Fundamental frequency discrimination in mixtures of high-frequency complex tones: Data and ideal-observer model predictions

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

Introduction

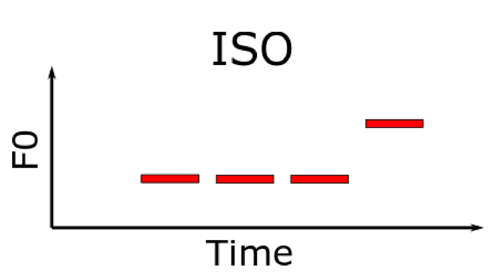
- Pitch perception of resolved complex tones can remain fairly accurate even when all harmonics are beyond the putative limits of phase locking [1, 2, 3, 4].
- Pitch perception of complex tones can also remain fairly accurate in the presence of complex tone maskers [5, 6, 7].
- However, is is unknown whether accurate pitch perception is possible with both (1) complex tone maskers and (2) targets entirely beyond the limits of phase locking.

Overview

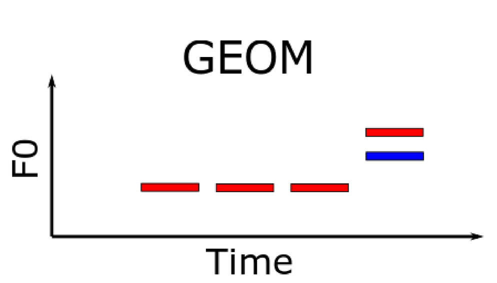
- Tested **Low Freq** (~ 1680-2800 Hz) and **High Freq** (~ 7000-14000 Hz) conditions
- **Behavior — F0 discrimination**
  - F0DLs with and without single simultaneous masker complex tone
  - Target-to-masker ratio (TMR) required for fixed F0 difference with two masker tones
- **Computational model — ideal observer**
  - Simulated firing rates in auditory-nerve models and excitatory-inhibitory coincidence detector models
  - Calculated ideal F0 discrimination thresholds based on simulations

Stimuli and task

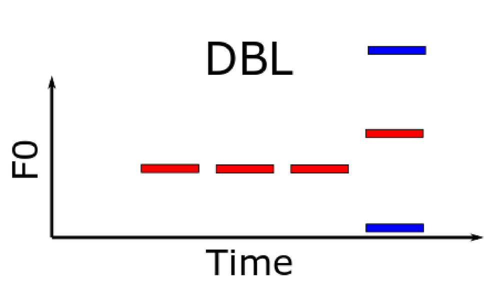
- **Targets:** 350 ms complex tones in threshold-equalizing noise (TEN) [8]
  - All harmonics of F0, bandpass filtered (12th-order zero-phase Butterworth, cutoffs at 5.5× and 10.5× nominal F0)
- **Maskers:** 350 ms complex tones
  - All harmonics of F0, bandpass filtered (12th-order zero-phase Butterworth, 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)
- **Levels:**
  - 50 ± 3 dB SPL per component (pre-filtering), TEN at 40 dB SPL in ERB around 1 kHz
- **Task:** "Was the last tone higher or lower?"
- **Masker conditions:**



Target tones without masker



Target tones with single masker tone centered geometrically between F0s of target tones



Target tones with two masker tones flanking the F0s of the target tones by 5-7 semitones (random uniform)

Procedure

- **Participants:** Young normal-hearing listeners with range of musical and psychoacoustical experience
- **Screening:**
  - Hearing status — ≤ 20 dB HL at audiometric frequencies from 250 Hz - 8 kHz
  - Audibility — Masked thresholds in TEN ≤ 50 dB SPL for pure tones at 16 and 18 kHz
  - Pitch — F0DLs $\leq 6\%$ at 280 Hz and $\leq 12\%$ at 1400 Hz for stimulus without TEN
- **Data collection**
  - F0DLs measured with seven 1-up-3-down adaptive staircases per condition per participant
  - TMRs measured with seven 1-up-3-down adaptive staircases per condition per participant

Results

Sub-semitone F0DLs at High Freq

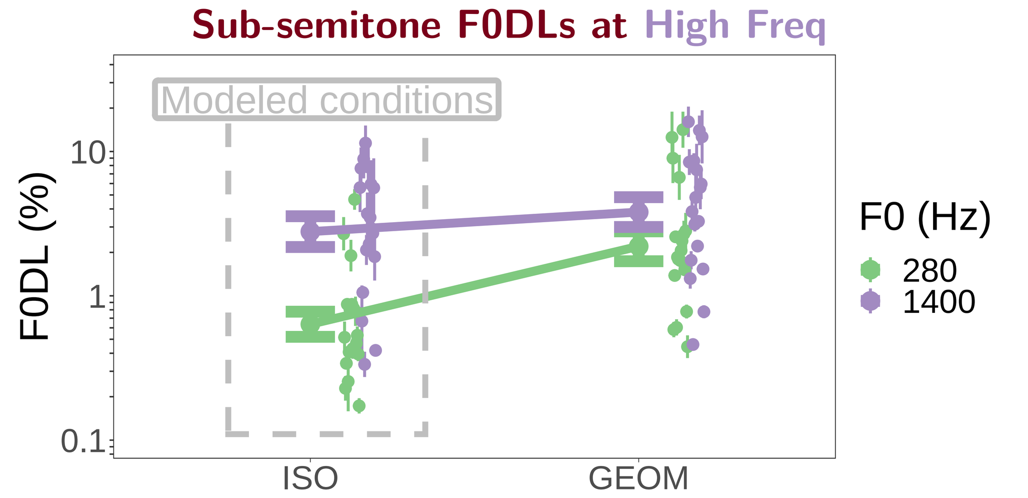


Figure 1: F0 difference limens (F0DLs) from 18 participants. Small points and error bars indicate individual means and ±1 SEM, while large points and error bars indicate group averages and ± 1 SEM.

Higher TMRs needed at High Freq in double complex-tone masker

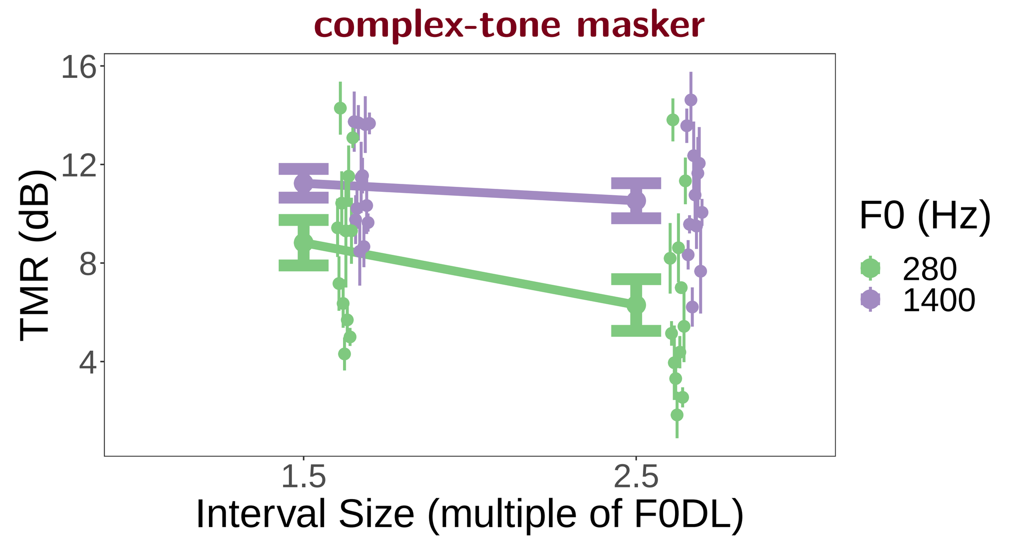


Figure 2: Target-to-masker ratio (TMR) required to achieve 79.4% correct F0 discrimination for the DBL masker with fixed interval sizes. Data are from 14 participants. Intervals between the reference and target tones were set at multiples of each listener's individual F0DL. Small points and error bars indicate individual means and ±1 SEM, while large points and error bars indicate group averages and ± 1 SEM.

Ideal observer

- Simulated firing rates for neurons with CFs between 5× and 11× F0 for ISO stimuli, analyzed with ideal observer [9]
- Two variants: **all-information** (temporal and average-rate information) and **rate-place** (only average-rate information)

$$X_i \sim \text{Poisson}(r_i(t, \theta)) \quad (1)$$

Spike times of i-th auditory nerve, X_i, are distributed Poisson according to time-varying rate r_i governed by parameter θ. Then, according to the Cramér-Rao lower bound...

$$\text{F0DL}_{\text{all-information}} = \left(\sum_i \int_0^T \frac{1}{r_i(t, F_0)} \left[\frac{\partial r_i(t, F_0)}{\partial F_0} \right]^2 dt \right)^{-1/2} \quad (2)$$

$$\text{F0DL}_{\text{rate-place}} = \left(\sum_i \int_0^T \frac{1}{\bar{r}_i(F_0)} \left[\frac{\partial \bar{r}_i(F_0)}{\partial F_0} \right]^2 dt \right)^{-1/2} \quad (3)$$

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., NIH F31 DC019247-01 awarded to D.R.G., and NSF NRT-UtB1734815
- A special thank you to Laurel Carney for suggestions and code for the excitatory-inhibitory neuron models

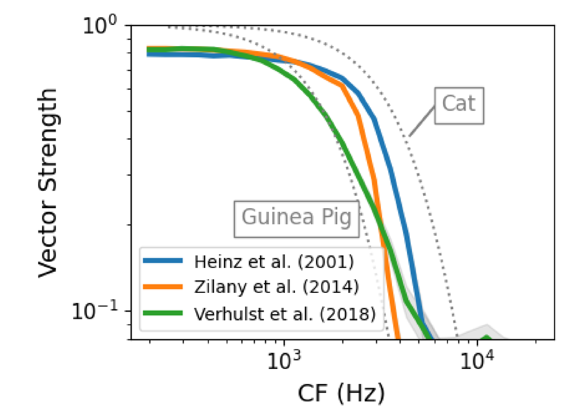
Bibliography

1. Oxenham, A. J., Micheyl, C., Keebler, M. V., Loper, A. & Santurette, S. *Proceedings of the National Academy of Sciences* **108,** 7629–7634 (2011).
2. Lau, B. K., Mehta, A. H. & Oxenham, A. J. *The Journal of Neuroscience* **37,** 9013–9021 (2017).
3. Gockel, H. E. & Carlyon, R. P. *Acta Acustica* **104,** 766–769 (2018).
4. Gockel, H. E., Moore, B. C. J. & Carlyon, R. P. *The Journal of the Acoustical Society of America* **148,** 3322–3333 (2020).
5. Micheyl, C., Keebler, M. V. & Oxenham, A. J. *The Journal of the Acoustical Society of America* **128,** 257–269 (2010).
6. Micheyl, C., Bernstein, J. G. W. & Oxenham, A. J. *The Journal of the Acoustical Society of America* **120,** 1493–1505 (2006).
7. Wang, J. *et al.* *The Journal of the Acoustical Society of America* **132,** 339–356 (2012).
8. Moore, B. C. J., Huss, M., Vickers, D. A., Glasberg, B. R. & Alcántra, J. I. *British Journal of Audiology* **34,** 205–224 (2000).
9. Heinz, M. G., Colburn, H. S. & Carney, L. H. *Neural Computation* **13,** 2273–2316 (2001).
10. Zilany, M. S. A., Bruce, I. C. & Carney, L. H. *The Journal of the Acoustical Society of America* **135,** 283–286 (2014).
11. Verhulst, S., Altoè, A. & Vasilikov, V. *Hearing Research* **360,** 55–75 (2018).
12. Weiss, T. F. & Rose, C. *Hearing Research* **33,** 175–180 (1988).
13. Oxenham, A. J. & Shera, C. A. *Journal of the Association for Research in Otolaryngology* **4,** 541–554 (2003).
14. Shera, C. A. & Guinan, J. J. *The Journal of the Acoustical Society of America* **113,** 2762–2772 (2003).
15. Krips, R. & Furst, M. *Neural Computation,* 2524–2553 (2009).
16. Saddler, M. R., Gonzalez, R. & McDermott, J. H. *bioRxiv* (2020).

Auditory nerve simulations

- Simulated high spontaneous rate fibers (approximately 40 CFs and 50 fibers per CF)
- **Filterbank models**
  - Heinz *et al.* [9]
  - Zilany *et al.* [10]
- **Transmission line models**
  - Verhulst *et al.* [11]

Vector strength



Filter tuning

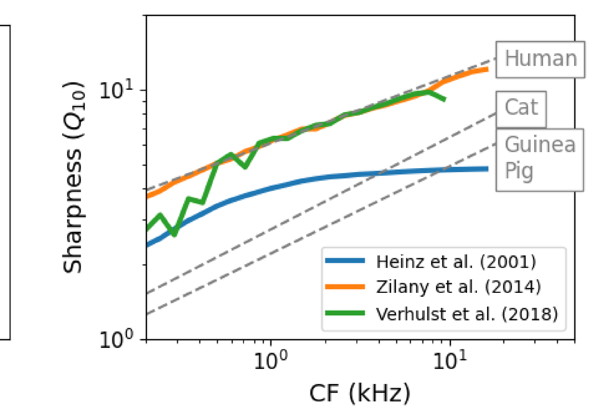


Figure 3: Vector strength (left) and Q_{10} (right) for each of the nerve models (colored lines). Animal data in the left side are model fits from Weiss & Rose [12]. Human and animal data in the right side are model fits from Oxenham & Shera [13] and Shera & Guinan [14], respectively.

Ideal observer thresholds (ISO conditions only)

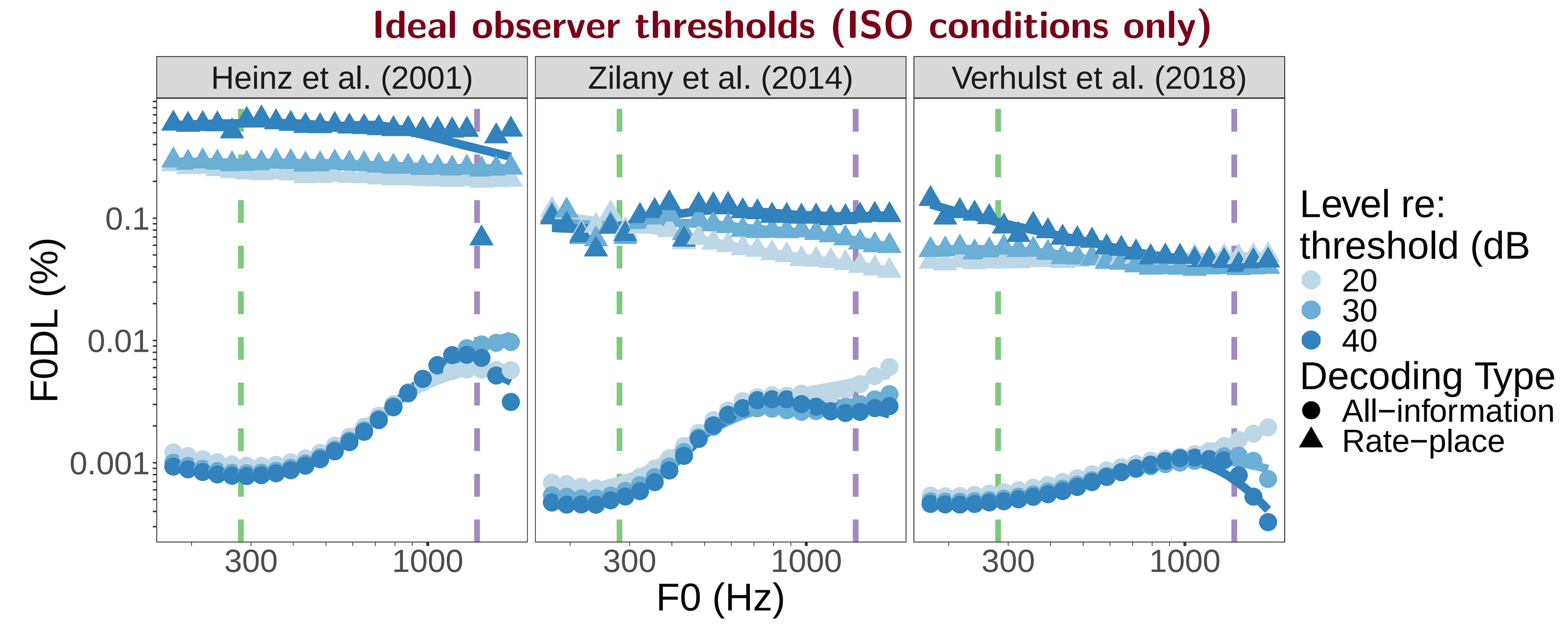


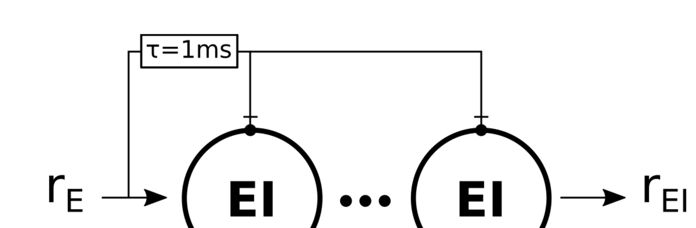
Figure 4: Ideal observer F0DLs for each of the tested auditory nerve models. Points show the F0DL estimates at each tested frequency while smooth lines show LOESS curves fit to the F0DL estimates. Vertical dashed lines indicate the nominal F0s tested in the behavioral tasks (280 Hz = green, 1400 Hz = purple).

Excitatory-inhibitory coincidence detector simulations

- Simulated excitatory-inhibitory coincidence detector neurons [15]
  - Excitatory input was firing rate from auditory nerve, inhibitory input was delayed copy of input at same CF
  - Output neuron spikes if excitatory input spikes and inhibitory input did not spike in the preceding Δ seconds

$$r_{EI}(t) = r_E(t) \left[1 - \int_{t-\Delta}^{t} r_E(\zeta - \tau) d\zeta \right]^M \quad (4)$$

Output rate is product of excitatory drive and inhibitory factor. Characteristics of inhibition are determined by M (number of inhibitory inputs), τ (delay), and Δ (integration time of inhibition)



Ideal observer thresholds (ISO conditions only)

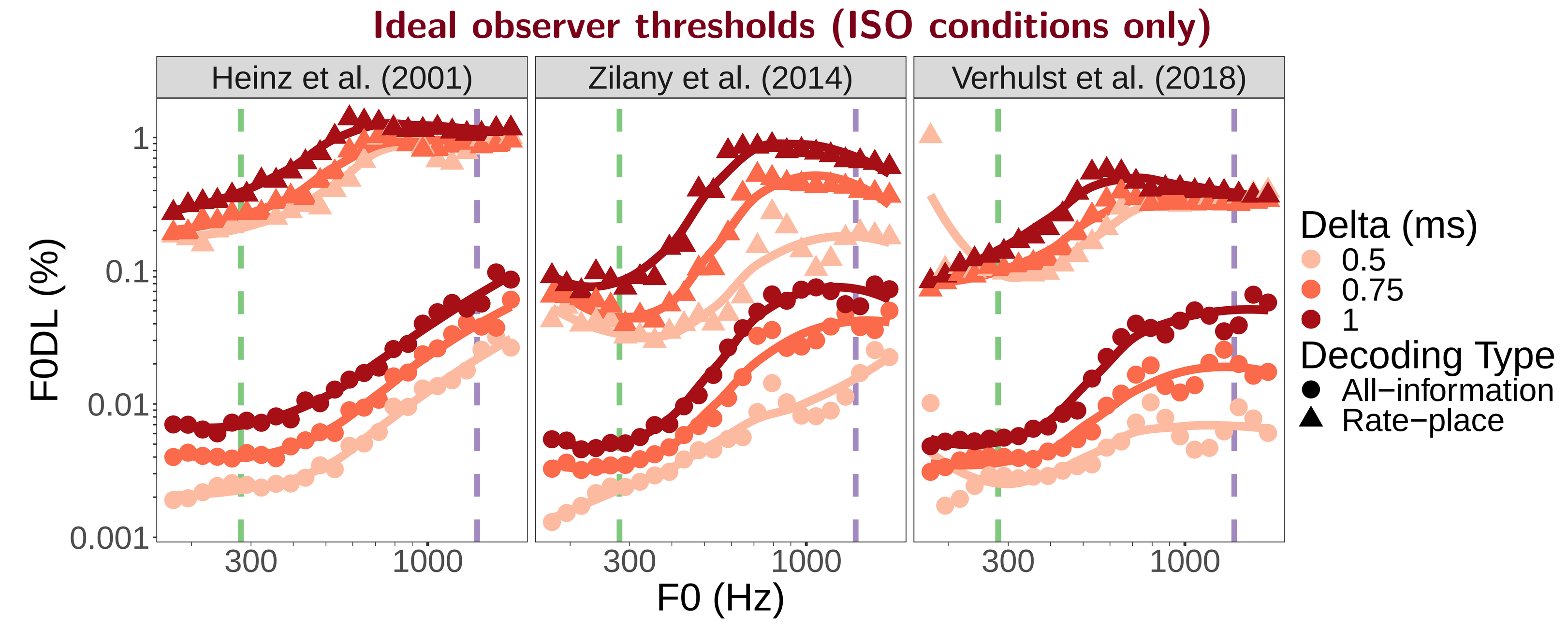


Figure 5: Ideal observer F0DLs for the auditory nerve models followed by an excitatory-inhibitory neuron model. Points show the F0DL estimates at each tested frequency while smooth lines show LOESS curves fit to the F0DL estimates. Vertical dashed lines indicate the nominal F0s tested in the behavioral tasks (280 Hz = green, 1400 Hz = purple). All simulations were at 30 dB re: threshold and used $\tau = 1$ ms and $M = 25$.

Model-behavior comparisons and conclusions

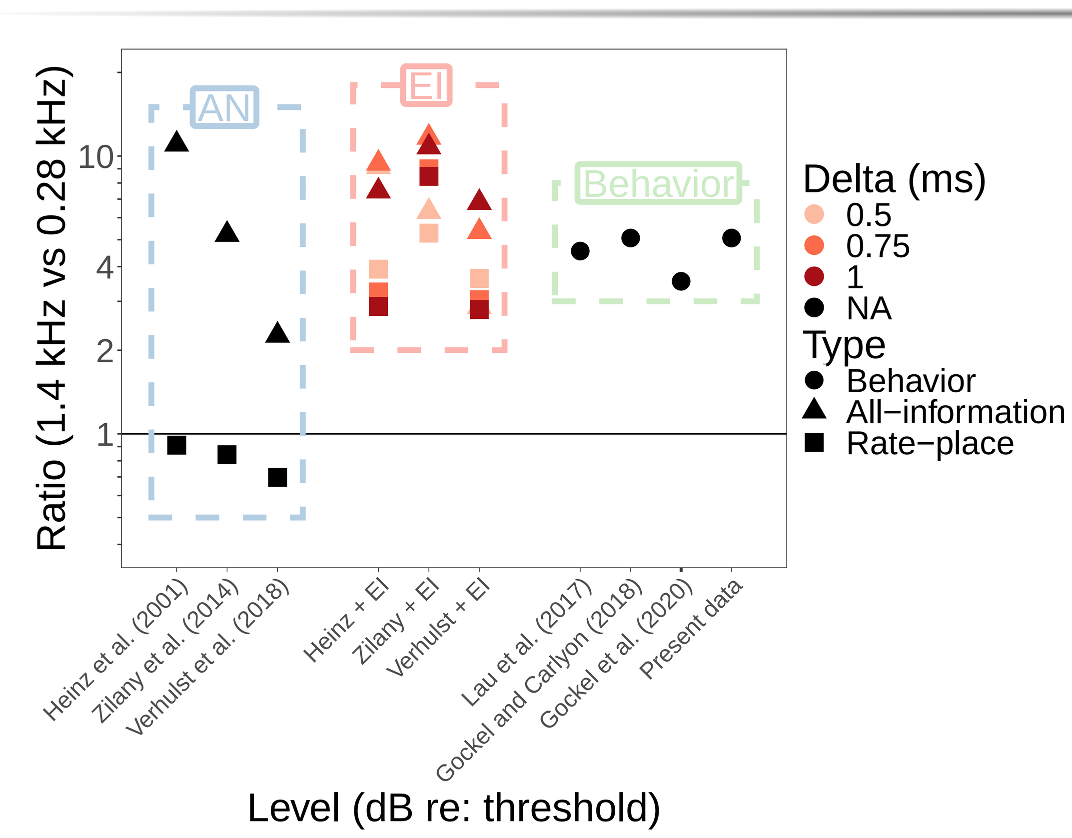


Figure 6: Ratios between F0DLs for F0s of 1.4 kHz and 0.28 kHz for the auditory nerve models (blue outline, left), auditory nerve models followed by an excitatory-inhibitory neuron layer (pink outline, middle), and behavior (green outline, right). The simulations were at 30 dB re: threshold, and the excitatory-inhibitory neurons used $\tau = 1$ ms, $M = 25$, and a range of inhibitory time constants (Δ; color). Behavioral data are from Lau *et al.* [2], Gockel & Carlyon [3], and Gockel *et al.* [4].

Conclusions

- **Behavior —** Accurate F0DLs were achieved at **High Freq**, but multiple concurrent complex-tone maskers impaired F0 discrimination more at **High Freq** than at **Low Freq**
- **Computational model —** Rolloff of F0DLs with increasing frequency could reflect a degrading temporal code in the auditory nerve and/or a transformation to a rate code at a later stage of processing
- **Future directions**
  - Explore whether auditory nerve or EI neuron simulations can account for impact of complex-tone maskers on F0 discrimination
  - Explore other modeling frameworks to relate neural simulations to behavior (e.g., deep neural networks; [16])
- Download me at `https://guestdaniel.github.io/download/GuestOxenhamARO2021.pdf`