Pitch perception of concurrent high-frequency complex tones
Modeling with auditory nerve simulations

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 fairly accurate even when all harmonics are beyond the putative limits of phase locking [8, 4, 1].
- Pitch perception of complex tones can also remain fairly accurate in the presence of complex tone maskers [6, 5, 10].
- However, is is unknown whether accurate pitch perception is possible with both (1) complex tone maskers and (2) targets entirely beyond the limits of phase locking.

Overview — behavior

- Tested Low Freq (~ 1680-2800 Hz) and High Freq (~ 7000-14000 Hz)
- **Pitch discrimination**
  - F0DLs with and without single masker complex tone
- **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

- **Targets:** Complex tones in threshold-equalizing noise (TEN) [7]
  - 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:**
  - Pitch discrimination — 50 ± 3 dB SPL per component (pre-filtering), TEN at 40 dB SPL in ERB around 1 kHz
  - Melody & major/minor discrimination — 55 ± 3 dB SPL per component (pre-filtering), TEN at 43 dB SPL in ERB around 1 kHz

Methods

- **Participants:** Young normal-hearing listeners
  - ≤ 20 dB HL at audiometric frequencies from 250 Hz - 8 kHz
- **Screening:**
  - Audibility — Masked thresholds in TEN ≤ 50 dB SPL for pure tones at 16 and 18 kHz
  - Pitch — F0DLs ≤ 6% at 280 Hz and ≤ 12% at 1400 Hz for stimulus without TEN
  - Melodies — ≥ 70% correct for melody discrimination for 280 Hz lowpass-filtered melodies
  - Major/minor — ≥ 70% correct for triad discrimination for lowpass-filtered 280 Hz triads
- **Data collection**
  - F0DLs measured with 7 1-up-3-down adaptive staircases per condition
  - Melody and major/minor discrimination measured with 10 blocks of 25 trials per condition

Acknowledgements

- This research was funded by a UMN College of Liberal Arts Graduate Fellowship awarded to D.R.G., NIH R01 DC005216 awarded to A.J.O., and NSF NRT-UtB1734815

Bibliography

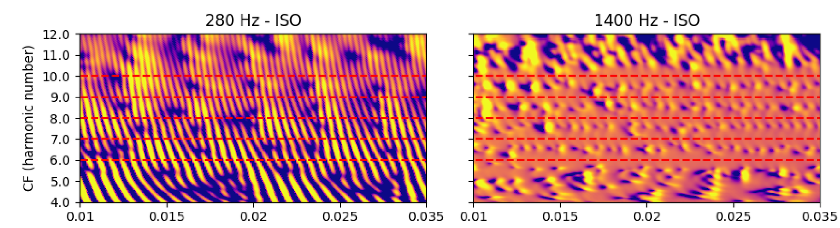
1. Gockel, H. E. & Carlyon, R. P. *Acta Acustica* **104,** 766–769 (2018).
2. Guest, D. R. & Oxenham, A. J. *The Journal of the Acoustical Society of America* **145,** 1782–1782 (2019).
3. Heinz, M. G. PhD thesis (Boston University, 2000).
4. Lau, B. K., Mehta, A. H. & Oxenham, A. J. *The Journal of Neuroscience* **37,** 9013–9021 (2017).
5. Micheyl, C., Bernstein, J. G. W. & Oxenham, A. J. *The Journal of the Acoustical Society of America* **120,** 1493–1505 (2006).
6. Micheyl, C., Keebler, M. V. & Oxenham, A. J. *The Journal of the Acoustical Society of America* **128,** 257–269 (2010).
7. Moore, B. C. J., Huss, M., Vickers, D. A., Glasberg, B. R. & Alcántra, J. I. *British Journal of Audiology* **34,** 205–224 (2000).
8. Oxenham, A. J., Micheyl, C., Keebler, M. V., Loper, A. & Santurette, S. *Proceedings of the National Academy of Science* **108,** 7629–7634 (2011).
9. Tan, Q. & Carney, L. H. *The Journal of the Acoustical Society of America* **120,** 1435–1445 (2006).
10. Wang, J. *et al.* *The Journal of the Acoustical Society of America* **132,** 339–356 (2012).
11. Zilany, M. S. A., Bruce, I. C. & Carney, L. H. *The Journal of the Acoustical Society of America* **135,** 283–286 (2014).

Overview — modeling

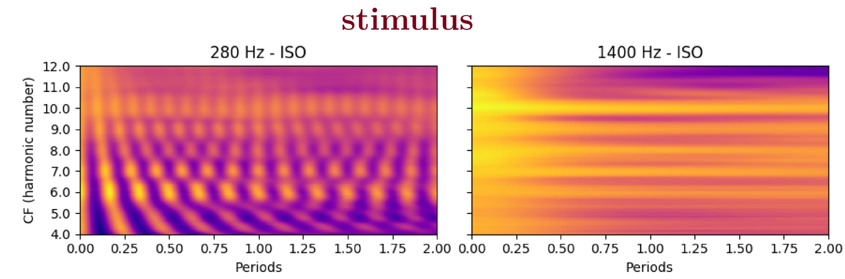
[1] Simulate auditory nerve responses for each stimulus configuration

- Firing rates from auditory nerve model of Zilany *et al.* [11]
- 80 CFs from 0.20 - 20 kHz, sampling rate of 300 kHz
- Mixture of 60% HSR, 20% MSR, and 20% LSR fibers

Time-CF firing rate response for ISO stimulus



Lag-CF autocorrelation of response for ISO stimulus



[2] Estimate F0DLs for each set of responses

- Model population activity of auditory nerve as joint distribution of nonhomogeneous Poisson processes [3, 9]
- Assume observer uses average response over many masker waveforms as template to assess competing hypotheses
  - Derive suboptimal "smart" observer by applying this constraint to form of optimal observer that has access to individual masker waveforms [3]

$$\mathrm{JND}_{F_0} = \left(\frac{\left(\sum_i \int_0^T \frac{1}{\bar{r}_i(t|F_0)} \left[\frac{\partial \bar{r}_i(t|F_0)}{\partial F_0} \right]^2 dt \right)^2}{\sum_i \int_0^T \frac{1}{\bar{r}_i(t|F_0)} \left[\frac{\partial \bar{r}_i(t|F_0)}{\partial F_0} \right]^2 dt + \mathrm{var}_w \left(\sum_i \int_0^T \frac{\partial \bar{r}_i(t|F_0)}{\partial F_0} \frac{r_i(t|F_0,w)}{\bar{r}_i(t|F_0)} dt \right)} \right)^{-1/2}$$

Details

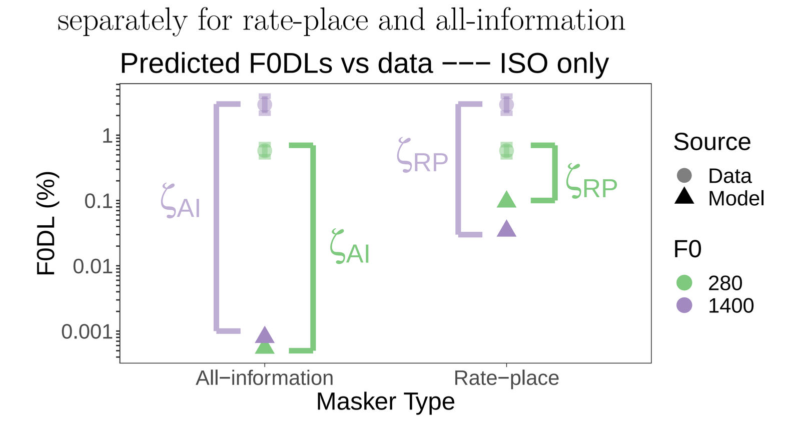
- r_i — firing rate of i-th nerve fiber
- w — index for random masker waveforms
- $\bar{r}_i$ — firing rate of i-th nerve fiber, averaged across random stimulus waveforms

Intuitions

- Change in firing rate w/ respect to F_0
- Variance due to Poisson randomness
- Variance due to randomness of masker waveforms

[3] Scale ideal observer F0DLs to behavioral range

- Use F0DL data from previous experiment [2]
- Estimate scaling factors to match model F0DLs to behavior
- Perform separately for Low Freq and High Freq, separately for rate-place and all-information

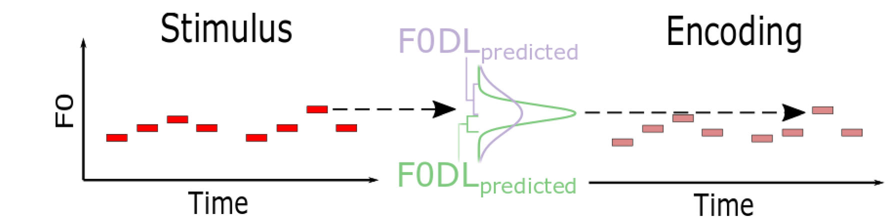


- Predict behavioral F0DLs as scaled copy of estimated model F0DLs

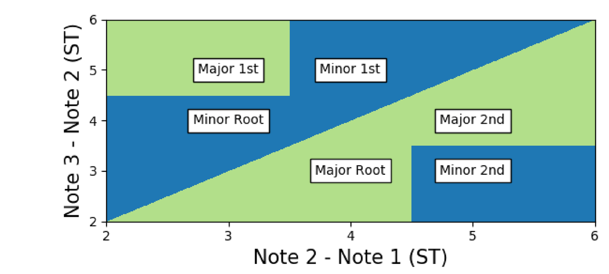
$$\mathrm{F0DL}_{\mathrm{predicted}} = 10^{\log_{10}(\mathrm{F0DL}_{\mathrm{model}})+\zeta}$$

[4] Transform F0DLs into percent correct for music tasks

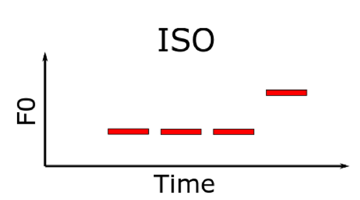
- Use predicted F0DLs as measure of internal noise for encoding notes in 10000 simulated trials of melodies and triads



- **Melodies:** Respond same if encodings of notes in two melodies are the same, respond different otherwise
- **Major/minor:** Respond with the key of the nearest triad in 2D interval space to encoded intervals

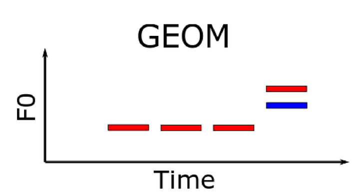


Results — Pitch discrimination



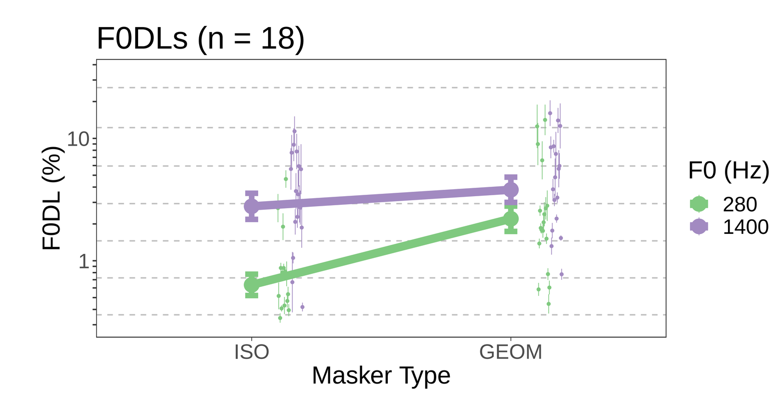
No masker complex tone

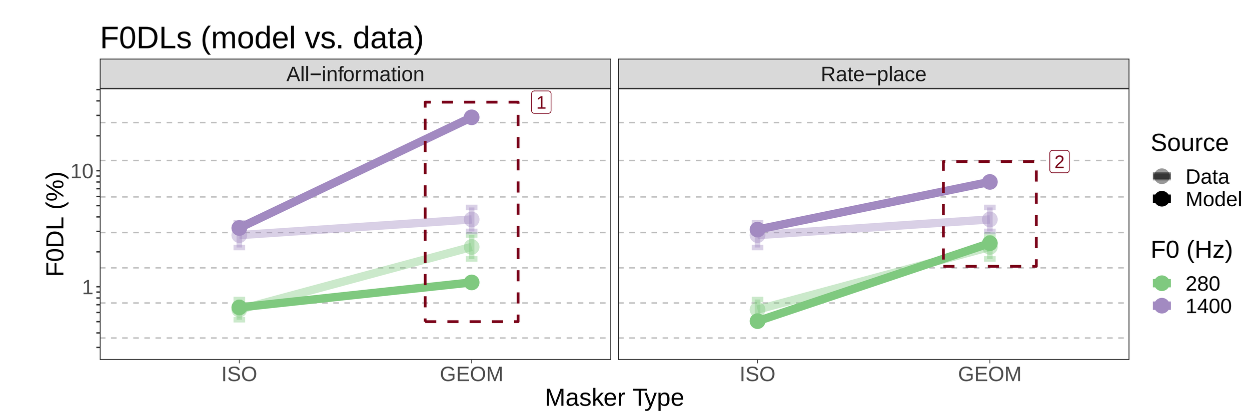
Indicate direction of ΔF0



Masker complex tone (centered geometrically between target and reference F0s)

Indicate direction of ΔF0



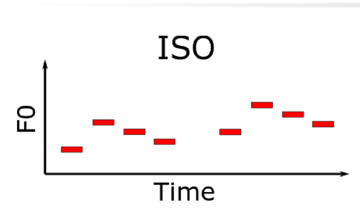


- **F0 discrimination significantly worse at High Freq and in presence of spectrally overlapping maskers**

❶ All-information observer accounts poorly for effect of masker on F0DLs

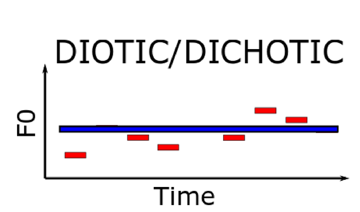
❷ Rate-place observer accounts better for effect of masker on F0DLs

Results — Melody discrimination



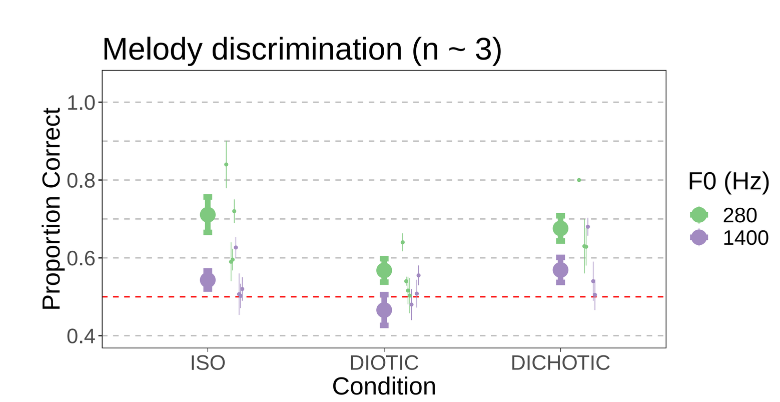
Two four-note melodies in major scale
- Randomly transposed 2-5 semitones across intervals

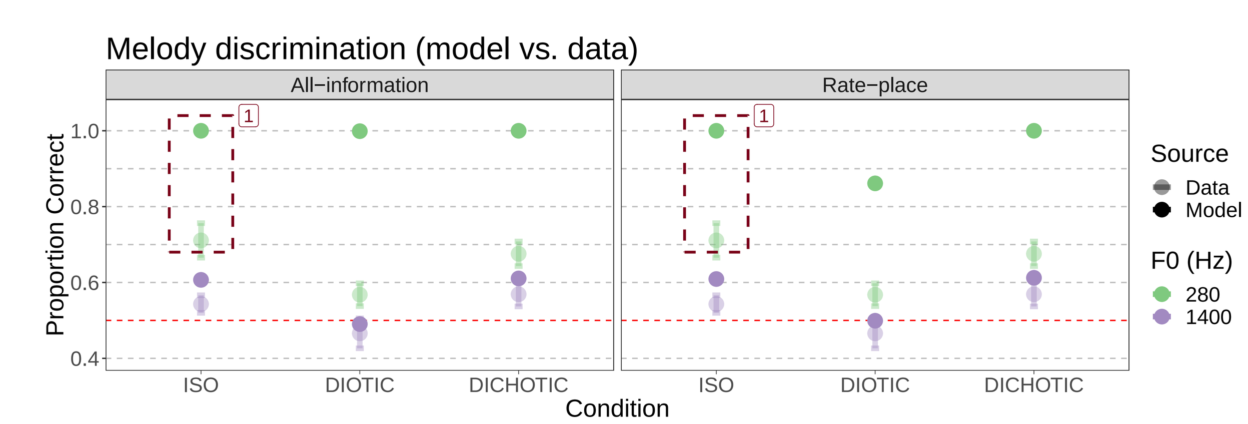
Indicate if melodies are same or different



Two four-note melodies in major scale with drone masker (centered geometrically at average of all notes)
- Diotic: masker in both ears
- Dichotic: masker in left, target in right

Indicate if melodies are same or different

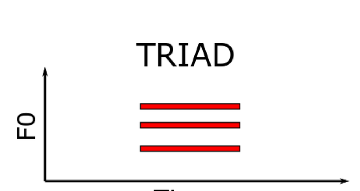




- **Melody discrimination significantly worse at High Freq and in presence of spectrally overlapping maskers**

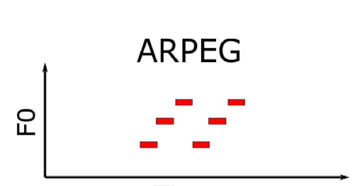
❶ Model overestimates performance, may suggest that non-peripheral limitations on melody discrimination play important role
- Elements of paradigm that were not modeled (e.g., transposition) may also explain some portion of gap

Results — Major/minor discrimination



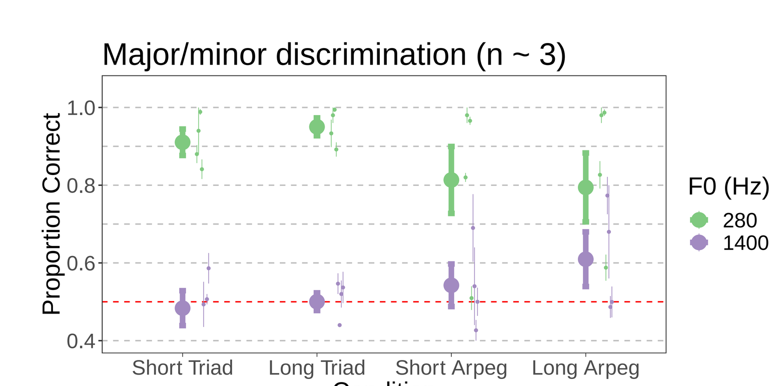
Major or minor triads (root, 1st, or 2nd inversion)

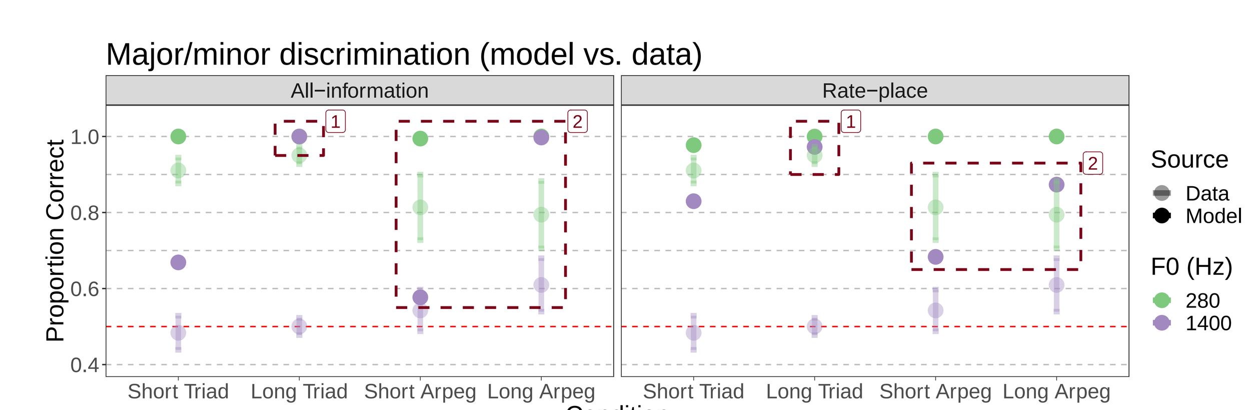
Indicate if major or minor



Six-note major or minor arpeggio (root, 1st, or 2nd inversion)

Indicate if major or minor





- **Listeners *could not* perform major/minor discrimination for High Freq triads**
- **Some listeners *could* perform major/minor discrimination of long High Freq arpeggios**

❶ Models predict long triads should be discriminable at Low Freq and High Freq, but listeners could only discriminate Low Freq triads

❷ All-information model predicts much larger benefit of duration than rate-place model — effect of duration in behavioral data unclear