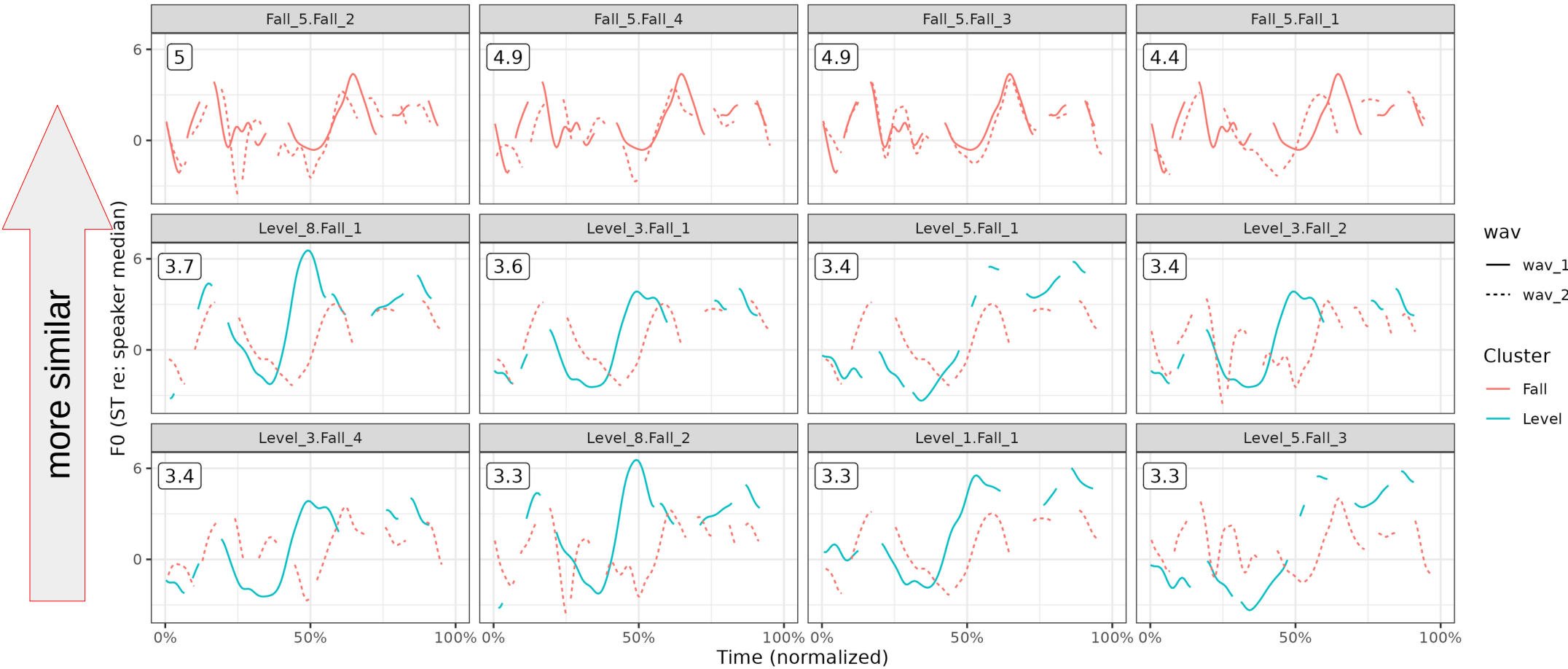


Very high ratings for very similar contours = the task is doable in principle

wav
— wav_1
.... wav_2

Cluster
— Fall
— Level

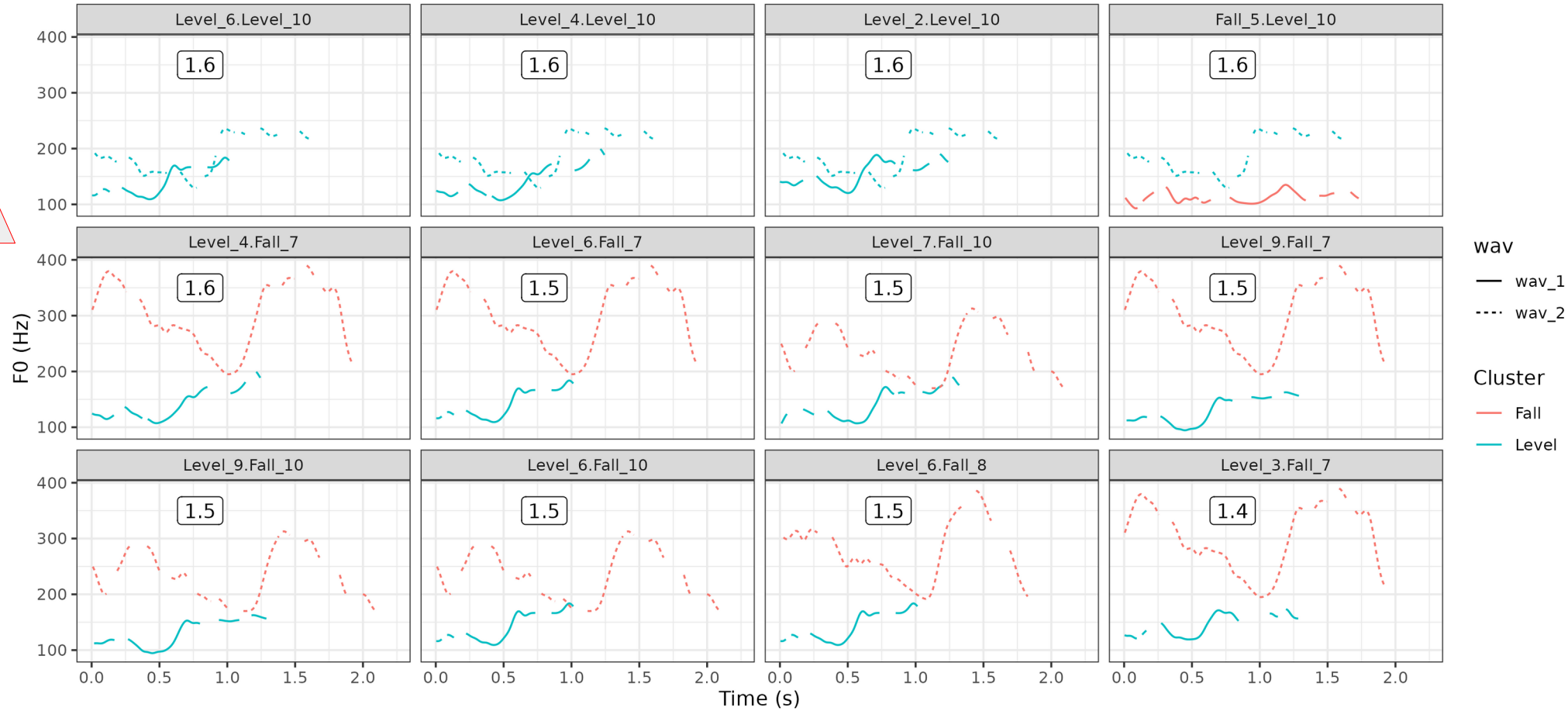
Overall results: 12 most similar contours (normalized time and F0)



Overall results: 12 least similar contours

(raw time and F0)

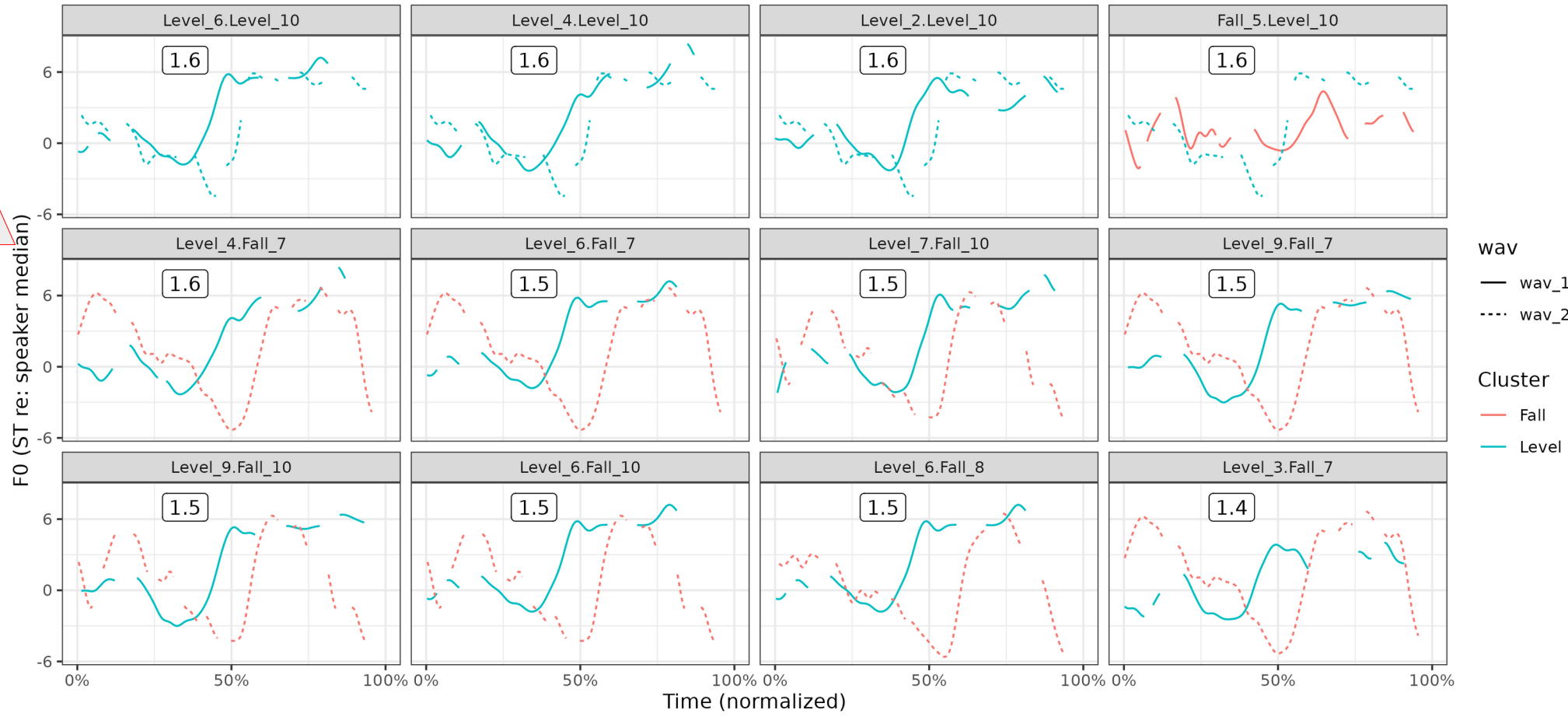
more similar



Overall results: 12 least similar contours

(normalized time and F0)

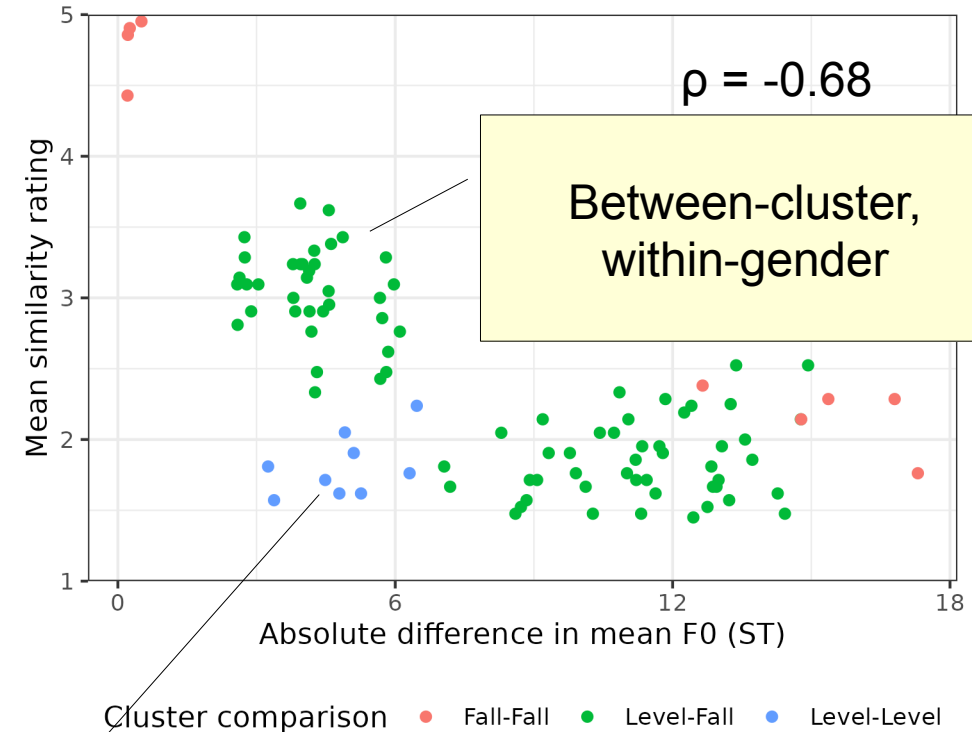
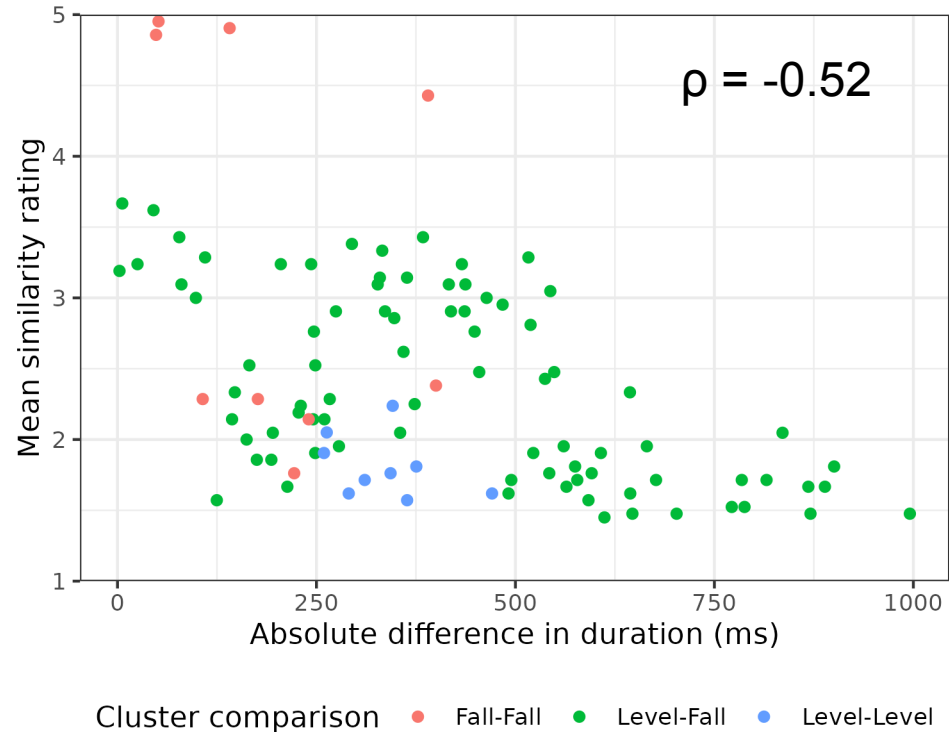
more similar



We present the results of linear mixed models for ease of interpretation. The results are the same with cumulative link mixed models.

- We wanted to test post-hoc if there are *only* gender effects.
- No: Gender and cluster membership are both strong predictors of similarity ratings (whether their interaction is included or not).
- We also explored correlations between continuous acoustic measures and similarity ratings.

Correlations: acoustics – ratings



Within-cluster, between-gender comparisons: low ratings despite fairly similar average F0

- Controlling for gender, difference between across and within cluster comparisons appears to be robust
→ Clusters are perceptually distinct.
- But: substantial influence from duration and F0 register (and speaker gender on its own) on perceptual similarity.
- Putative H-L% sticks out perceptually, but along several dimensions (pitch movement, pitch register *and* duration).

Acknowledgments

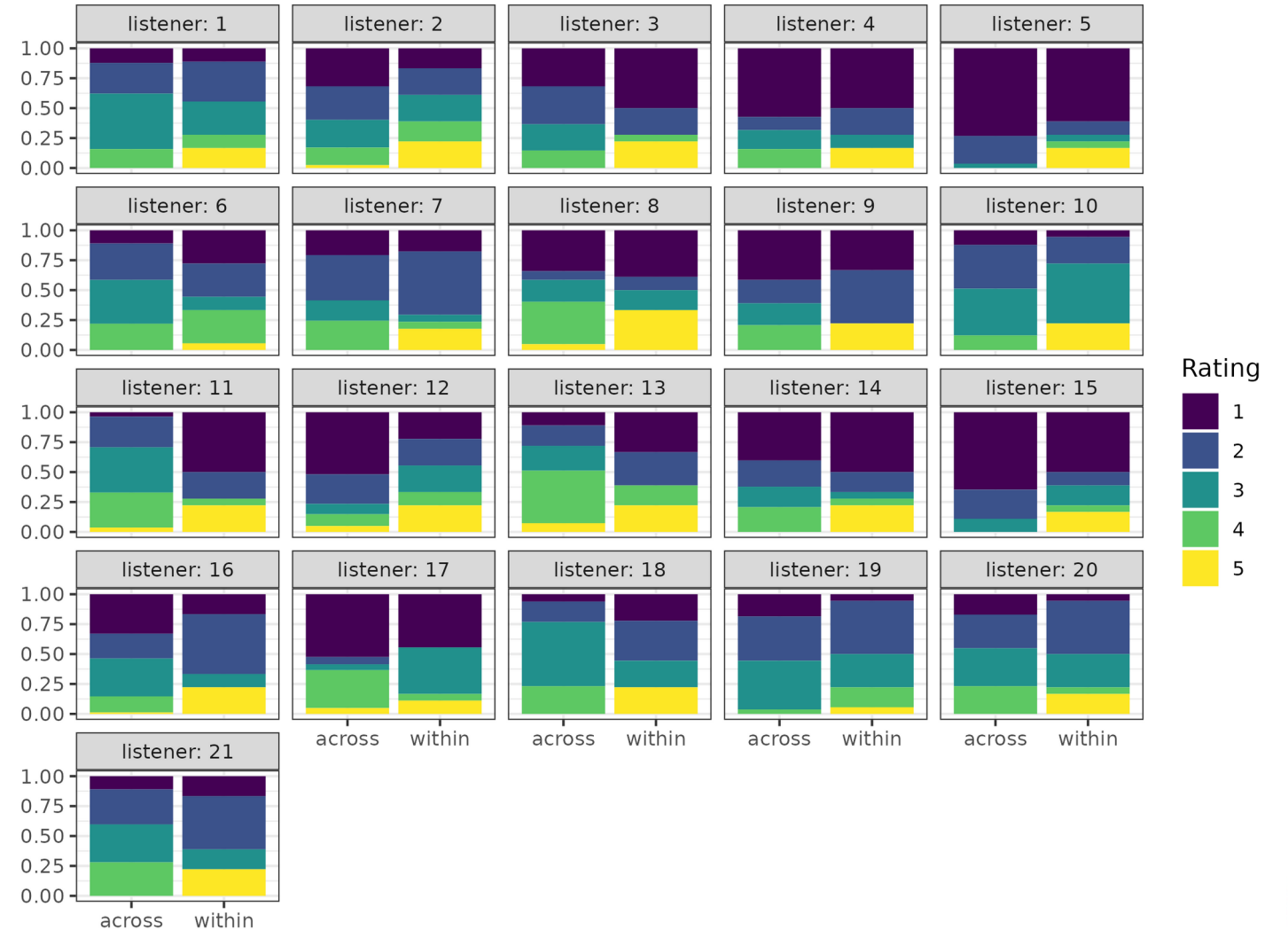
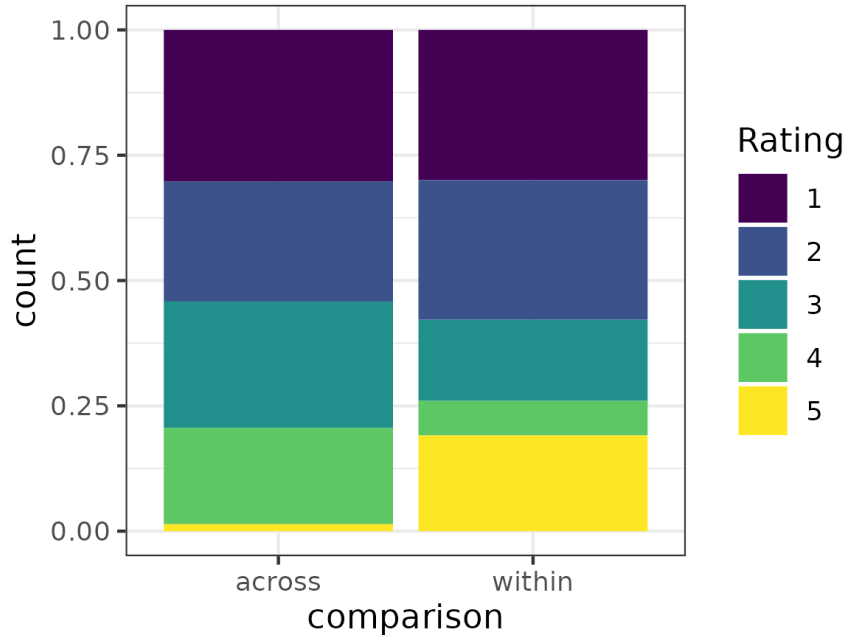
We wish to thank Aleksandra Kopotilova for her contributions to the perception study.

The research for this paper has been funded by the German Research Foundation (DFG) – Project-ID 281511265 – as part of the SFB 1252 “Prominence in Language” in the projects A03 and A06 at the University of Cologne.

References

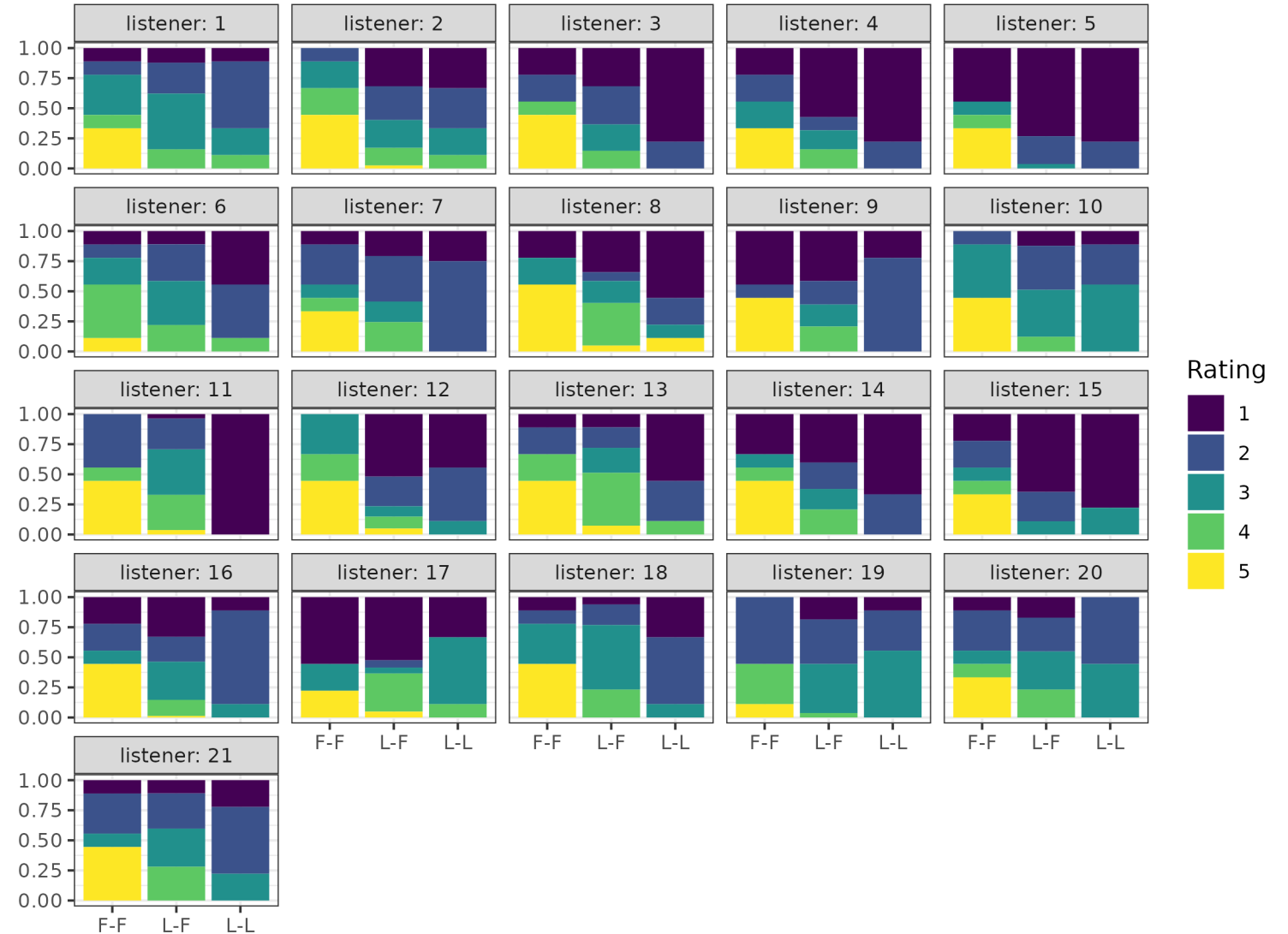
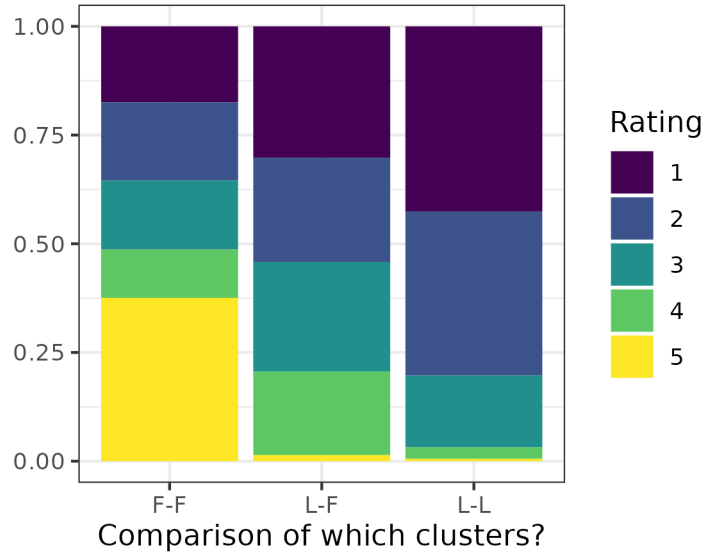
- Kaland, C. (2023). Contour clustering: A field-data-driven approach for documenting and analysing prototypical f0 contours. *Journal of the International Phonetic Association* 53(1), 159–188.
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44(2), 314–324.
- Repp, S., & Seeliger, H. (subm.). Contrast + givenness, local + non-local. The influence of complex information-structural settings on the prenuclear, nuclear and post-nuclear regions in exclamatives and questions. Submitted to *Language and Speech*.
- Seeliger, H., & Kaland, C. (2022). Boundary tones in German wh-questions and wh-exclamatives – A cluster based approach. *Proceedings of Speech Prosody 2022*, 27–31.

Appendix: listener-specific variation?



Within or across cluster comparison?

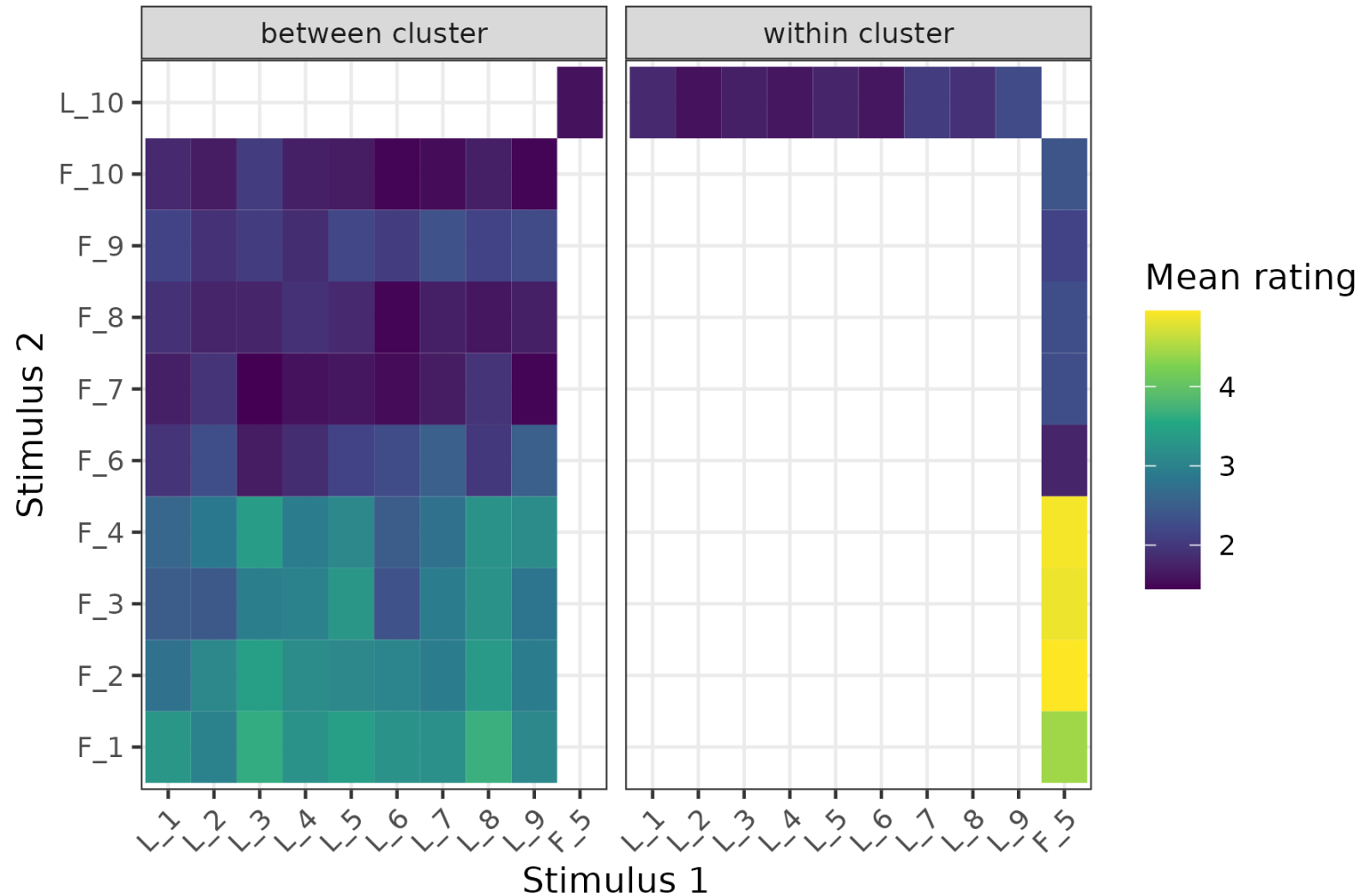
Appendix



Comparison of which clusters?



Overall results: cluster comparison



Overall results: gender comparison

