Perception of melodies and triads at high frequencies

Daniel R. Guest and Andrew J. Oxenham
University of Minnesota, Department of Psychology, Auditory Perception and Cognition Lab

Introduction

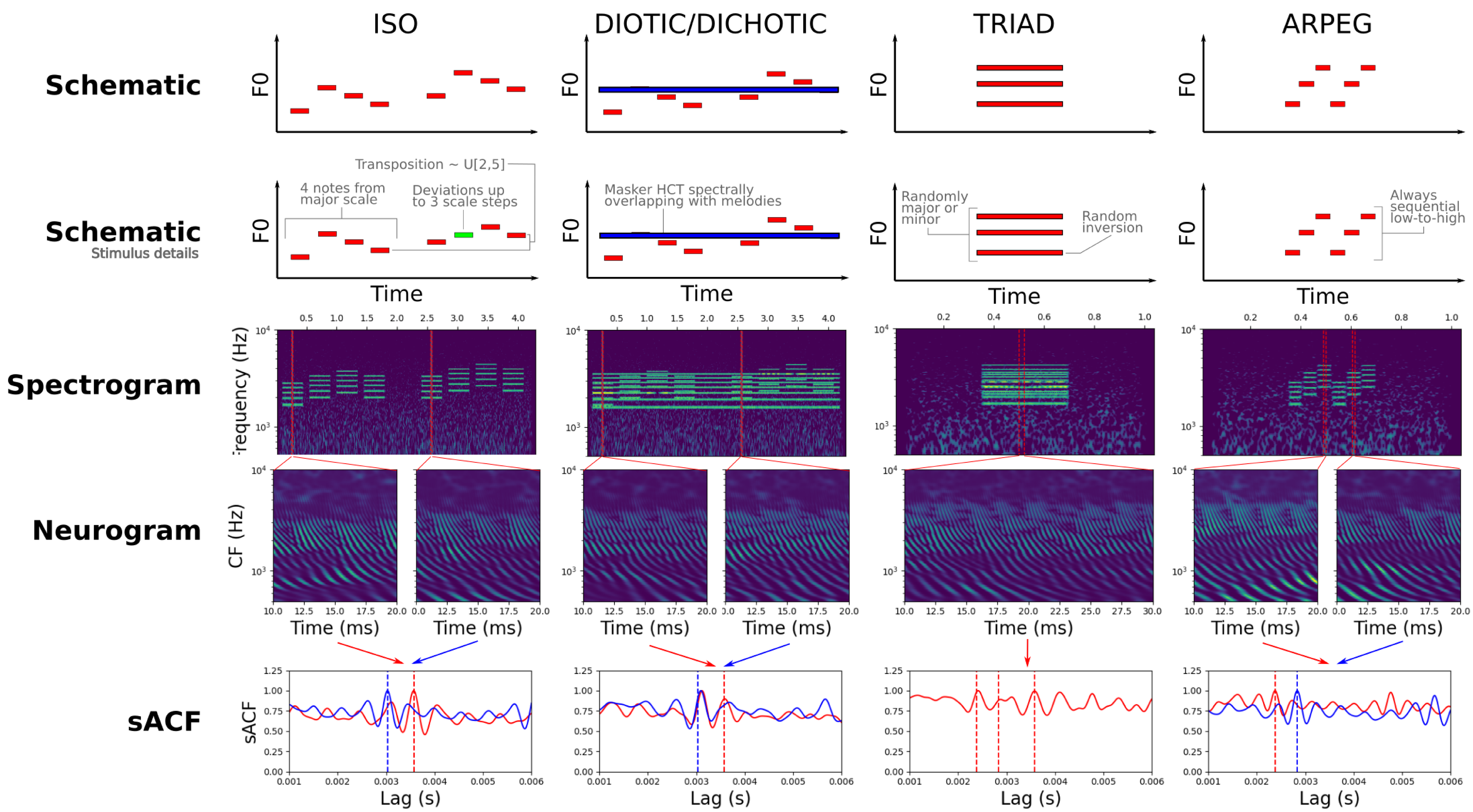
- Pitch perception of resolved complex tones can remain accurate even when (1) all harmonics are beyond the putative limits of phase locking [6, 2, 1] and when (2) tones are presented in the context of complex tone maskers [4, 3, 8]
- However, is is unknown whether accurate F0DLs in these cases translate to useful pitch perception in more realistic tasks
- **Can listeners extract and utilize pitch to perform musical tasks at high frequencies and in the context of complex tone maskers?**

Overview

- Tested **Low Freq** (~ 1680-2800 Hz) and **High Freq** (~ 7000-14000 Hz)
- **Melody discrimination**
  - Same-different identification for four-note melodies with and without single masker complex tone
- **Major/minor discrimination**
  - Major-minor discrimination for simultaneous and arpeggiated (sequential) triads

Stimuli & Methods

- **Targets:** Complex tones in threshold-equalizing noise (TEN) [5]
  - All harmonics of F0, bandpass filtered (12th-order zero-phase Butterworth, cutoffs at 5.5× and 10.5× nominal F0)
- **Maskers:** Complex tones
  - All harmonics of F0, bandpass filtered (12th-order zero-phase Butterworth, cutoffs at 4× and 12× nominal F0)
- **Frequency range:**
  - **Low Freq** (nominal F0 = 280 Hz ± 10% rove)
  - **High Freq** (nominal F0 = 1400 Hz ± 10% rove)
- **Durations:**
  - Pitch & melody discrimination — 350 ms per tone
  - Major/minor discrimination (triads) — 750 ms (short), 2250 ms (long)
  - Major/minor discrimination (arpeggios) — 125 ms per tone (short), 375 ms per tone (long)
- **Levels:**
  - 55 ± 3 dB SPL per component (pre-filtering), TEN at 43 dB SPL in ERB at 1 kHz
- **Procedure:** Percent correct measured in constant stimulus procedure
  - 25 trials per block, blocks presented in randomized order
  - 10 blocks per condition (but some data collection terminated early due to COVID-19)



Perception of melodies and triads at high frequencies

Daniel R. Guest and Andrew J. Oxenham
University of Minnesota, Department of Psychology, Auditory Perception and Cognition Lab

Melody discrimination at High Freq at chance, poor performance in context of complex tone maskers

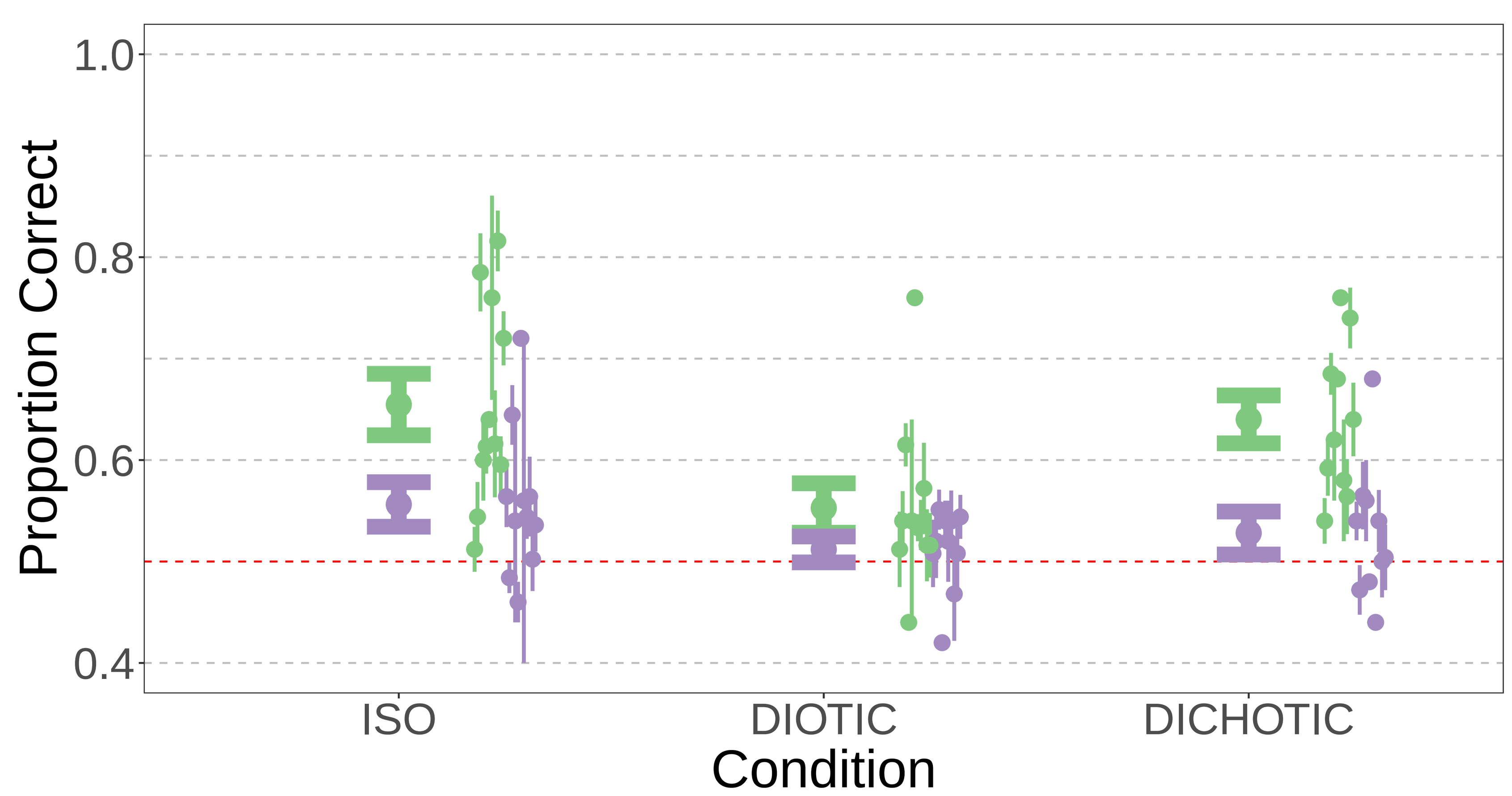


Figure 1: Means and ±1 SEM in the melody discrimination task. Listeners could discriminate same-different melodies at Low Freq, but not at High Freq, when melodies were presented in isolation (ISO) or with a contralateral masker (DICHOTIC). At High Freq few listeners performed above chance. When maskers were presented in the same ear (DIOTIC), most listeners performed near chance.

Listeners only detected a change in the melodies when they contained a contour change

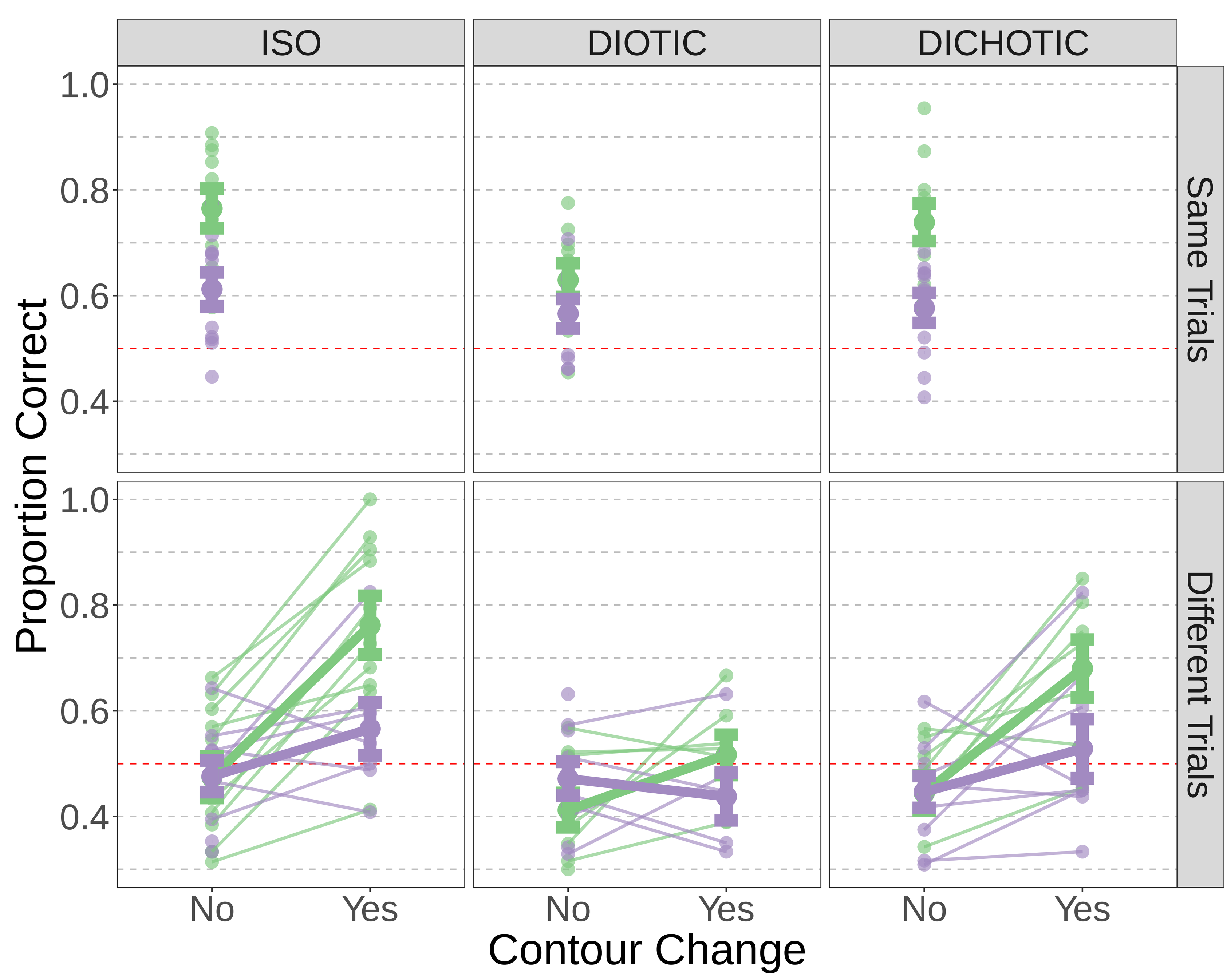


Figure 2: Means and ±1 SEM in the melody discrimination task, but separated by whether or not the trial contained a contour change in from one melody to another. A contour change was defined as the change of an upward interval to a downward interval, a downward interval to an upward interval, or no interval (i.e., two notes with the same frequency) to any non-zero interval. Listeners only reliably detected a different melody when it contained a contour change.

Temporal coding accurate at Low Freq and robust to the presence of maskers; little temporal information available at High Freq

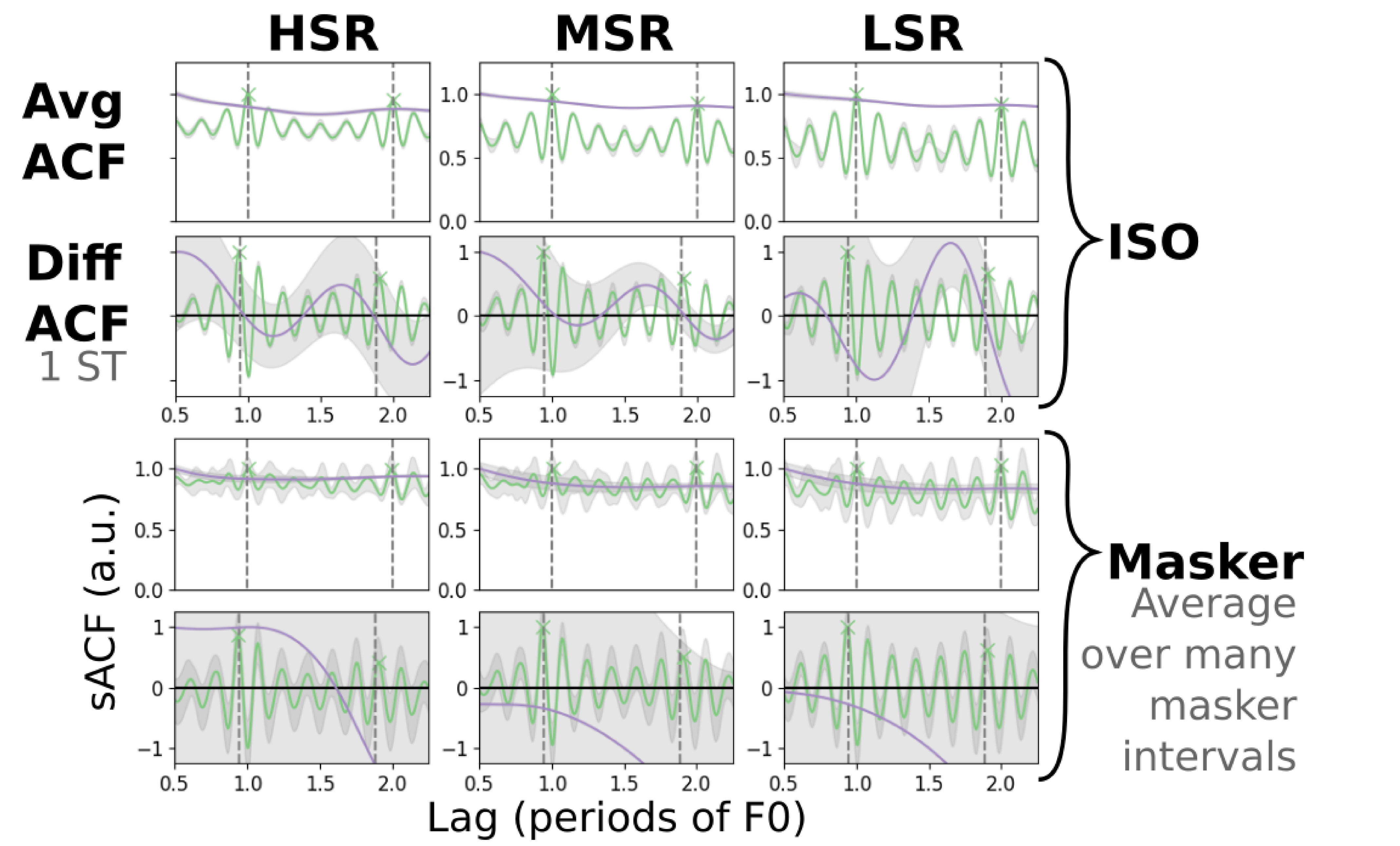


Figure 3: Normalized summary autocorrelation functions (ACFs) and changes in ACFs associated with a 1 ST shift in target F0 for simulated auditory nerve fibers responding to single notes from the melody stimuli. Full methods for the simulations are available at the end of the poster. Error bars indicate ±1 standard deviation. Temporal coding of F0 is accurate at Low Freq for isolated tones (top half) and, to some extent, for tones in the context of complex-tone maskers (bottom half).

Reliable rate-place cues for F0 changes in MSR and LSR fibers, even in the presence of a masker complex tone

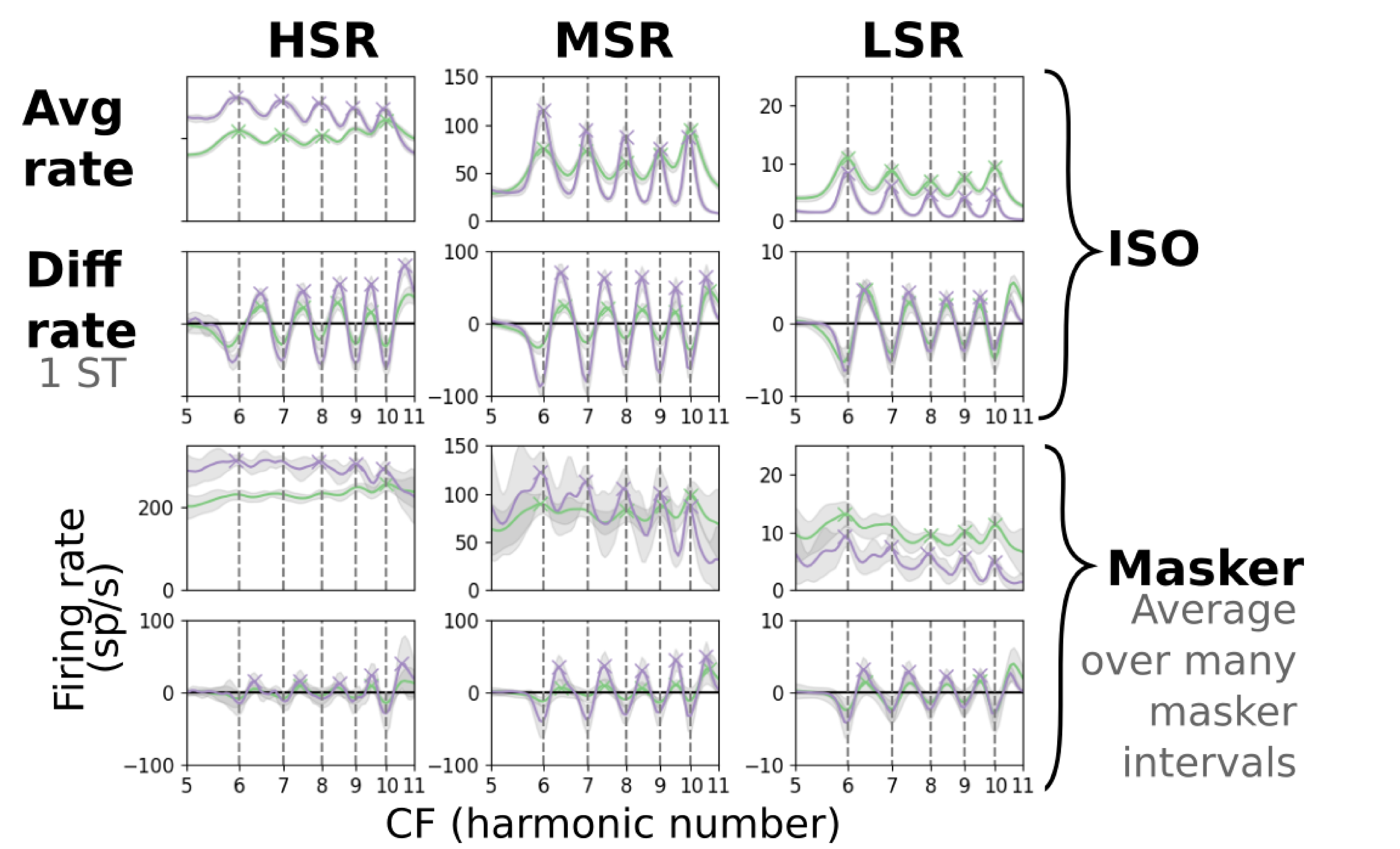


Figure 4: Excitation patterns and changes in excitation pattern associated with a 1 ST shift in target F0 for simulated auditory nerve fibers responding to single notes from the melody stimuli. Error bars indicate ±1 standard deviation. Rate-place coding of target harmonics, which is accurate for isolated tones (top half), is not robust to the presence of the masker complex tone (bottom half). However, a 1 ST change in the target F0 still produces reliable shifts in the excitation patterns consistent with an increase in F0, at least in MSR and LSR fibers.

Perception of melodies and triads at high frequencies

Daniel R. Guest and Andrew J. Oxenham
University of Minnesota, Department of Psychology, Auditory Perception and Cognition Lab

Listeners unable to discriminate major-minor triads and arpeggios at High Freq

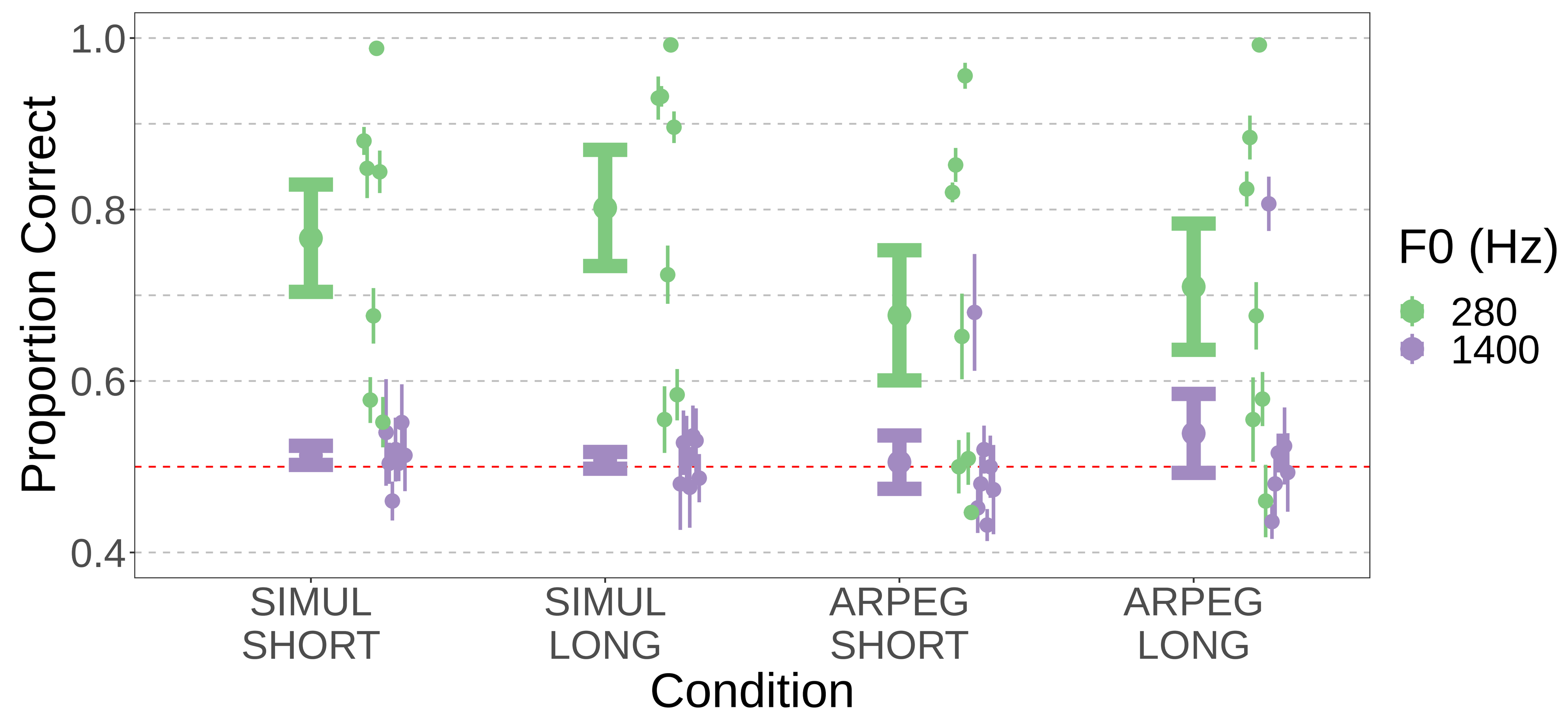


Figure 5: Means and ±1 SEM in the major-minor discrimination task. Most (but not all) listeners could discriminate major-minor triads and arpeggios at Low Freq. At High Freq, virtually no listeners could discriminate major-minor triads or arpeggios.

Comparable performance across chord qualities and inversions

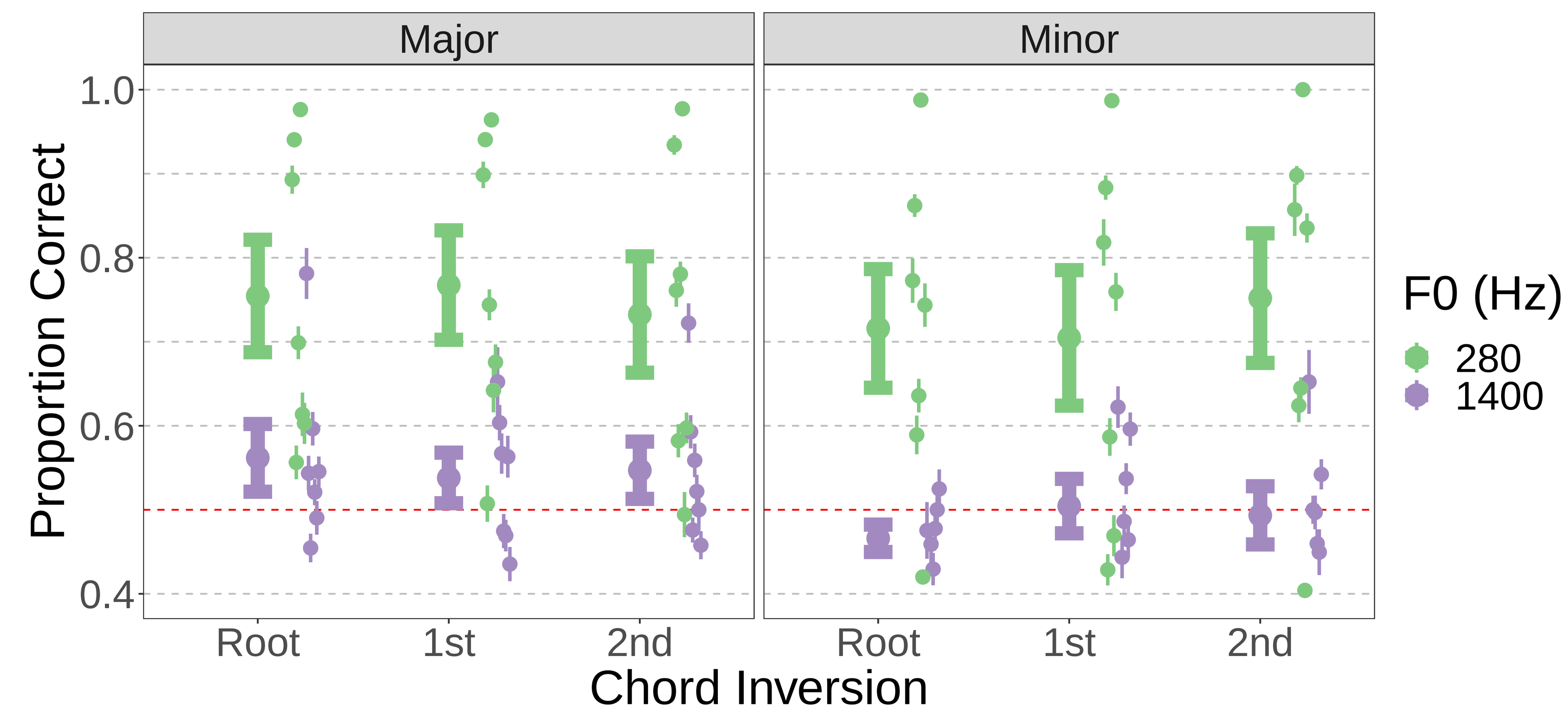


Figure 6: Means and ±1 SEM in the major-minor discrimination task, but separated by the quality and inversion of the chord and pooled across conditions. Quality and inversion had little reliable effect on discrimination performance.

Average rate and interspike interval counts both contained information about the quality of chords at Low Freq and High Freq

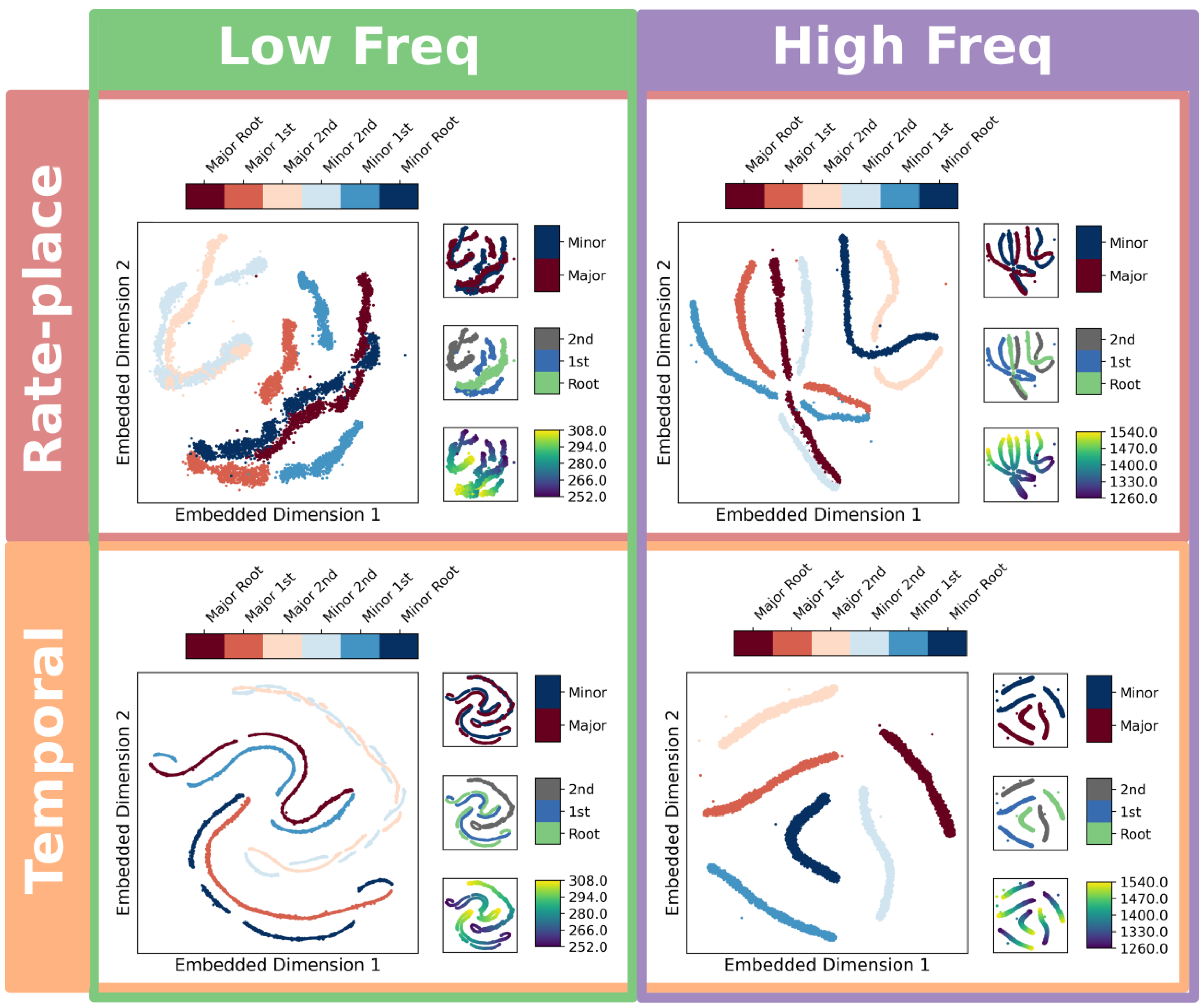


Figure 7: Embeddings of neural metrics derived from simulated spikes from high spontaneous rate auditory nerve fiber responses to the chord stimuli. The metrics were either average rate profiles (rate-place) or summary inter-spike interval histograms (temporal) and were embedded in a 2D space by PCA followed by t-distributed stochastic neighborhood embedding. The bottom-right subpanels in each section are colored according to the F0 of the middle voice

Neural network successfully trained to classify triads using average rate or interspike interval counts at Low Freq and High Freq

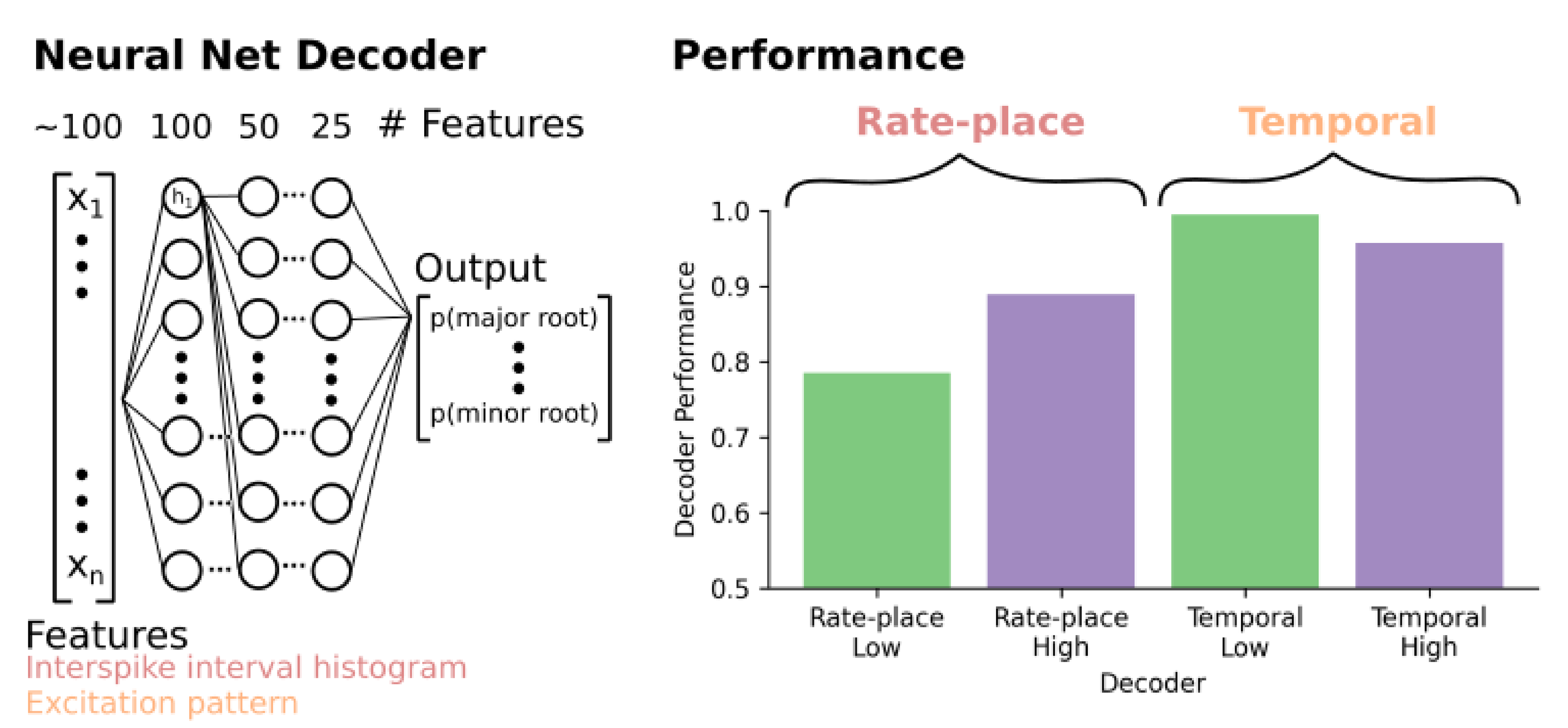


Figure 8: Schematic for and performance of a fully-connected neural network with three hidden layers trained to decode the chord type from the simulated spike trains from Figure 7. The results shown are for a held-out validation dataset composed of 1200 simulations, while the network was trained on a full dataset composed of 9000 simulations.

High Freq temporal decoding performance likely relies on decoding beat frequencies from interspike intervals

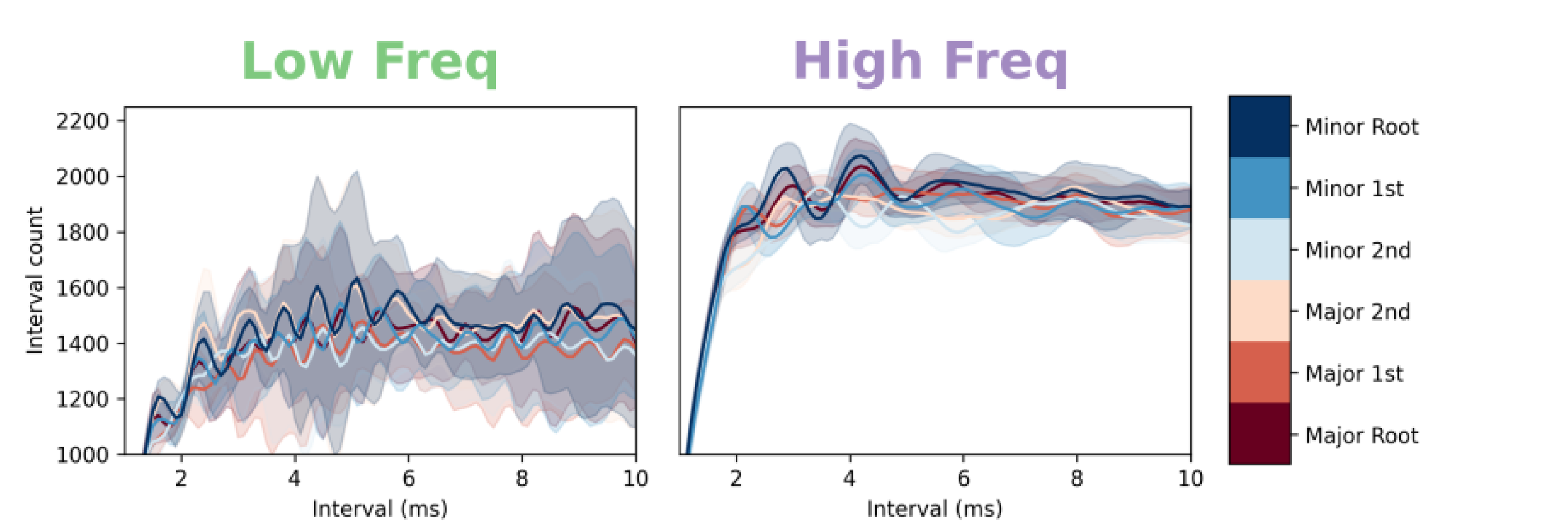


Figure 9: Average interspike interval histograms for the triad stimuli used as features in Figure 8. Line color indicates the chord type.

Perception of melodies and triads at high frequencies

Daniel R. Guest and Andrew J. Oxenham
University of Minnesota, Department of Psychology, Auditory Perception and Cognition Lab

Simulation methods

- Auditory nerve instantaneous firing rate and spike trains simulated using Zilany *et al.* [9] model implemented in Rudnicki *et al.* [7]

Figure	*Type*	*Output*	*Notes*
Figure 3	Firing rate	Autocorrelation (inverse Fourier transform of power spectrum) of instantaneous rate averaged across fibers	Average of 15 simulations in each panel
Figure 4	Firing rate	Average firing rate	Average of 15 simulations in each panel
Figure 7	Spikes	2D embedding of interspike interval histogram (0.1 ms bins, lag times less than 10 ms, summed across fibers) or average firing rate	9000 simulations per panel, `scikit-learn` PCA (n=50), and *t*-SNE (n=2, perplexity=60) used to embed
Figure 8	Spikes	Decoding performance for neural network trained on interspike interval histograms and average firing rates	Network trained on 9000 simulations, tested on test set of 1200 simulations; network was fully-connected with ReLU units and trained over 100 epochs
Figure 9	Spikes	Interspike interval histograms	Average of 1500 simulations per category

Acknowledgements

- A special thank you our research assistant, Tommy Tobin, for assisting with pilot testing, literature review, and data collection
- This work was supported by the following funding sources: UMN College of Liberal Arts Graduate Fellowship award to D.R.G., UMN Department of Psychology Summer Graduate Fellowship awarded to D.R.G., NSF NRT-UtB1734815, NIH R01 DC005216 awarded to A.J.O, and NIH F31 DC019247-01 awarded to D.R.G.

Conclusions

- Few listeners had above-chance melody discrimination at **High Freq** or in the context of maskers, and all listeners appeared to rely on contour changes to perform task
- No listeners had above-chance triad discrimination at **High Freq**, even when individual notes were isolated in time (and F0DLs are known to be good [2])
- No clear peripheral explanation of observed deficits
  - Sufficient information was likely available at the level of the auditory nerve to support accurate melody discrimination at **Low Freq** and **High Freq**, even in the context of a masker
  - Dimensionality reduction and neural network decoders demonstrated that sufficient information was available to discriminate major and minor triads at **Low Freq** and at **High Freq**

Bibliography

1. Gockel, H. E. & Carlyon, R. P. *Acta Acustica* **104,** 766–769 (2018).
2. Lau, B. K., Mehta, A. H. & Oxenham, A. J. *The Journal of Neuroscience* **37,** 9013–9021 (2017).
3. Micheyl, C., Bernstein, J. G. W. & Oxenham, A. J. *The Journal of the Acoustical Society of America* **120,** 1493–1505 (2006).
4. Micheyl, C., Keebler, M. V. & Oxenham, A. J. *The Journal of the Acoustical Society of America* **128,** 257–269 (2010).
5. Moore, B. C. J., Huss, M., Vickers, D. A., Glasberg, B. R. & Alcántra, J. I. *British Journal of Audiology* **34,** 205–224 (2000).
6. Oxenham, A. J., Micheyl, C., Keebler, M. V., Loper, A. & Santurette, S. *Proceedings of the National Academy of Science* **108,** 7629–7634 (2011).
7. Rudnicki, M., Schoppe, O., Isik, M., Völk, F. & Hemmert, W. *Cell Tissue Research* **361,** 159–175 (2015).
8. Wang, J. *et al.* *The Journal of the Acoustical Society of America* **132,** 339–356 (2012).
9. Zilany, M. S. A., Bruce, I. C. & Carney, L. H. *The Journal of the Acoustical Society of America* **135,** 283–286 (2014).

The Auditory Perception and Cognition Lab, 2019