

Hierarchical processing of naturalistic music during production and perception

Riesa Cassano¹, Jamal Williams², Marius Cătălin Iordan², Uri Hasson², Elise A. Piazza¹

¹Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627

²Princeton Neuroscience Institute and Department of Psychology, Princeton, NJ 08544

Contact: riesa_cassano@urmc.rochester.edu

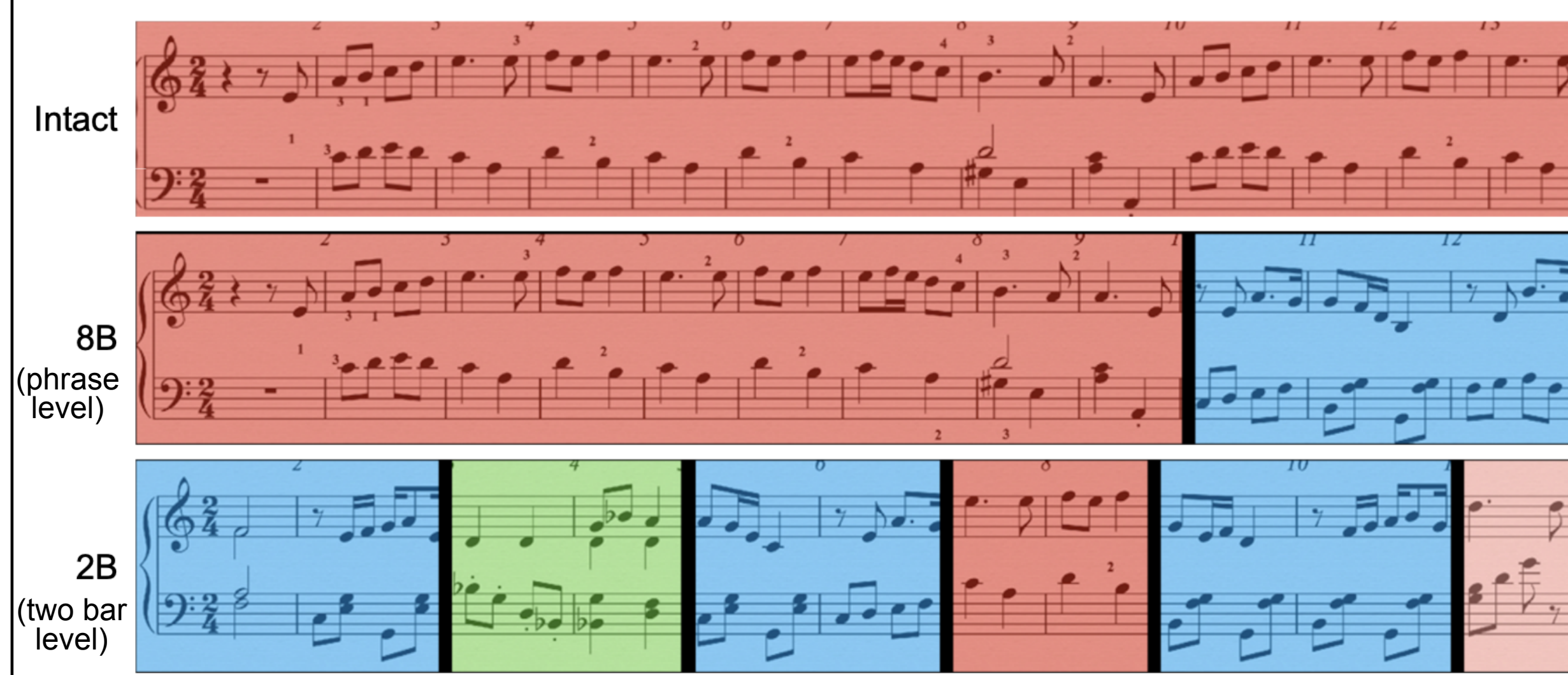


Background

- Western Classical music is typically organized into nested hierarchical structure, spanning notes, phrases, and sections.
- Listeners have shown sensitivity to structural features at the local (compared to the global) level (Granot and Jacoby, 2011).
- Study 1: We use scrambled natural music to investigate processing of musical structure at multiple timescales. We use an event segmentation task to probe perception of distinct events in intact music, as well as music scrambled in phrases and smaller units.
- Different parts of the brain process music at different timescales during perception (Farbood et al., 2015).
- Study 2: In an ongoing fMRI study, we extend the findings of Farbood et al. (2015) to music production and directly compare production and perception in the same subjects.

Methods

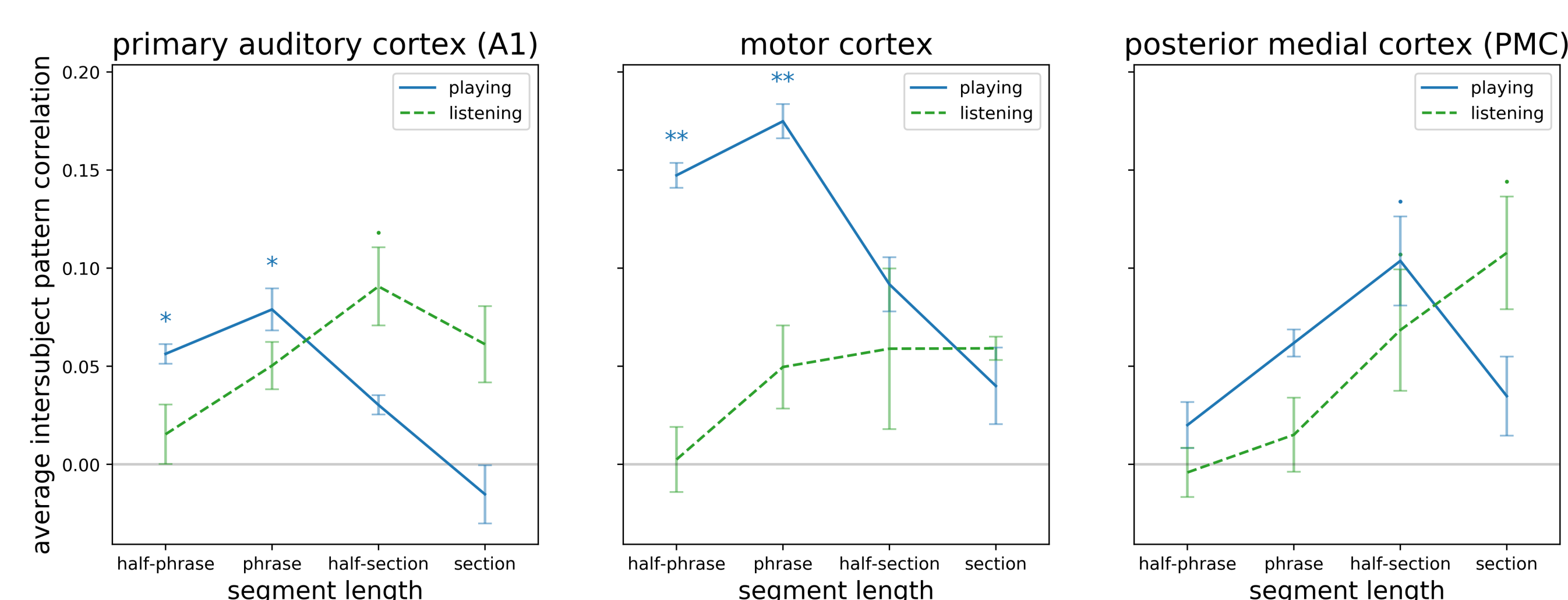
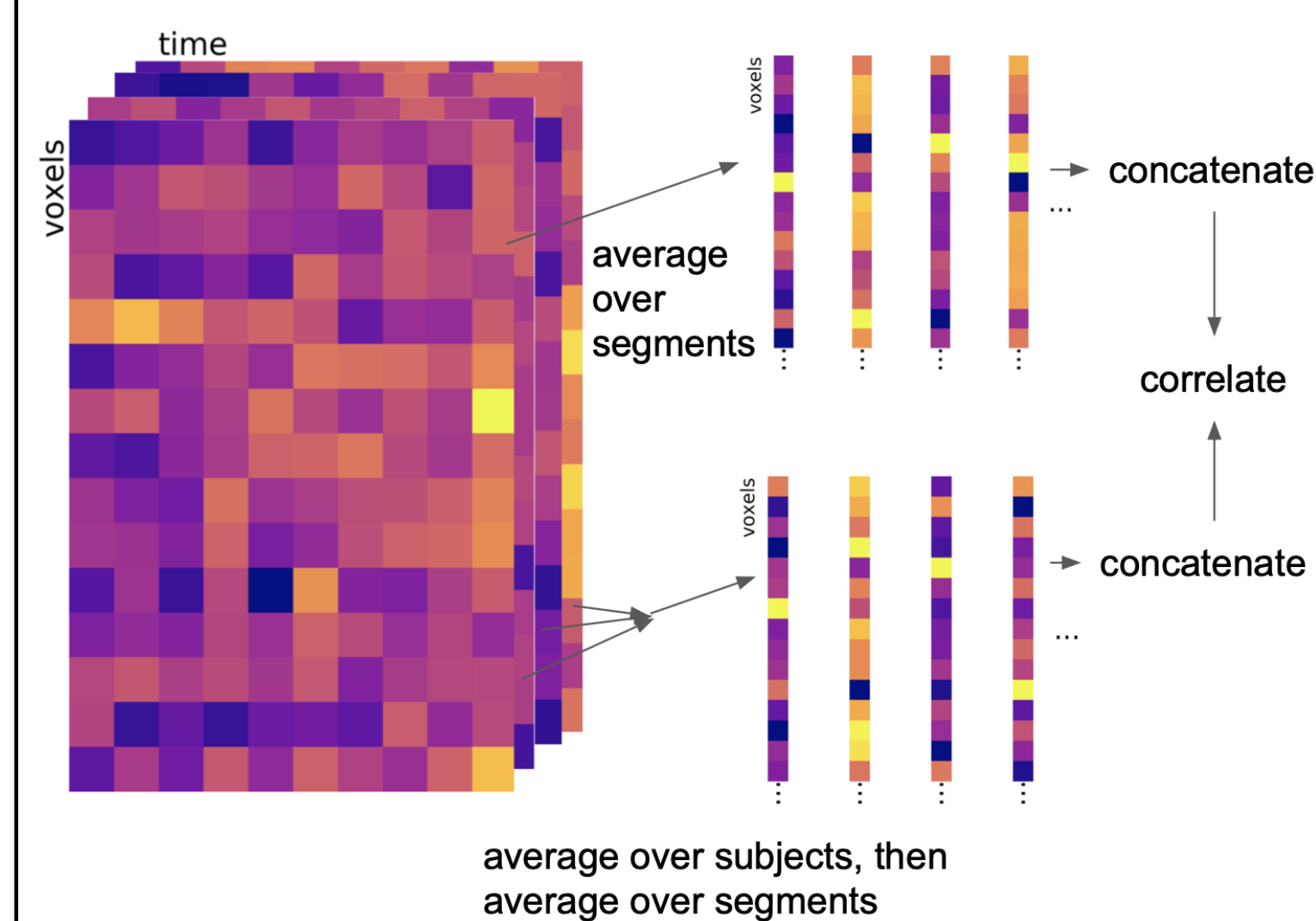
- In an online behavioral study, subjects listened to versions of a medley scrambled at different timescales. They identified when they heard a "meaningful change" in the music by pressing the space bar.
- In a related ongoing fMRI study, expert pianists played a non-ferromagnetic three-octave keyboard in the MRI scanner.



Colors correspond to which section of the original medley the bar comes from.

- Phrases generally contain 8 bars each. Phrases in music are the basic unit of meaning, analogous to sentences in narrative.
- Each section contains 4 phrases. "Half-sections" contain 2 phrases.

Preliminary fMRI results



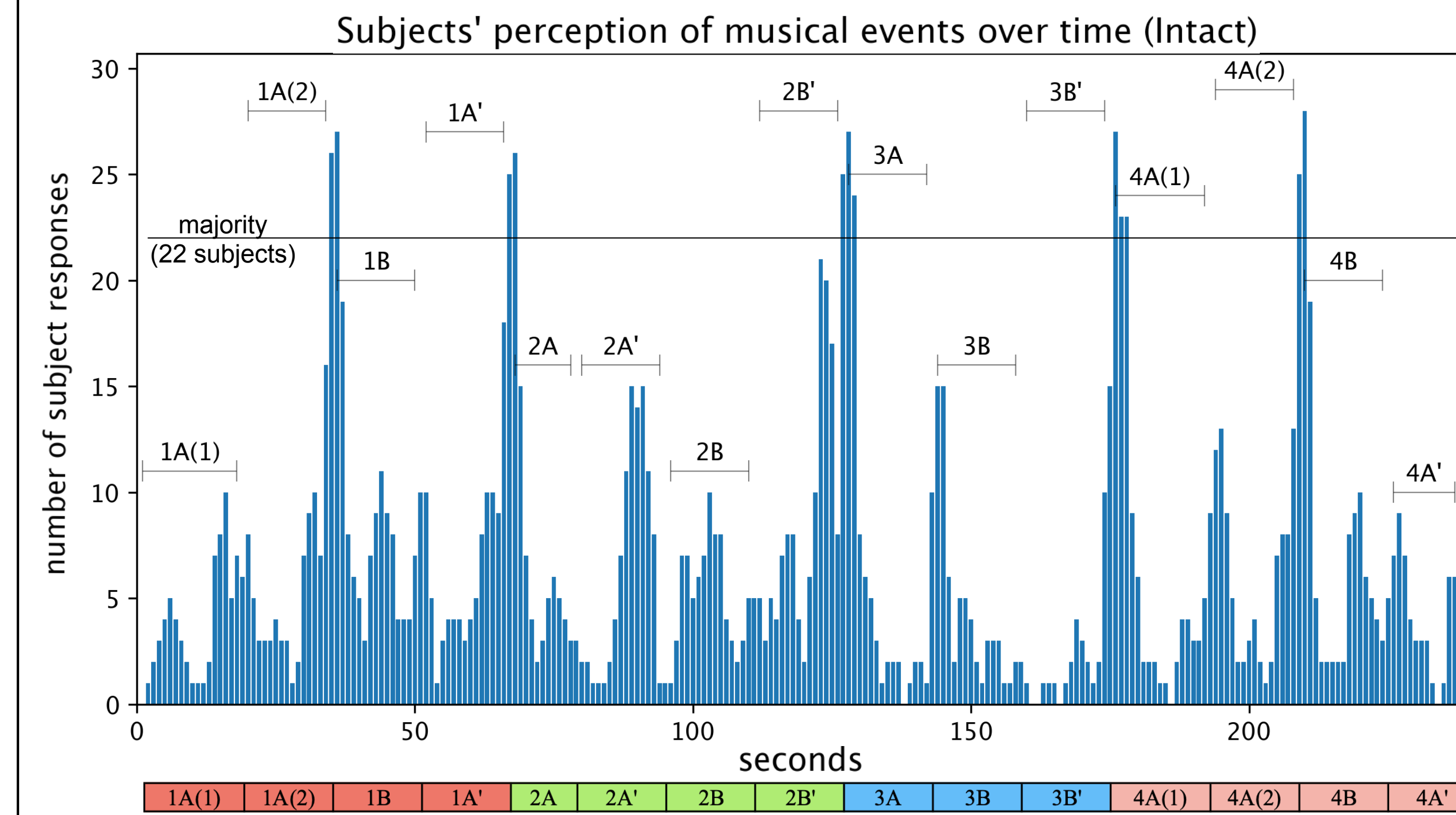
- Intersubject correlation (ISC) measures response reliability across subjects.

- ISC above chance in a brain region suggests that the stimulus drives a response that is shared across subjects in that region.

During playing, A1 and motor cortex show reliable patterns of activity over shorter segments, suggesting that subjects are chunking what they are playing into shorter segments (like phrases).

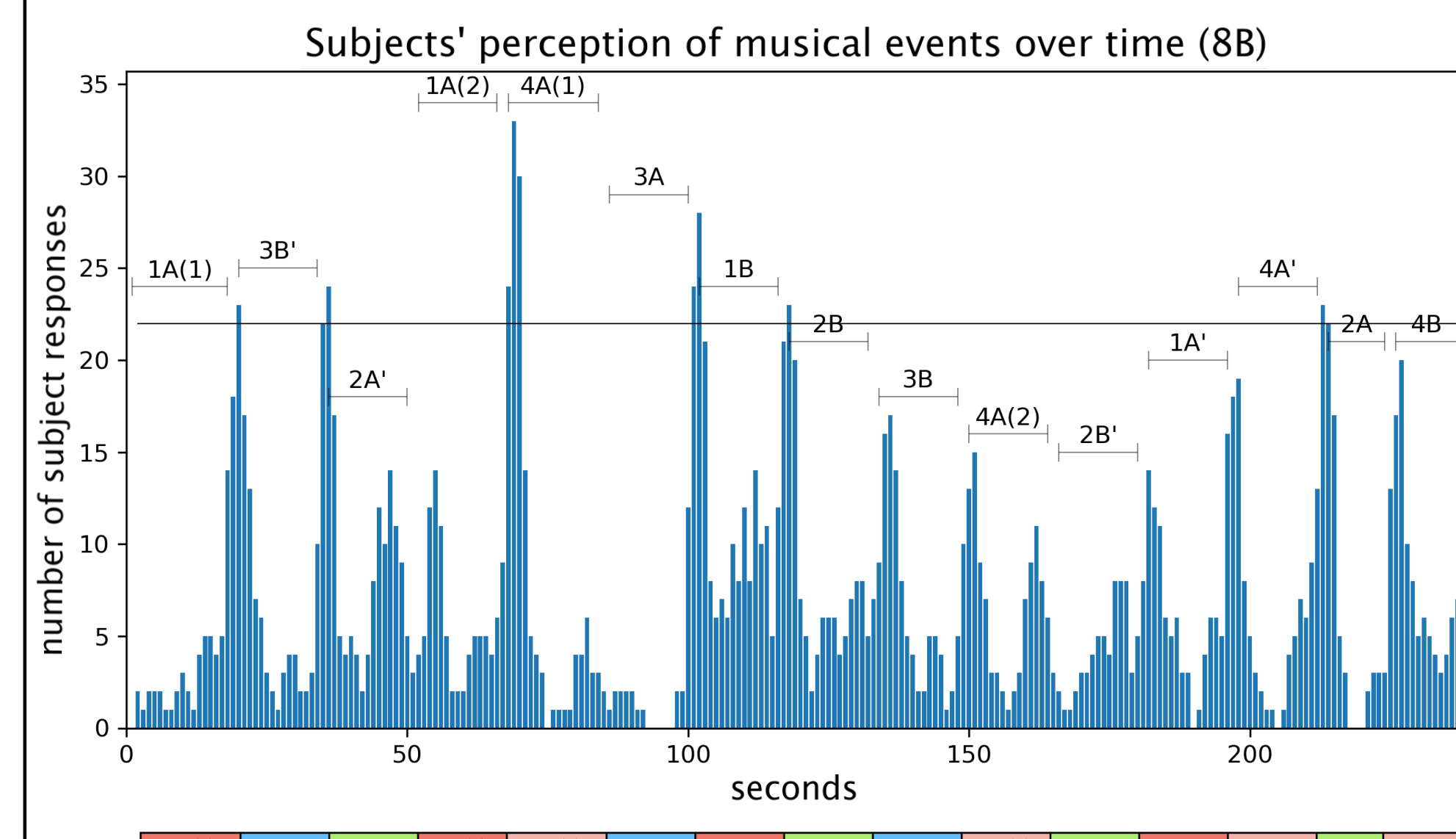
During listening, A1 and posterior medial cortex (a higher-order association area) show more reliable response patterns over longer segments, suggesting that subjects might devote more attention to longer-term chunking while listening.

Behavioral results

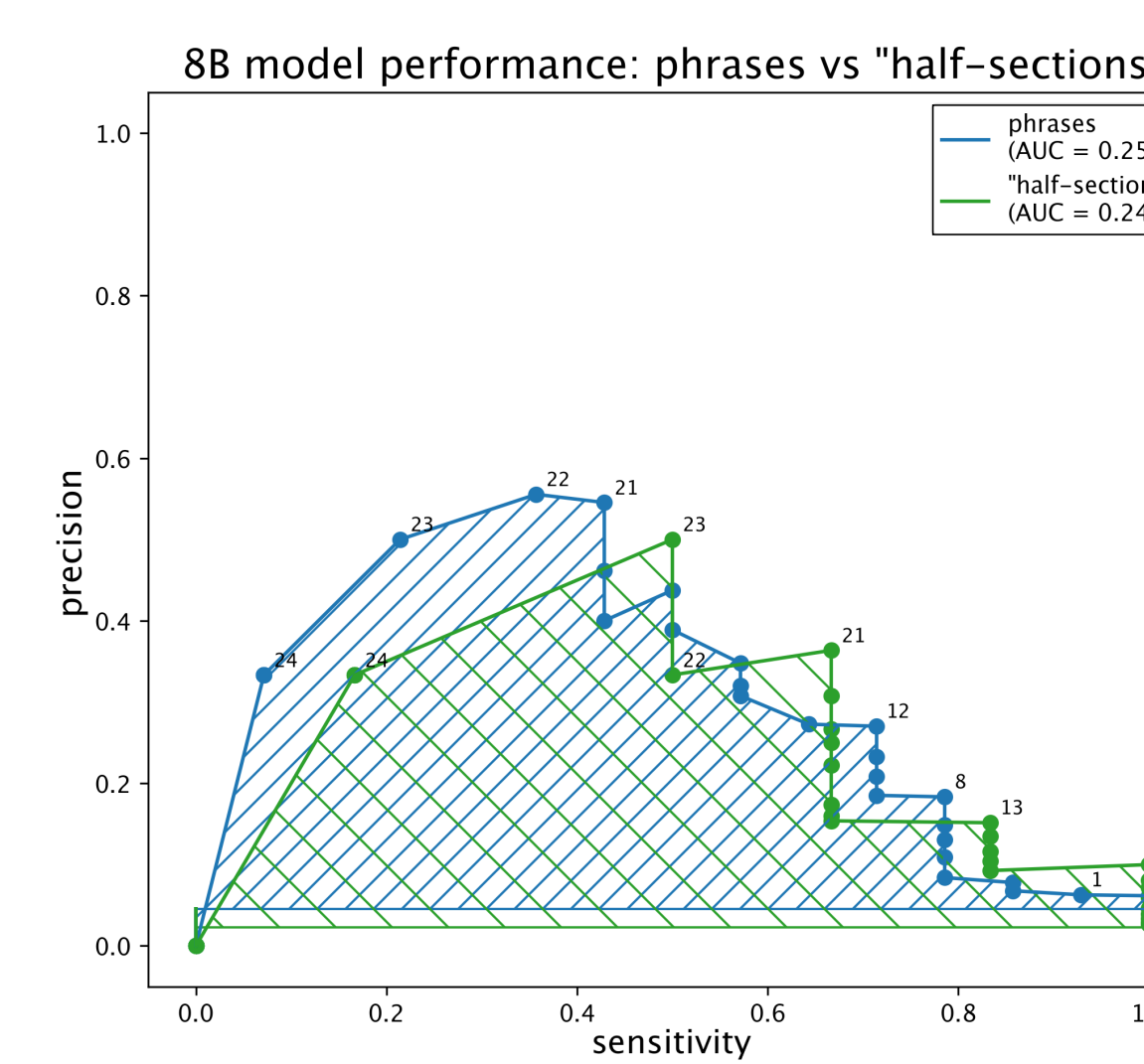


For natural, Intact music, we expected subjects to respond at phrase boundaries, since phrases are the basic unit of meaning in music.

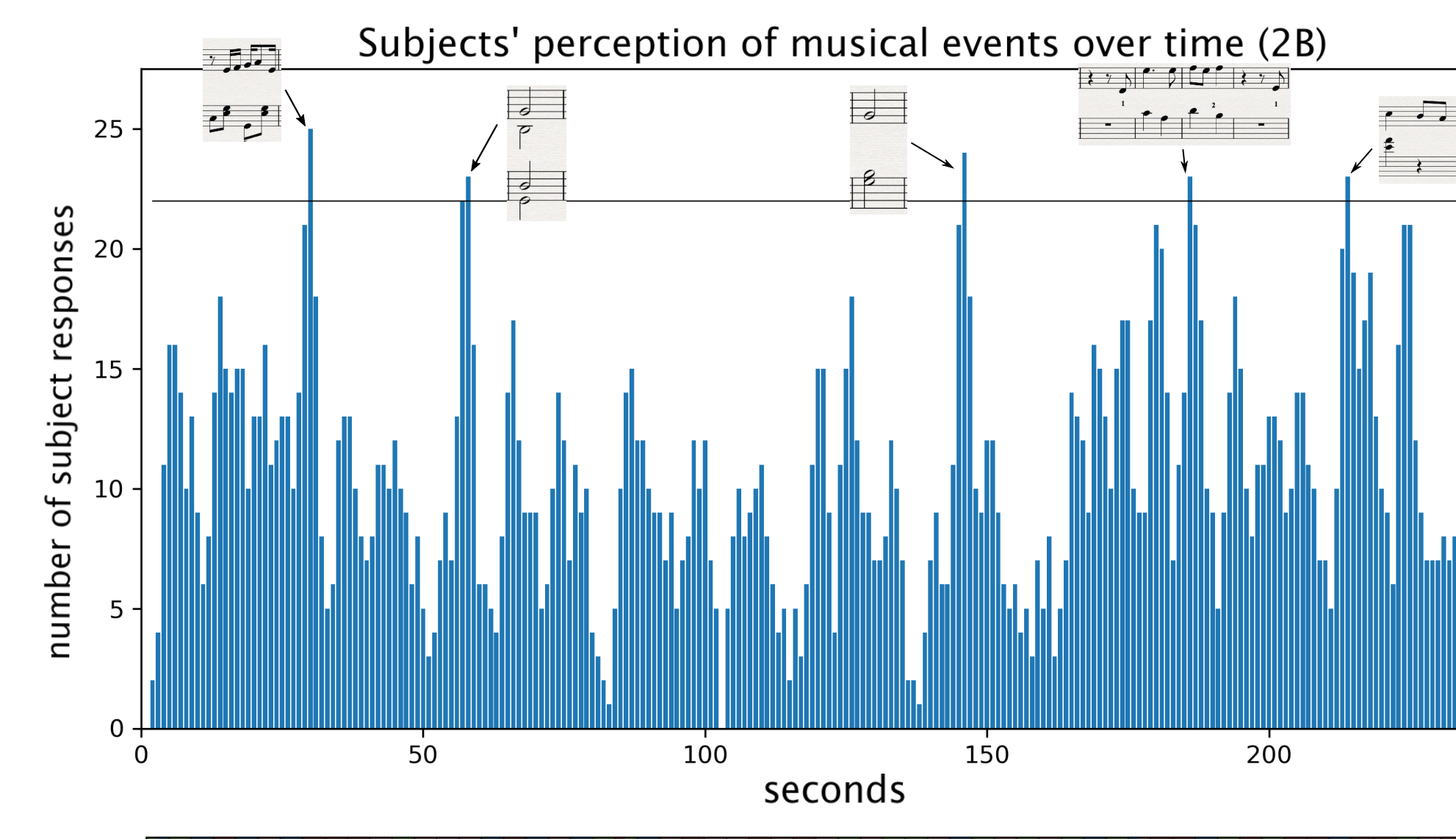
However, a majority of subjects respond to section boundaries and half-section (in sections 1 and 4) boundaries.



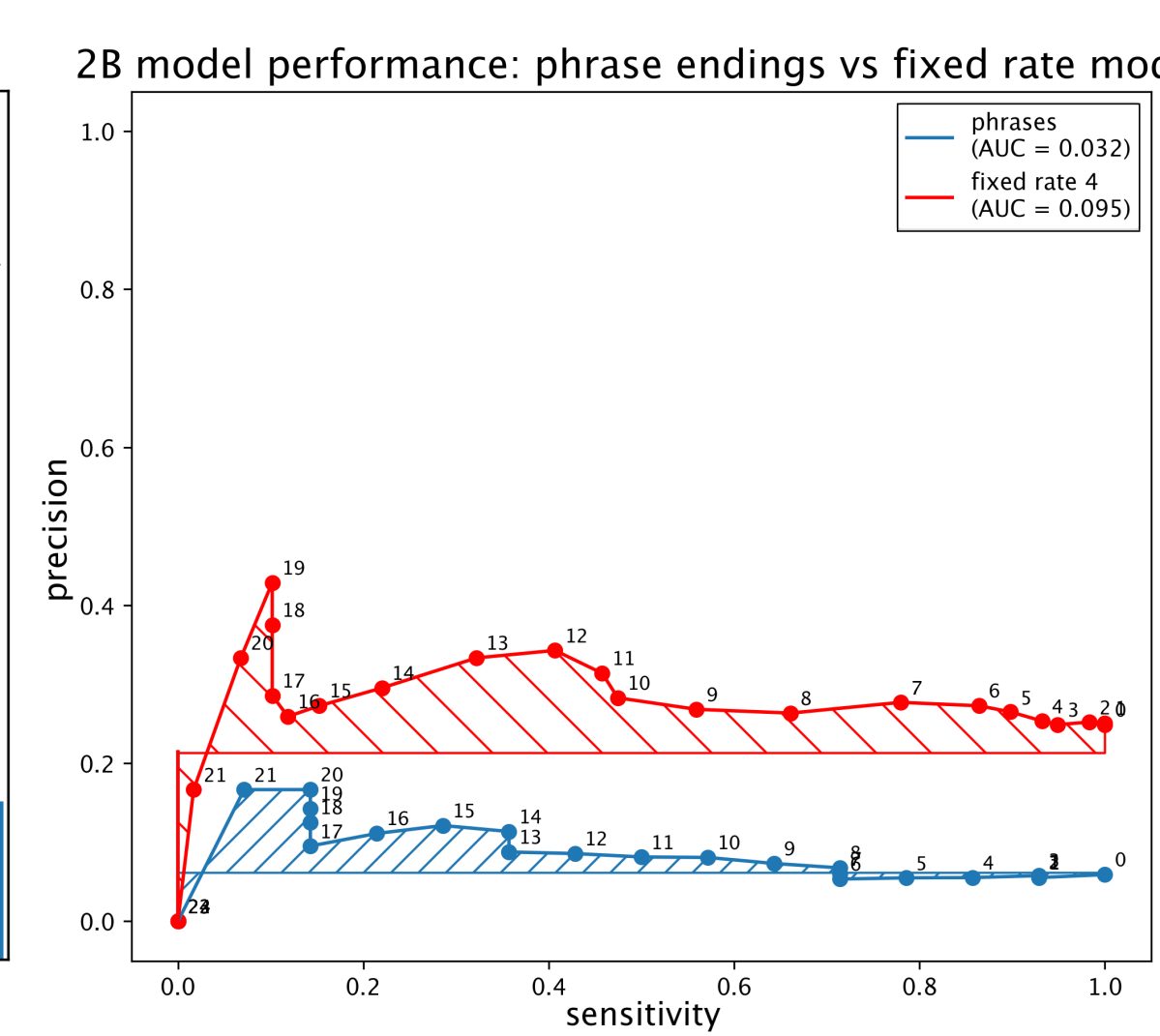
When section-level organization breaks down in phrase-scrambled music, subjects respond to phrase boundaries, but less reliably compared to Intact.



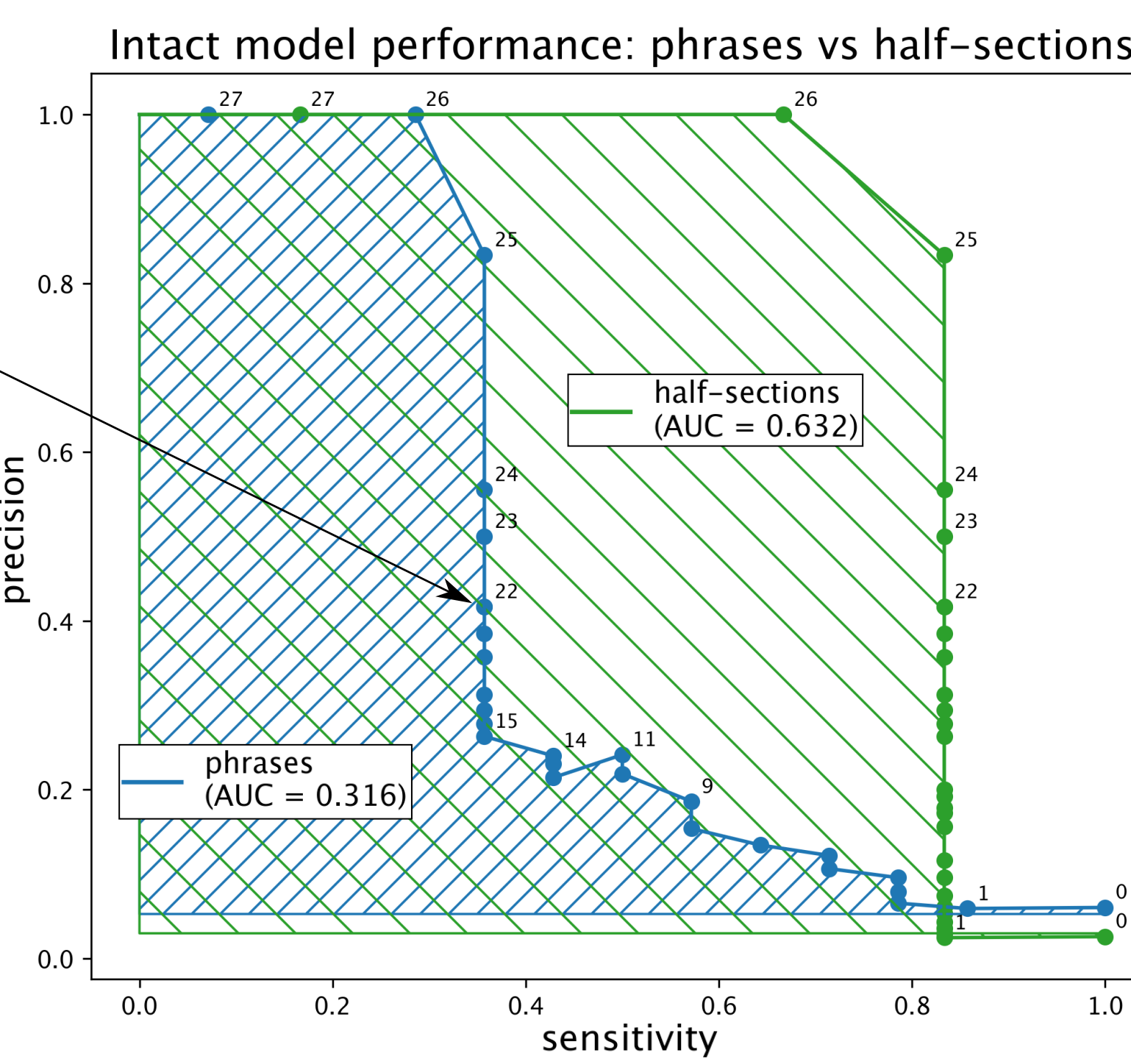
As expected, phrase boundaries explain response data better than "half-section" boundaries, since that level of structure was destroyed by the scramble.



When phrase-level organization breaks down in two-bar-scrambled music, subjects seem to respond to clear "ending-like" bars.



However, a fixed rate model (using the most common inter-response interval of 4s) explains the data better than bars with phrase endings.



Using measures of precision and sensitivity, section/half-section boundaries are a better model of subject responses compared to phrase boundaries.

Precision and sensitivity

What level of hierarchical structure is the data following most closely?

- Create different "ground truth" models of the stimuli using 1s and 0s.
- seconds with a boundary: 1, seconds without: 0

- For each model, compute precision and sensitivity at a range of thresholds.

What proportion of seconds captured by this threshold are seconds with true boundaries?

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{\text{true boundaries above threshold}}{\text{total seconds above threshold}}$$

What proportion of boundaries are captured by this threshold?

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{\text{true boundaries above threshold}}{\text{total true boundaries}}$$

- Assess model performance using area under the curve (AUC) and compare models.

- higher AUC: model better explains the data
- AUC range: 0.0 (null model) to 1.0 (perfect model)
- chance depends on number of boundaries

Summary

- Subjects' boundary judgements tend to agree best for the most intact music. When the music is more finely scrambled, subjects agree less and use more variable strategies.

- In Intact music, subjects perceive boundaries at half-section and section boundaries (the levels above phrases in the hierarchy).

- In 8B (phrase scrambled) music, subjects perceive boundaries at phrase boundaries, but the number of subjects that respond varies.

- In 2B (highly scrambled) music, subjects perceive boundaries at obvious closing gestures (e.g. half notes) as well as regular intervals of two bars (four seconds).

- Preliminary fMRI results suggest that expert pianists chunk music into shorter segments (phrases and half-phrases) when playing and longer segments (sections and half-sections) when listening.

Future directions

- Further analysis: does behavior differ depending on musical expertise, or more specifically piano experience?

- Expand stimulus set: Do these findings generalize to music with more complex structure? Is there an association between perceived phrase ending "strength" and subjects' responses?

- Expand fMRI dataset: scan more pianists to increase power for fMRI analyses.

- Event segmentation of neural data using Hidden Markov Models. Expanding on results of Williams et al. (2022): how does the brain segment music during production?

References

Farbood, M. M., Heeger, D. J., Marcus, G., Hasson, U., & Lerner, Y. (2015). The neural processing of hierarchical structure in music and speech at different timescales. *Frontiers in Neuroscience*, 9. <https://doi.org/10.3389/fnins.2015.00157>

Granot, R. Y., & Jacoby, N. (2011). Musically puzzling I: Sensitivity to overall structure in the sonata form? *Musicae Scientiae*, 15(3), 365–386. <https://doi.org/10.1177/1029864911409508>

Williams, J. A., Margulis, E. H., Nastase, S. A., Chen, J., Hasson, U., Norman, K. A., & Baldassano, C. (2022). High-Order Areas and Auditory Cortex Both Represent the High-Level Event Structure of Music. *Journal of Cognitive Neuroscience*, 34(4), 699–714. https://doi.org/10.1162/jocn_a_01815