Pitch perception of concurrent high-frequency complex tones

Modeling behavior with auditory nerve simulations

Daniel R. Guest and Andrew J. Oxenham

University of Minnesota, Department of Psychology, Auditory Perception and Cognition Lab

Introduction

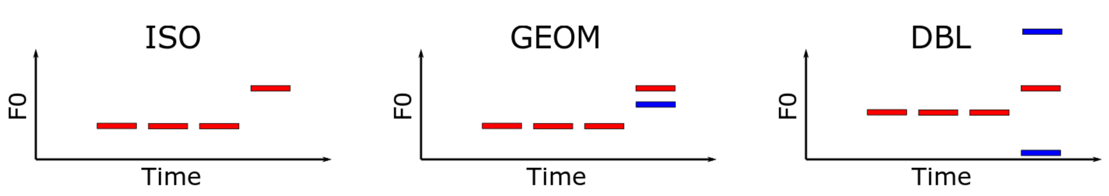
- Pitch perception can remain fairly accurate even when...
  - ...all harmonics are beyond the limits of phase locking [2] OR
  - ...targets are presented in the presence of complex tone maskers [3]
- This research aims to determine whether accurate pitch perception is possible under both of these conditions simultaneously, and to investigate the neural mechanisms underlying any such ability

Overview — Behavior

- **Paradigm:** Listeners heard three tones with same F0 (reference) followed by one tone with different F0 (target) mixed with maskers and indicated direction of F0 change
- **Experiments:**
  - **Exp. 1a and Exp. 1b:** F0DLs w/ and w/o masker tone
  - **Exp. 2:** Target-to-masker ratio (TMR) required for fixed interval w/ two masker tones

Methods

- **Targets:** Complex tones in threshold-equalizing noise (TEN) [4]
  - **Exp. 1a** harmonics 6-10 of F0
  - **Exp. 1b and Exp. 2:** all harmonics of F0, bandpass filtered (5.5× to 10.5× nominal F0) with 12th order Butterworth
  - 50 ± 3 dB SPL per component, random phase, 350 ms w/ 50 ms gaps, TEN at 40 dB SPL in 1 kHz ERB
- **Maskers:** Complex tones
  - **Exp. 1a:** harmonics 5-11 of F0
  - **Exp. 1b and Exp. 2:** all harmonics of F0, bandpass filtered (4× to 12× nominal F0) with 12th order Butterworth
  - 50 ± 3 dB SPL per component, random phase, 350 ms w/ 50 ms gaps
- **Frequency conditions:**
  - Low Freq (nominal F0 = 280 Hz ± 10% rove)
  - High Freq (nominal F0 = 1400 Hz ± 10% rove)
- **Masker conditions:**



No masker | Masker tone geometrically centered between target and reference | One masker tone above and one masker tone below target (at least 5.25 semitones)

- **Interval sizes:**
  - **Exp. 1:** Adaptive
  - **Exp. 2:** 1.5× and 2.5× F0DL from Exp. 1 **ISO**
- **Participants:**
  - Normal-hearing students at University of Minnesota with pure tone detection thresholds better than 35 dB SPL in TEN

Results

F0DLs worse in High Freq than Low Freq & GEOM masker had larger effect in Low Freq than High Freq

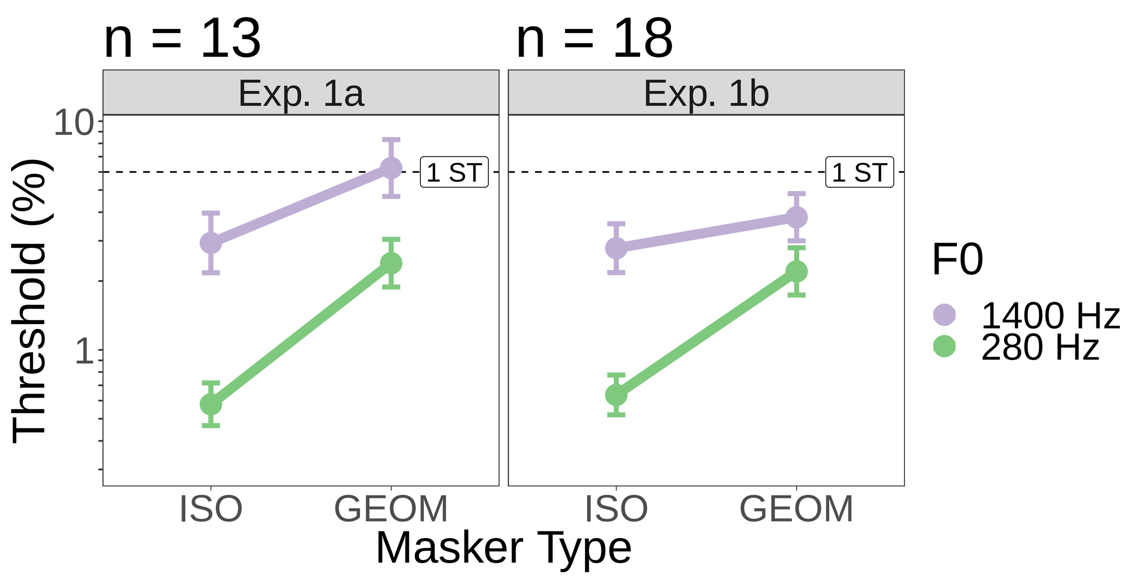


Figure 1: Results from 7 runs per condition for Exp. 1a and Exp. 1b. Error bars indicate ±1 SEM.

Higher TMRs required in High Freq than Low Freq to achieve same performance

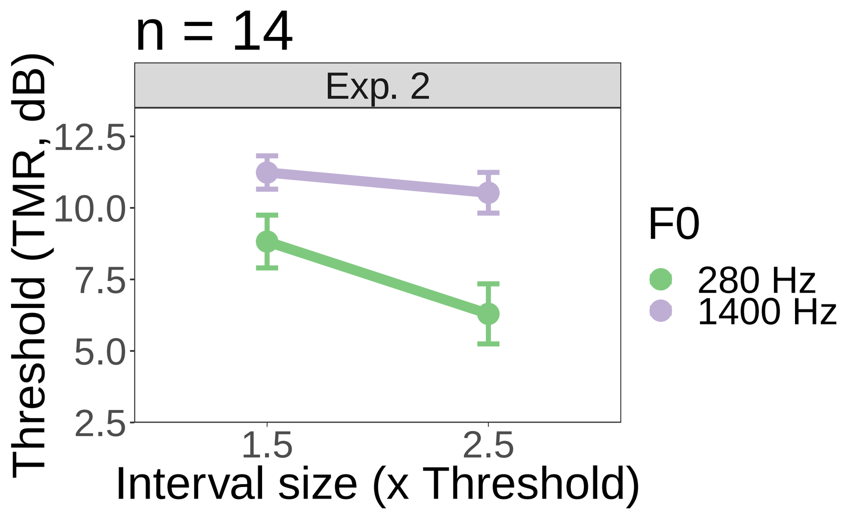


Figure 2: Results from 7 runs per condition for Exp. 2. Error bars indicate ±1 SEM. Masker type for all conditions was **DBL**.

Bibliography

[1] Heinz, M. G. PhD thesis. Boston University, 2000.

[2] Lau, B. K., Mehta, A. H., and Oxenham, A. J. In: *The Journal of Neuroscience* (2017). DOI: 10.1523/JNEUROSCI.1507-17.2017.

[3] Micheyl, C., Keebler, M. V., and Oxenham, A. J. In: *The Journal of the Acoustical Society of America* 128.1 (2010), pp. 257–269. DOI: 10.1121/1.3372751.

[4] Moore, B. C. J. et al. In: *British Journal of Audiology* 34.4 (2000), pp. 205–224. DOI: 10.3109/03005364000000131.

[5] Shera, C. J., Guinan, J. J., and Oxenham, A. J. In: *Proceedings of the National Academy of Sciences* 99.5 (2002), pp. 3318–3323. DOI: 10.1073/pnas.032675099.

[6] Tan, Q. and Carney, L. H. In: *The Journal of the Acoustical Society of America* 120.3 (2006), pp. 1435–1445. DOI: 10.1121/1.2225858.

[7] Zilany, M. S. A., Bruce, I. C., and Carney, L. H. In: *The Journal of the Acoustical Society of America* 135.1 (2014), pp. 283–286. DOI: 10.1121/1.4837815.

Overview — Modeling

[1] Simulate neural responses
- Simulate population responses to the stimuli at the level of the auditory nerve
  - Auditory nerve simulation of Zilany, Bruce, and Carney [7]

→

[2] Decode neural responses
- Define statistical model for population activity
- Derive sensible estimator for F0 and use to predict thresholds

→

[3] Predict and explore
- Predict behavioral findings:
  - High Freq < Low Freq
  - **GEOM < ISO**
  - F0 × masker interaction
- Explore effect of key parameters

[1] Simulate neural responses

- Auditory nerve simulations
  - 80 characteristic frequencies (CFs) log distributed from 200 Hz to 20 kHz
  - Each CF innervated by multiple nerve fibers
    - Counts selected assuming total of 20000 fibers (60% HSR, 20% MSR, 20% LSR from 200 Hz to 20 kHz)
  - Freq. tuning estimates from Shera, Guinan, and Oxenham [5]
- Stimulus parameters
  - Stimulus parameters generally matched experiment, except:
    - Target level/phase fixed
    - **GEOM** masker interval fixed at 1% and 3% in Low Freq and High Freq (respectively) to roughly match behavioral task
  - 300 kHz sampling rate

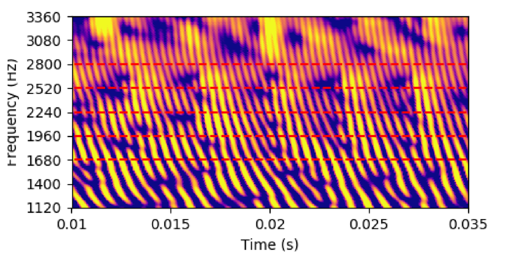


Figure 3: Simulated neurogram for a 280 Hz **ISO** stimulus

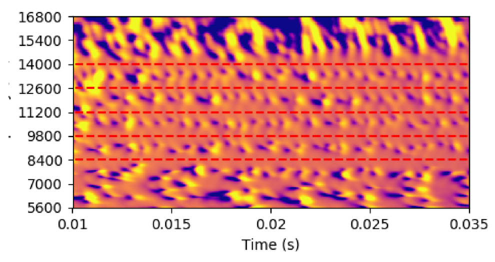


Figure 4: Simulated neurogram for a 1400 Hz **ISO** stimulus

[2] Decode neural responses

- Model population activity of auditory nerve as joint distribution of nonhomogeneous Poisson processes [1, 6]
- Assume observer uses average neural response over many random 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 [1]

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

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 (generally 100)

Intuitions

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

[3] Predict and explore

Decoding rate *and* timing cues predicts sign of main effects in Exp. 1a but underestimates difference between Low Freq and High Freq and does not correctly predict F0 × masker interaction

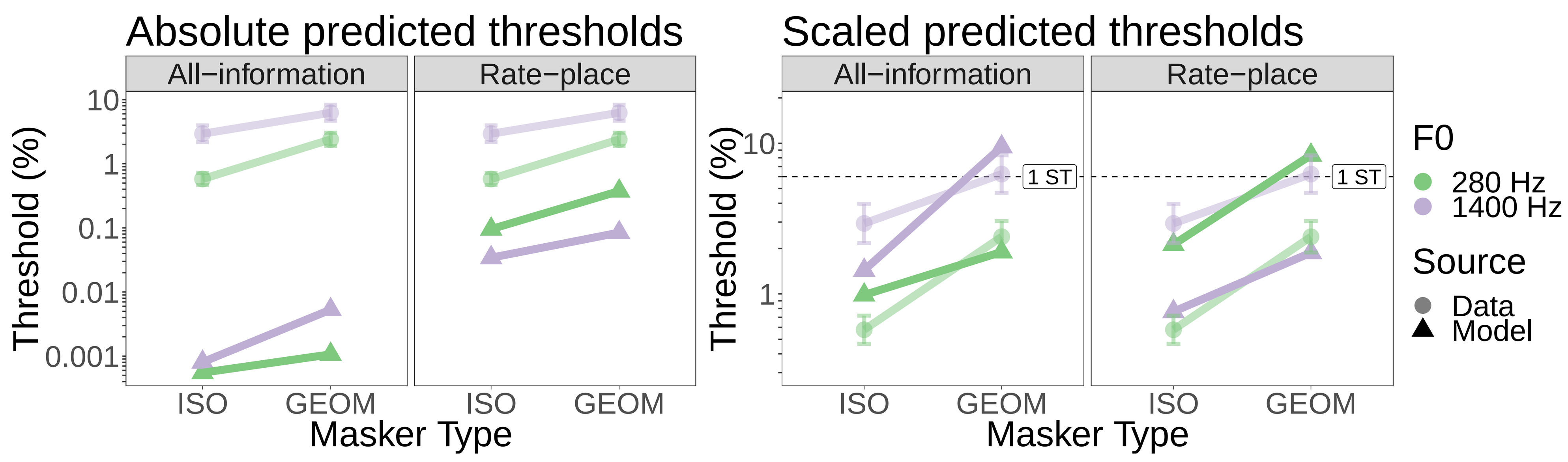


Figure 5: Predicted thresholds, both absolute (left) and scaled by a constant factor (separately for each decoding type) to minimize mean square error between predicted and actual thresholds (right).

When F0-specific scaling factors are applied, decoding rates alone provides best account of GEOM masker — scaling factors generalize to new stimulus variant

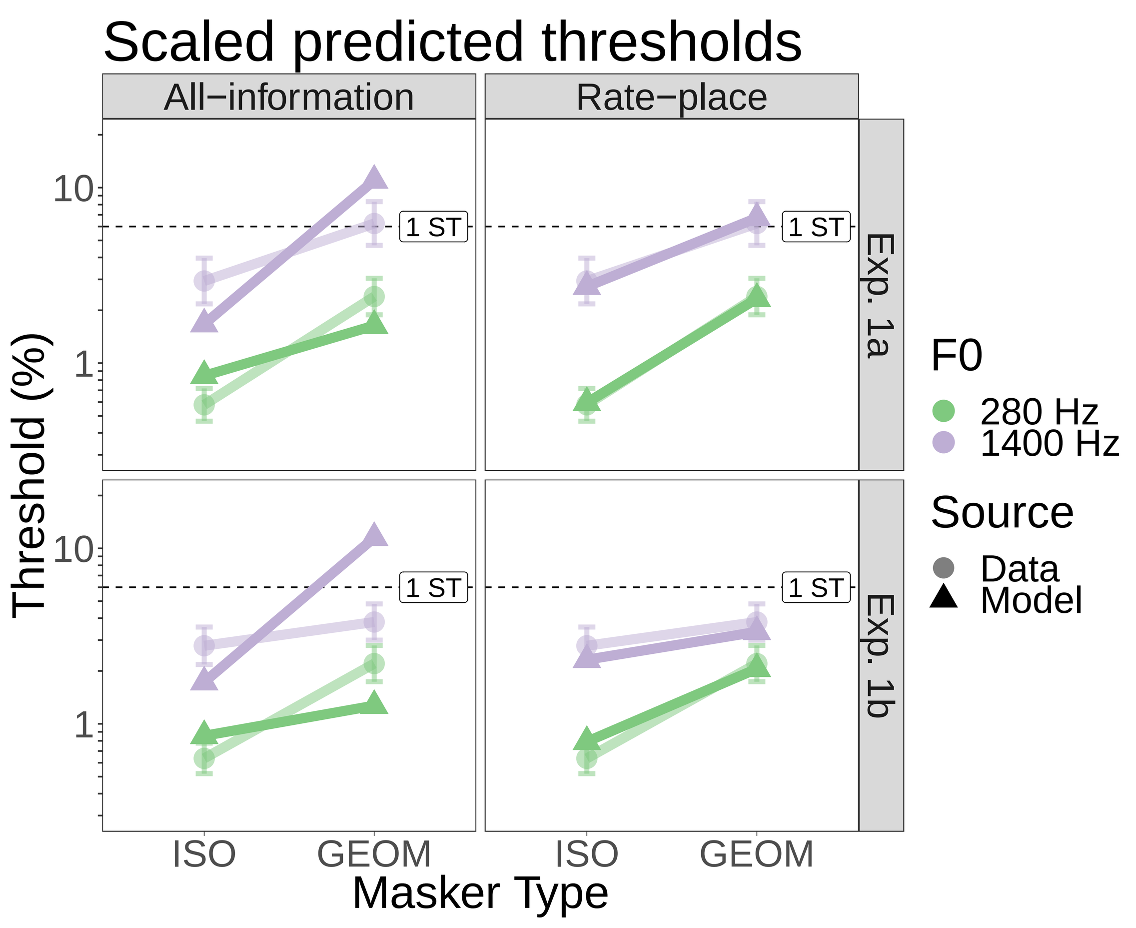


Figure 6: Predicted thresholds for Exp. 1a and 1b. Scaling factors were estimated to minimize mean square error between model predictions and data from Exp. 1a, separately for each decoding type *and* for each F0. Then, these scaling factors were applied to simulations from both Exp. 1a and Exp. 1b. F0-specific scaling factors could correspond to decoding efficiency that differs between F0s or spectral regions.

Better spectral resolvability of harmonics in High Freq (particularly in LSR fibers) may underlie better performance of rate-place observer at High Freq

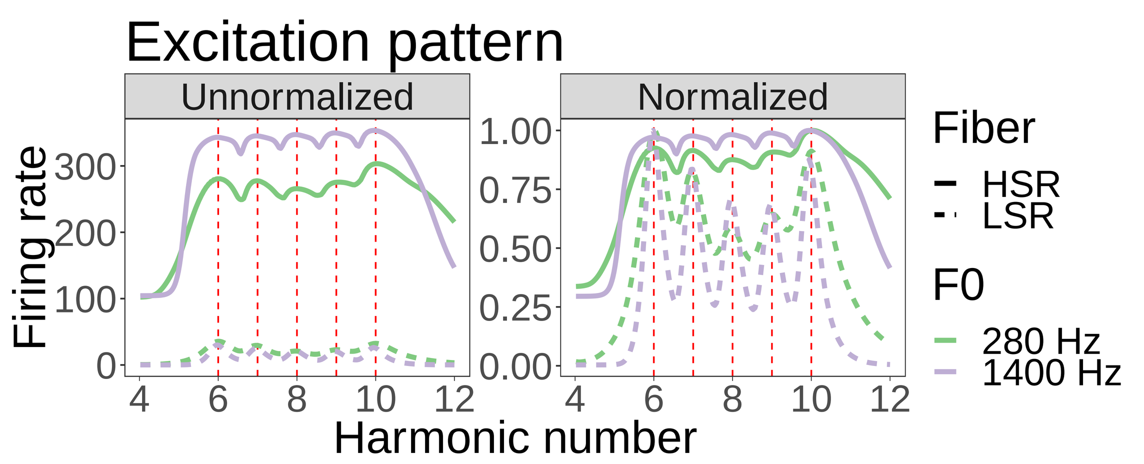


Figure 7: Excitation patterns (average firing rate as a function of CF) produced by **ISO** stimulus without TEN. Based on a simulation with 200 CFs spanning from 4× to 12× F0.

High Freq timing information comes from envelope-locked response at F0

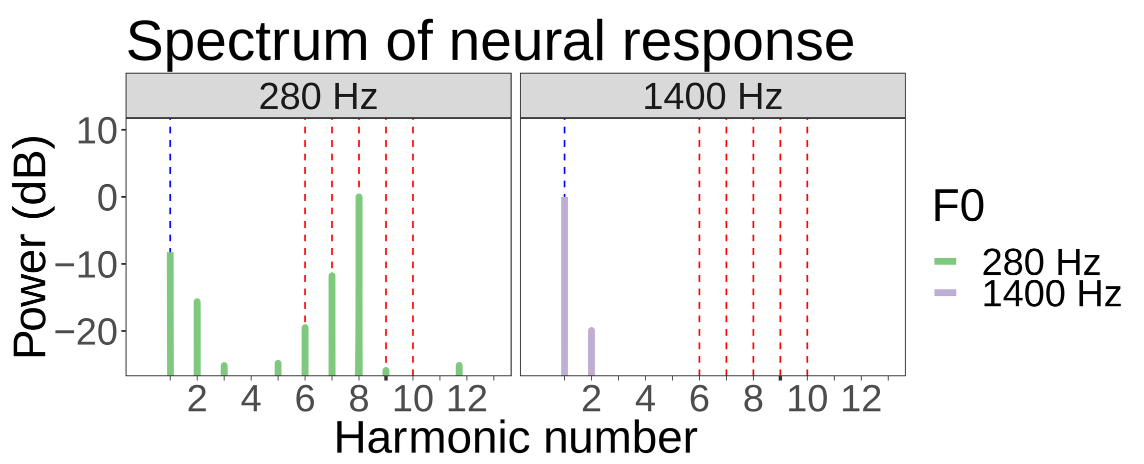


Figure 8: Log power spectrum of response of HSR fiber tuned to 8th harmonic for **ISO** stimulus (averaged over 20 waveforms). Responses were demeaned before calculation to eliminate DC component. Previous behavioral findings suggest that, although the neural observer may use this cue, humans likely do not [2].

Conclusions

- Decoding of auditory nerve rate information alone predicts impact of complex tone maskers on pitch perception at both Low Freq and High Freq
- Assumption of poorer decoding efficiency at High Freq is needed to fit all data with rate information alone
  - With no such assumptions, neither all-information nor rate-place model fit well — some task variance is likely non-peripheral
- Present modeling strategy can provide general insight into neural coding of complex pitch
  - Gap between all-information and rate-place predictions → significant timing information at both Low Freq and High Freq (but listeners seem not to always use it [2])
  - Sharper auditory filter tuning at high frequencies may play key role in explaining Low Freq and High Freq differences [5]

Future directions

- Extend modeling methodology to auditory brainstem/midbrain
- Model behavioral results from Exp. 2
- Use complex stimuli with other pitch tasks (e.g., melody perception)

Acknowledgements

- Funding from: UMN College of Liberal Arts Graduate Fellowship, NIH R01 DC005216, NSF NRT-UtB1734815