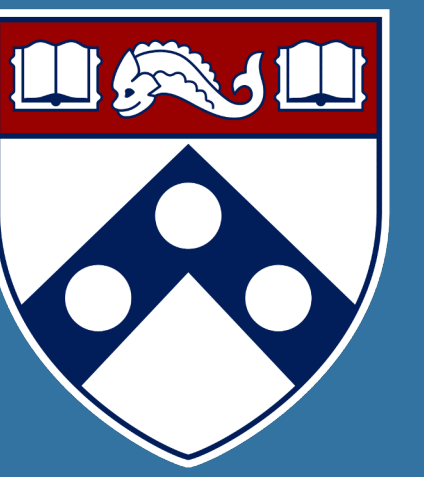


Information-theoretic hypothesis generation of relative cue weighting for the voicing contrast

Annika Heuser & Jianjing Kuang
University of Pennsylvania



Problem

- How effective cues are for differentiating a phonological contrast depends on **context**
 - How do kids learn to attend to the cues that are most effective for the context?
- Need to conduct perceptual experiments to figure this out

Method

- Feature importances from decision trees as a proxy for perceptual cue weights may help inform experimental designs
- Use Standard American English (SAE) voicing as a case study to verify this
- I.e. checked that feature importances from TIMIT [1] data reflected well-known differences in voicing realization between onsets/codas, and stops/fricatives (e.g. [2-4])

Acoustic features

cl_dur = closure duration

partial_voice = proportion of voicing during oral closure

trans_f0/f1/etc = transition f0/formants; average of first/last 5% of adjacent vowel

adj_vow_f0/f1/etc = average across vowel

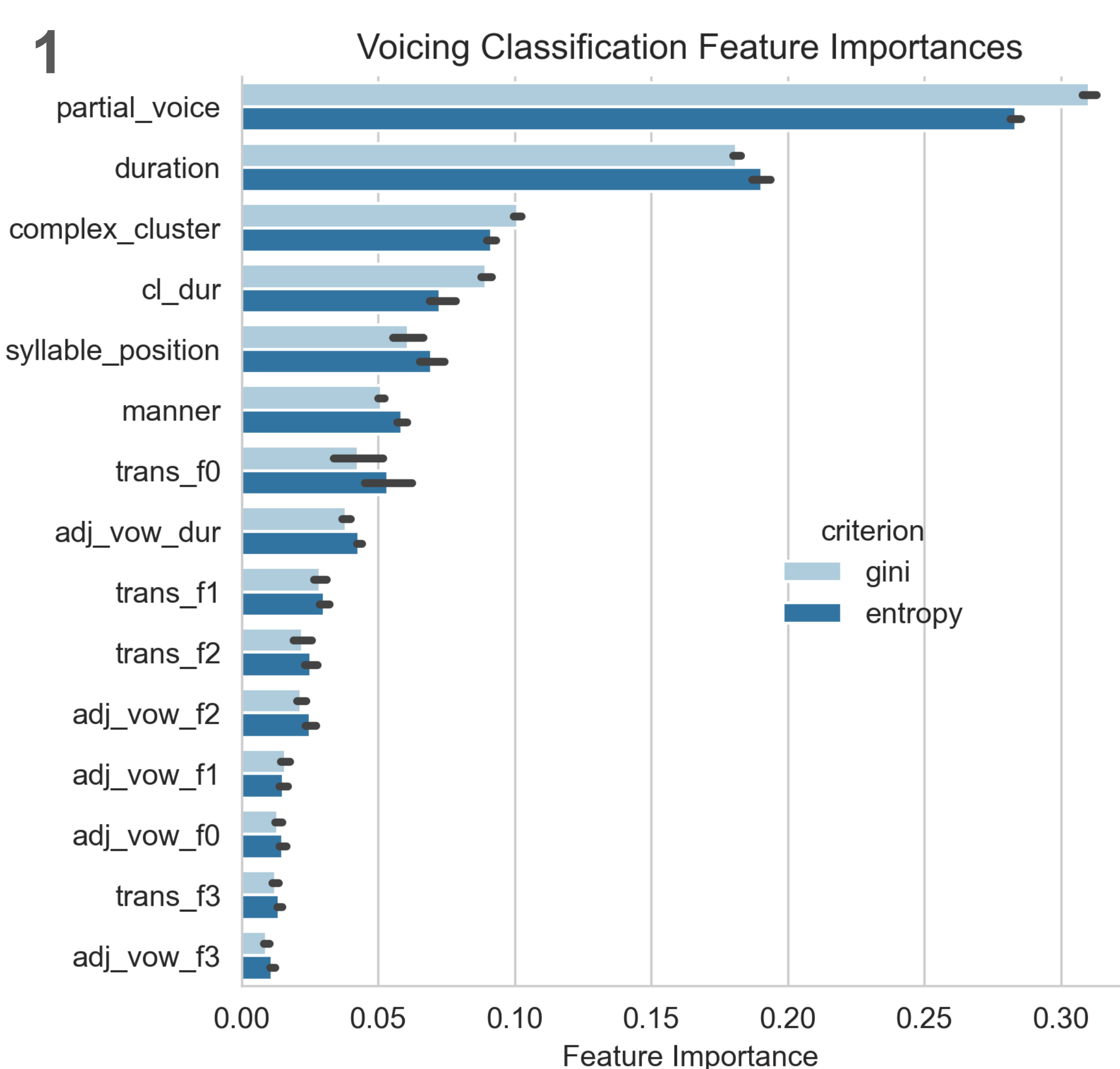
manner = stop or fricative

Results

- Accuracy of decision trees from which we calculated feature importances consistently ~85%

Figure 1 (above): From classifying all voiced/voiceless consonants together

- Expected that partial voicing and duration would be highly ranked



Contextual Variation

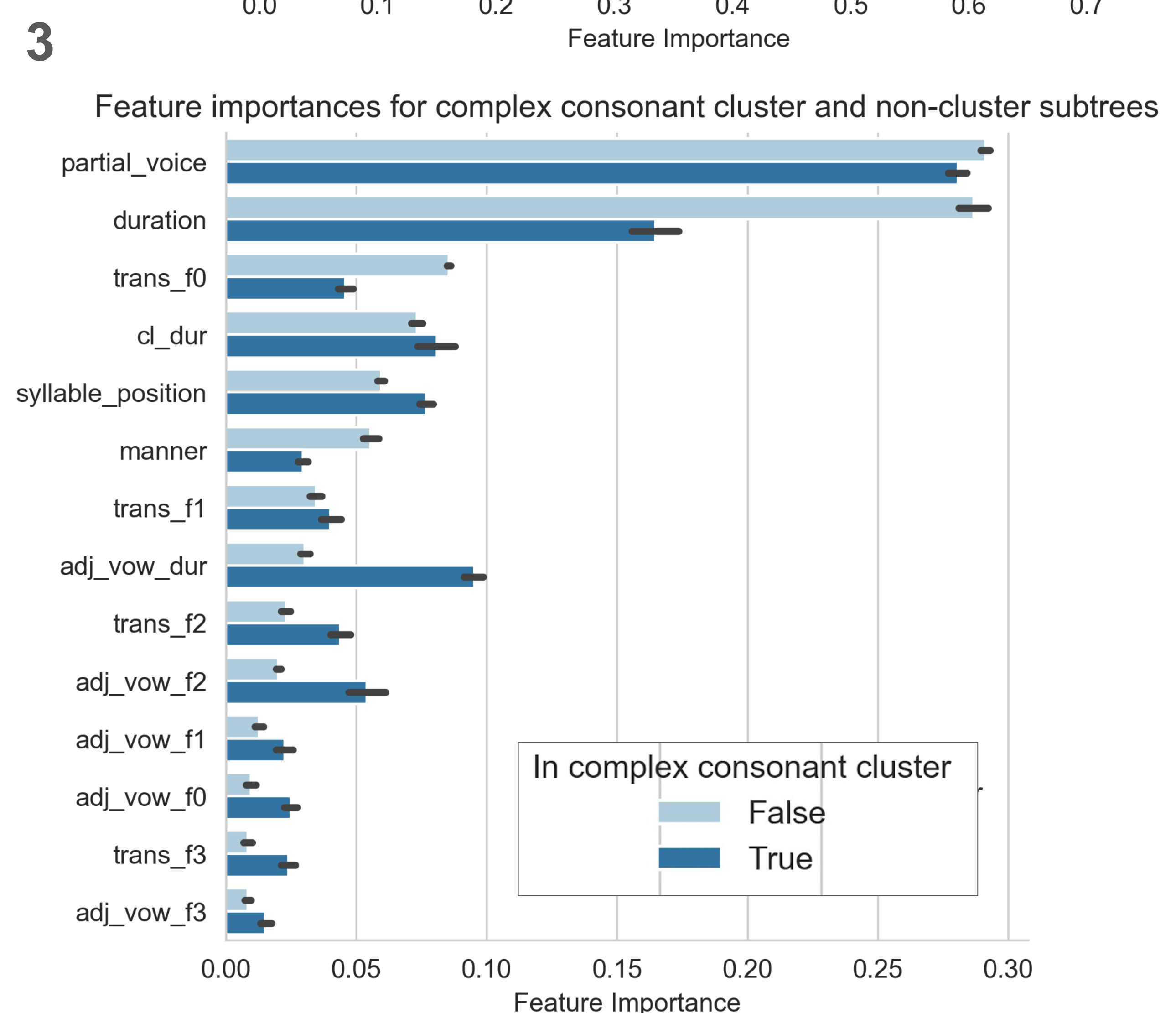
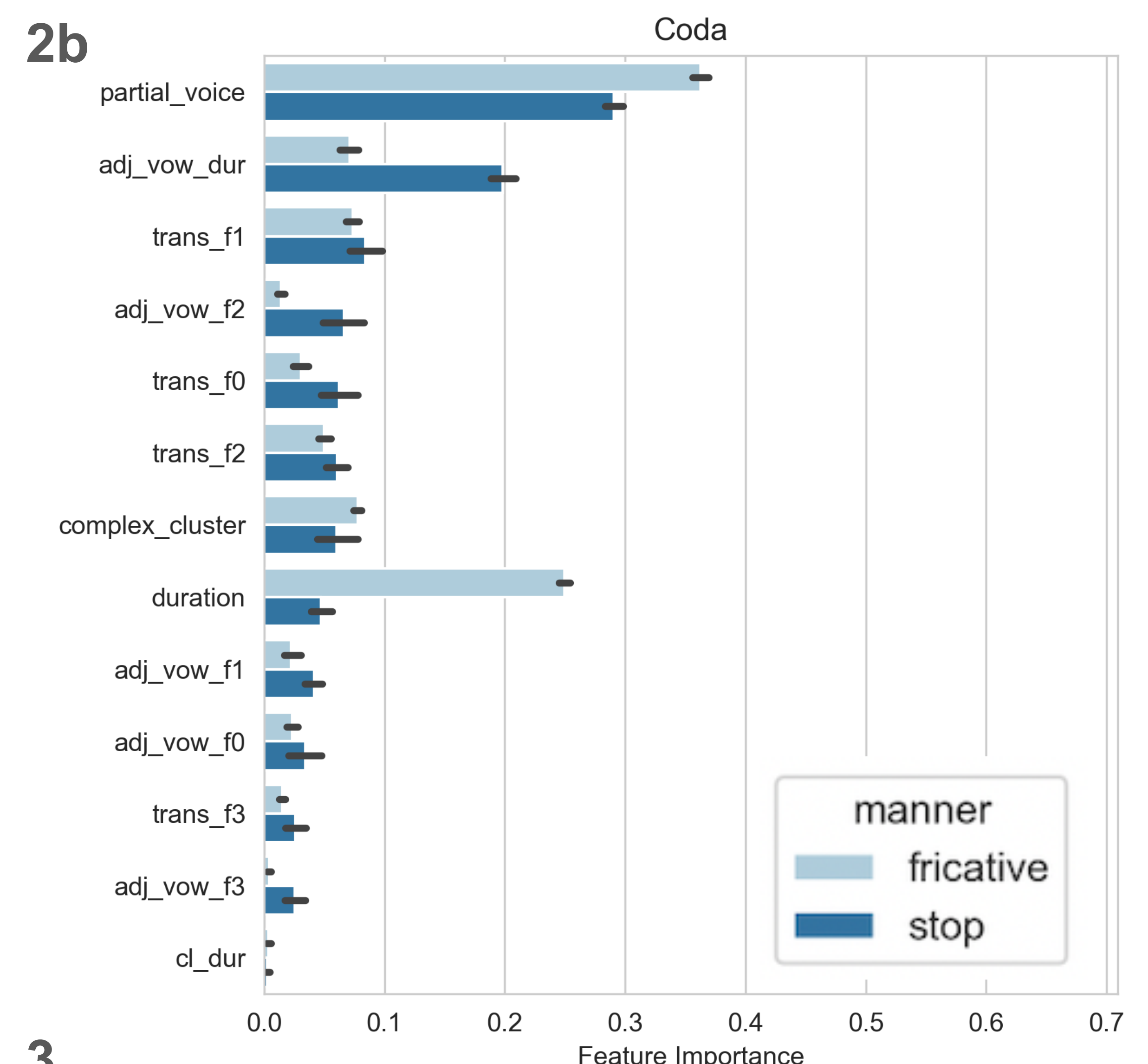
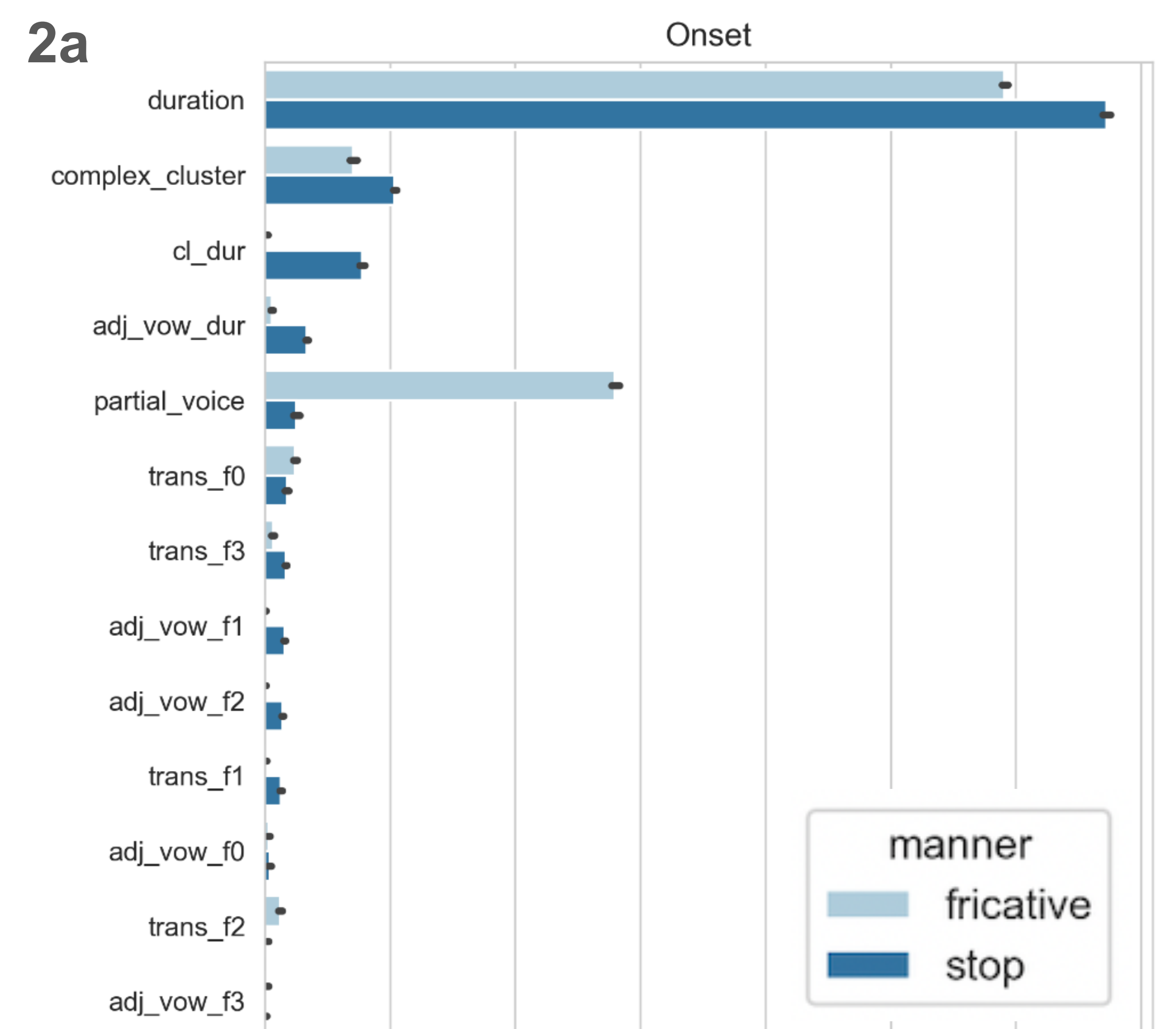
Figure 2 (top right): Feature importances from only classifying **fricative** or **stops** in word onsets (2a) or codas (2b), respectively

- Many expected findings, e.g. VOT (stop durations in onsets) is most important

Figure 3 (right): Classifying consonants in **complex clusters** (e.g. “s” or “t” in “star”) separately from **lone consonants**

- Less studied so this provides new insight!

Results – contextual variation



Contributions and Conclusions

- Validated that decision tree feature importances make accurate cue weighting predictions for SAE voicing
- Decision trees = low resource and high explainability
- Hypotheses from predicted relative cue weightings can inform future experiments aiming to understand the perception of phonological contrasts in different contexts

References

- [1] J. S. Garofolo, "Timit acoustic phonetic continuous speech corpus," Linguistic Data Consortium, 1993, 1993 [2] L. Davidson, "Variability in the implementation of voicing in American English obstruents," Journal of Phonetics, vol. 54, pp. 35–50, 2016. [3] L. Lisker, "Closure duration and the intervocalic voiced-voiceless distinction in English," Language, vol. 33, no. 1, pp. 42–49, 1957. [4] K. N. Stevens and D. H. Klatt, "Role of formant transitions in the voiced-voiceless distinction for stops," The Journal of the Acoustical Society of America, vol. 55, no. 3, pp. 653–659, 1974.