Active Learning for Behavior

Dennis L. Barbour, MD, PhD

Department of Biomedical Engineering
Washington University in St. Louis

October 22, 2018

BONAURIA

Disclosure

Dennis Barbour has an ownership interest in Bonauria, LLC, and may financially benefit if the company is successful in commercializing products that are related to this research.

E

F P

T O Z

L P E D

P E C F D

E D F C Z P

F E L O P Z D

D E F P O T E C

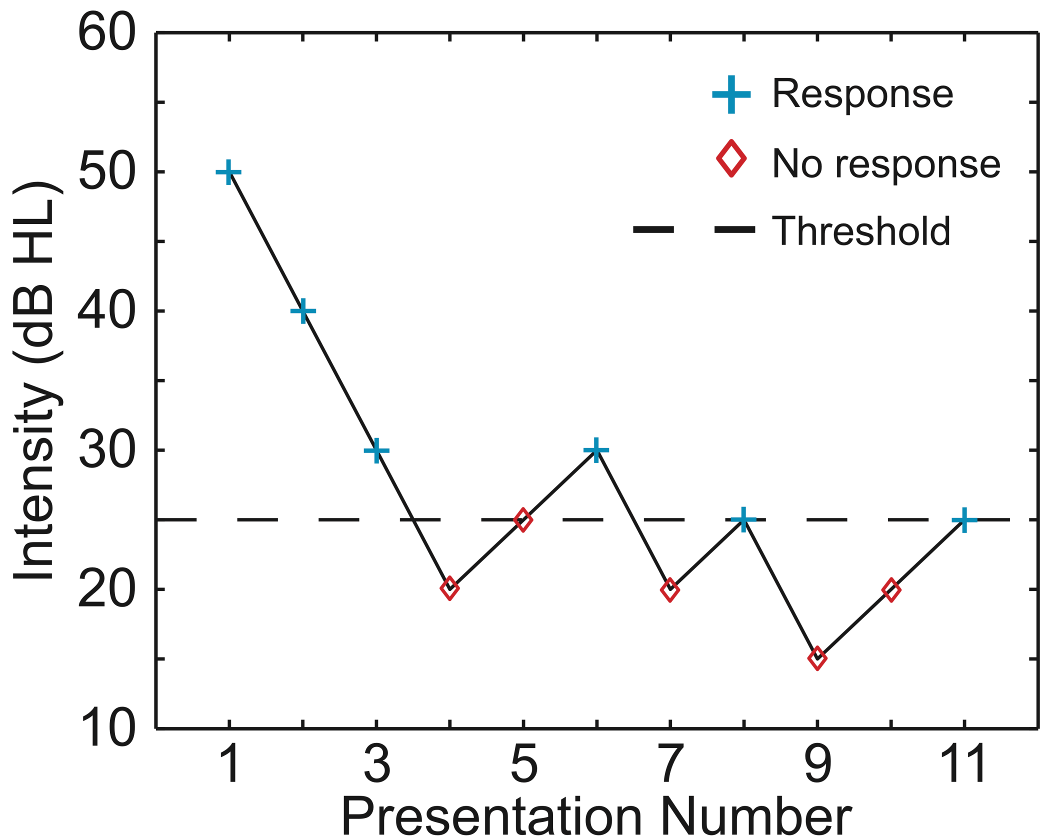
L E F O D P C T

F D P L T C E O

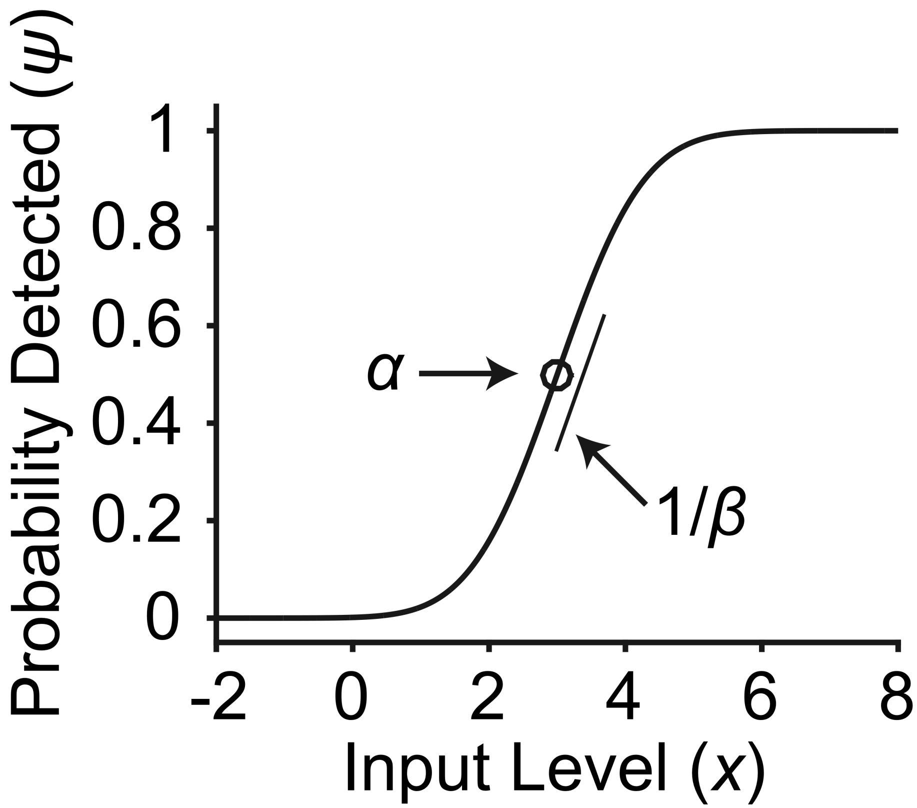
P E Z O L C F T D

D L Z V P O C J F L Z C P O F C D

Up-down procedures estimate thresholds only



Psychometric functions are traditionally parameterized by threshold and spread

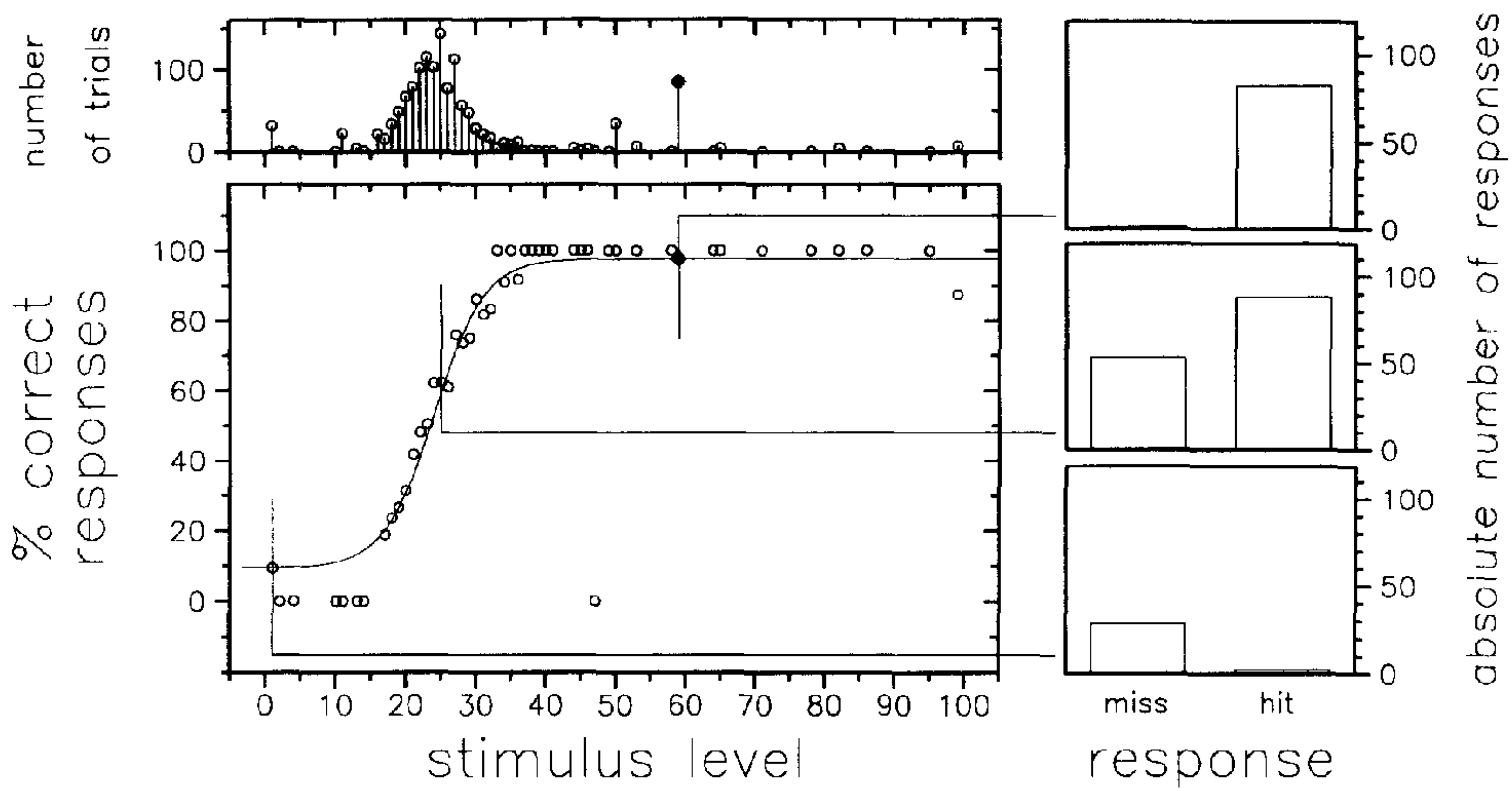


$$\psi\left(x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{\lambda-\alpha}{\beta}\right)^2} d\lambda$$

$$\alpha = \text{threshold}$$

$$\beta = \text{spread}$$

Classical psychophysics often involves estimating a Bernoulli probability distribution as a function of an independent variable



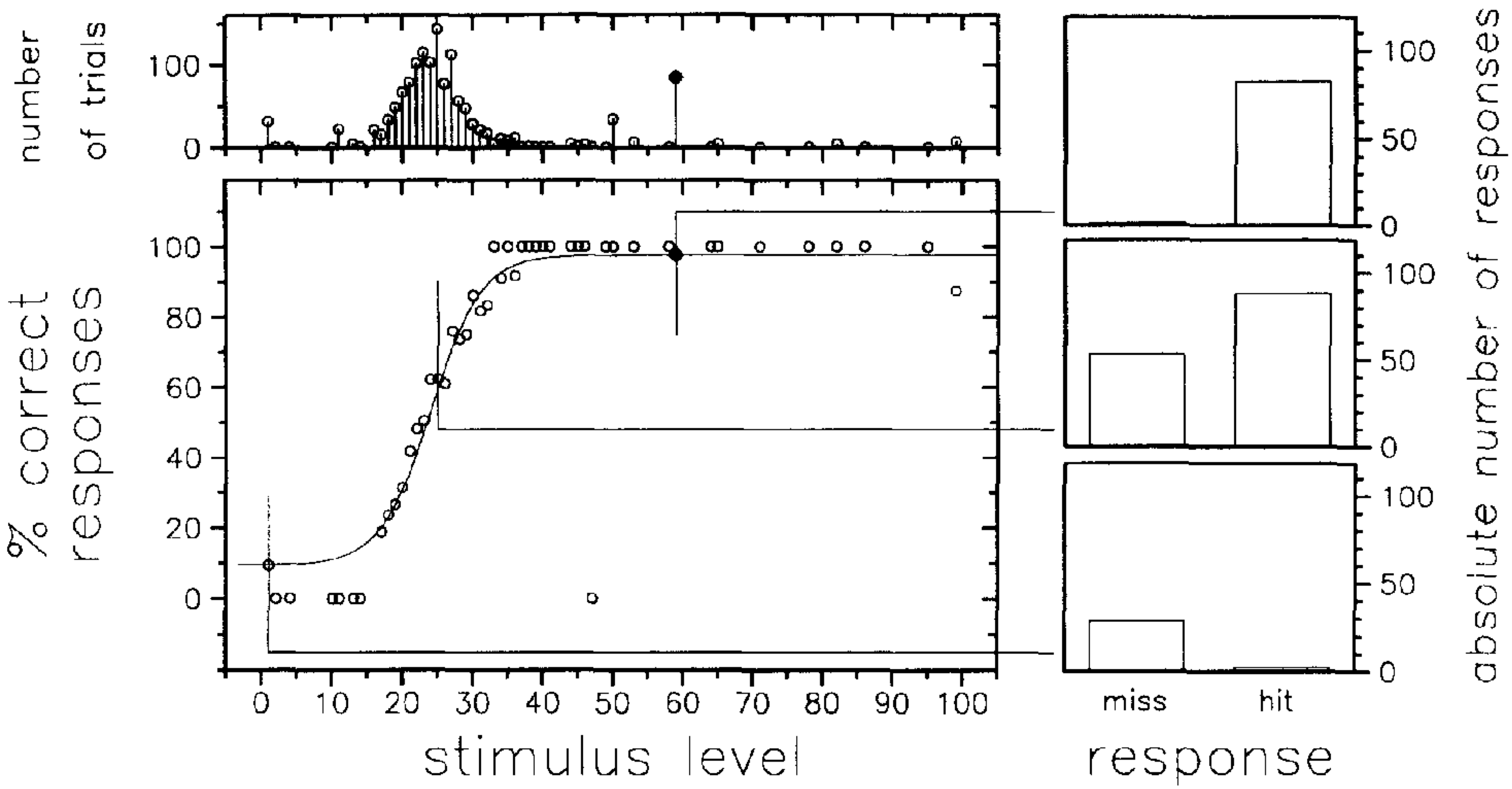
Treutwein, *Vision Res*, 35(17): 2503-22, 1995

Classical psychophysics involves estimating a Bernoulli probability distribution as a function of an independent variable

An example of a psychometric function with results from a forced-choice experiment with nine spatial alternatives is given in Fig. 1. Here, the percentage correct assignments of the stimulus location has been plotted against the stimulus level, which in this case was the duration of a temporal break in one of nine simultaneously displayed stimuli. The plotted results are cumulative data of 35 sessions, i.e. repetitions of the experiment with the same stimulus setup. '

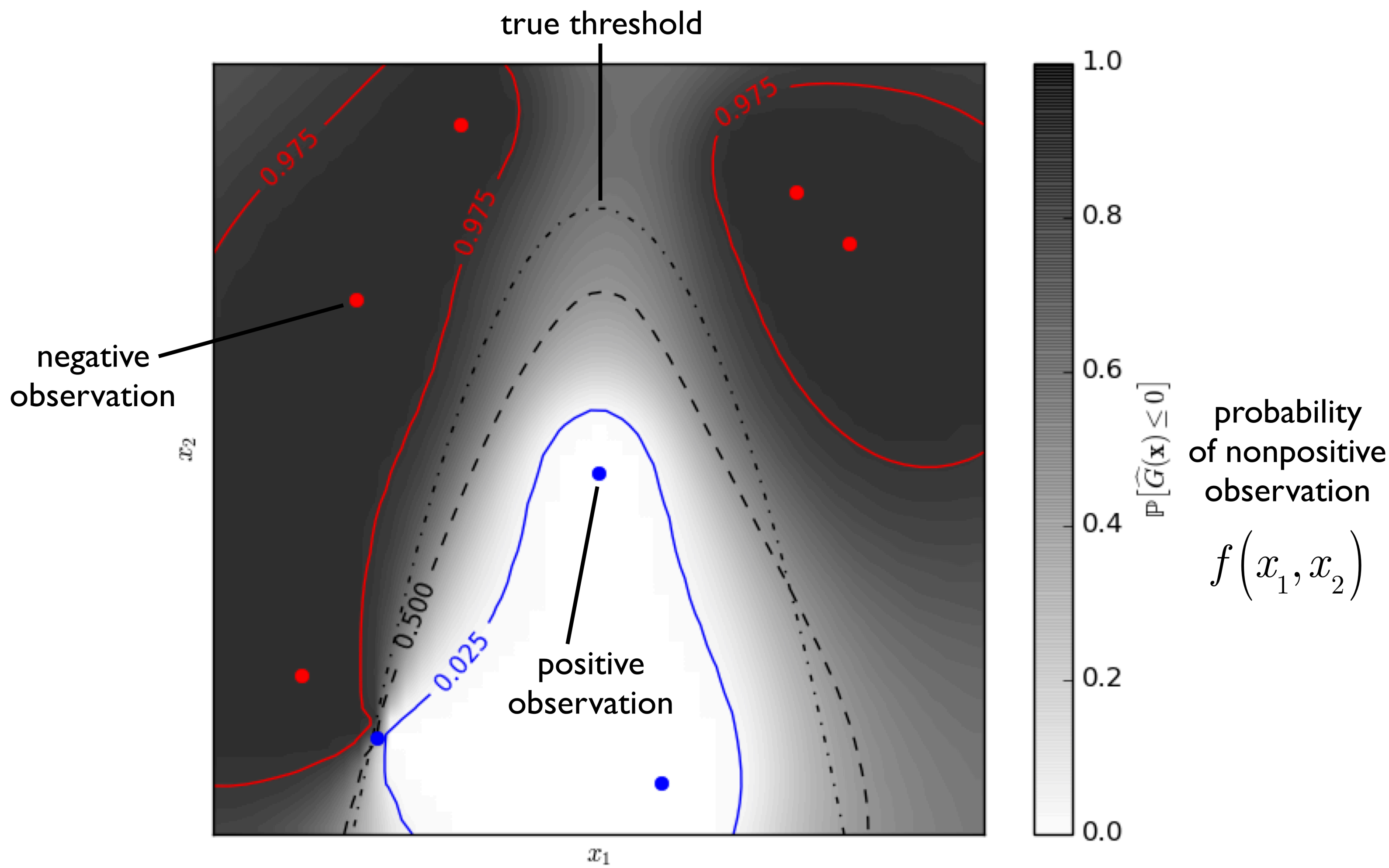
Treutwein, *Vision Res*, 35(17): 2503-22, 1995

Classical psychophysics often involves estimating a Bernoulli probability distribution as a function of an independent variable



Treutwein, *Vision Res*, 35(17): 2503-22, 1995

Probabilistic classification generates continuous estimates of class boundaries



scikit-learn.org

Bayes' theorem provides a key inferential framework

likelihood   prior over *f*

posterior

$$p\left(f \middle| x, y\right) = \frac{p\left(y \middle| f\right) p\left(f \middle| x\right)}{p\left(y \middle| x\right)}$$

prior over *y*

$$= \frac{p\left(y \middle| f\right) p\left(f \middle| x\right)}{\int p\left(y \middle| f\right) p\left(f \middle| x\right) df}$$

Bayes' theorem:
the posterior distribution (the probability of model given the observations) equals
the prior distribution (the probability of the model) times
the likelihood (the probability of the observations given the model) normalized by
the marginal likelihood (the probability of the observations or the model evidence)

Gaussian processes represent a powerful implementation of Bayesian inference

regression example

$$y(x) = f(x) + \varepsilon(x)$$

$$f(x) \sim \mathcal{GP}\left(\mu(x), K(x, x')\right)$$

$$\varepsilon(x) = \varepsilon \sim \mathcal{N}\left(0, \sigma^2\right)$$

classification example

$$y(x) \sim \mathrm{Bernoulli}\left(\Phi\left(f(x)\right)\right)$$

$$f(x) \sim \mathcal{GP}\left(\mu(x), K(x, x')\right)$$

$$\Phi(f) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{f} e^{-\frac{z^2}{2}}\, dz$$

mean function example: constant

$$\mu(x) = c$$

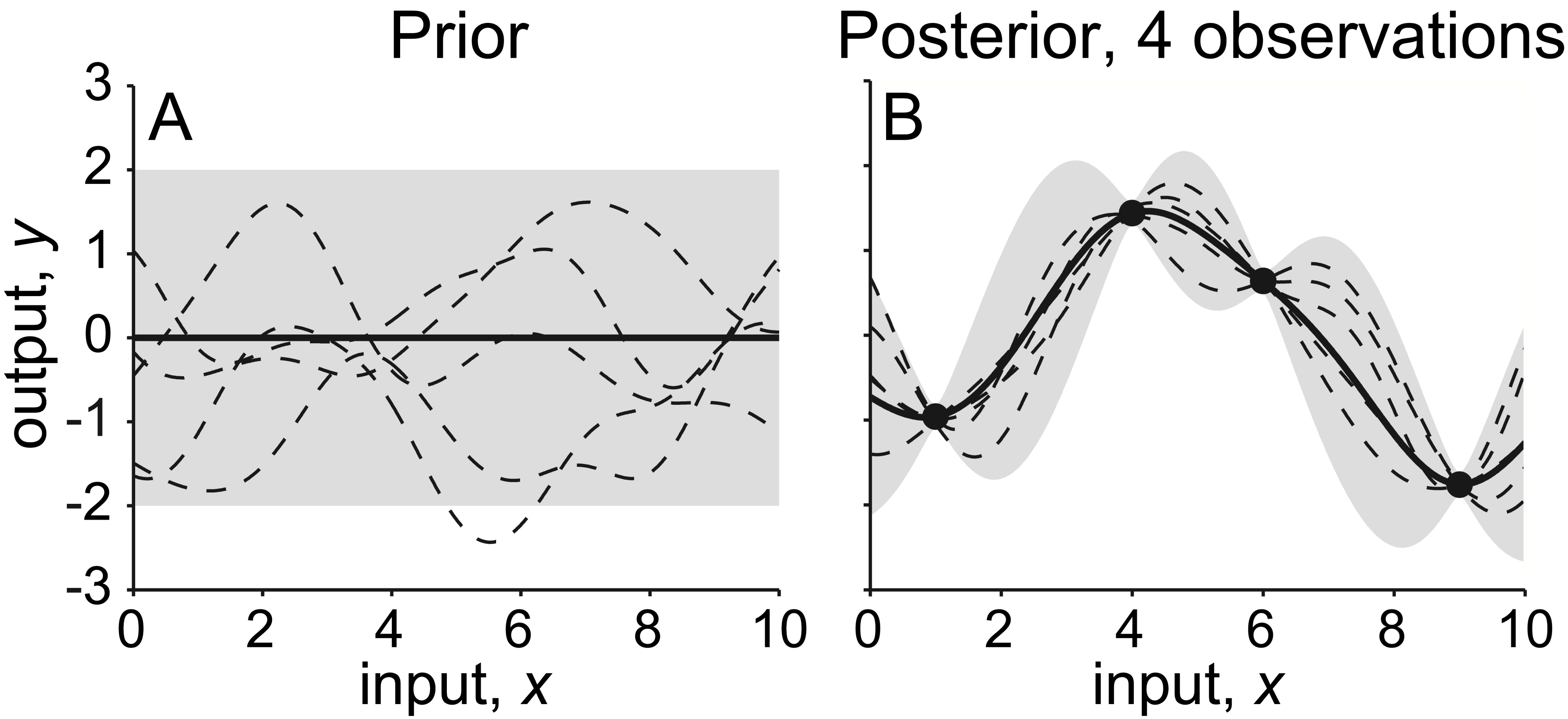
kernel example: linear

$$K(x, x') = s_1^2 (x - x')$$

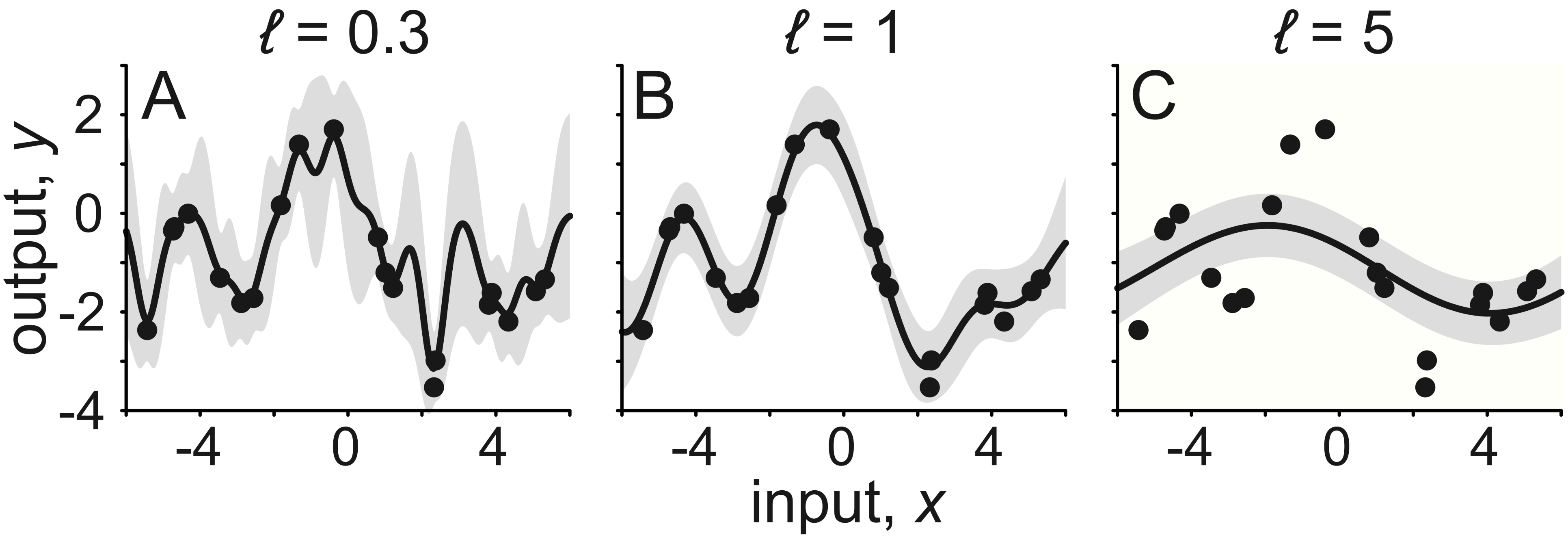
kernel example: squared exponential

$$K(x, x') = s_2^2 \exp\left(\frac{-(x - x')^2}{2\ell^2}\right)$$

Prior beliefs + observations lead to posterior beliefs



Constraints on the GP covariance function determine details of the posterior belief



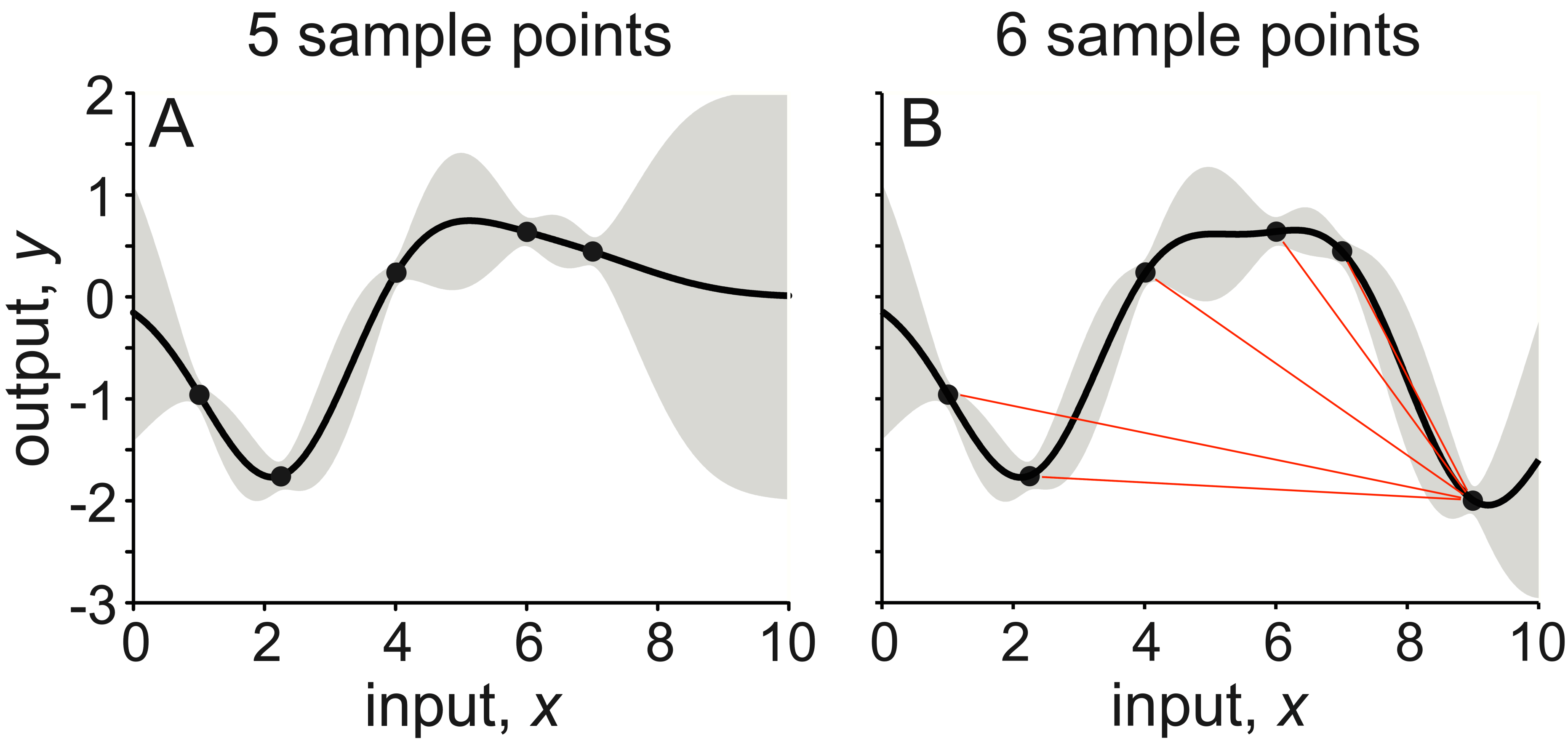
Gaussian process estimates enable active learning

$$\arg\max_{\mathbf{x}^*} H\left[\Theta \mid D\right] - E_{y^* \sim p\left(y^* \mid \mathbf{x}^*, D\right)} \left[H\left[\Theta \mid y^*, \mathbf{x}^*, D\right]\right]$$

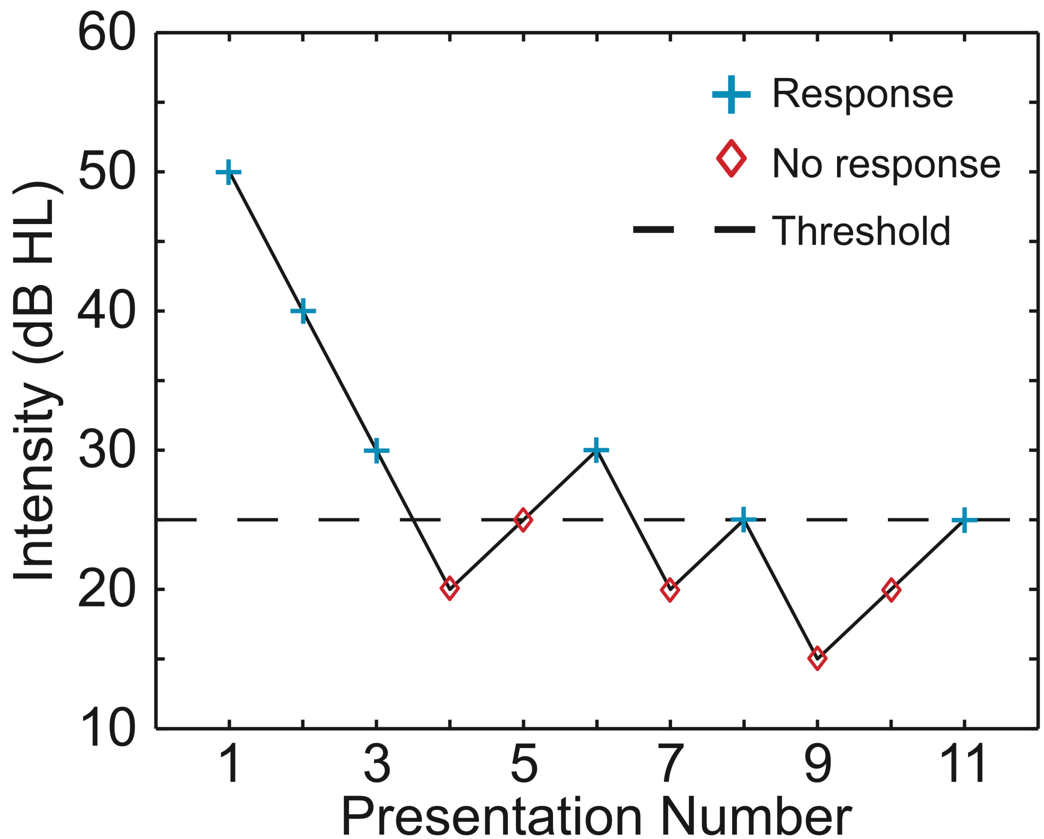
hyperparameters	Θ
test data	$\left(\mathbf{x}^*, y^*\right)$
training data	$D = \left\{\mathbf{x}_i, y_i\right\}_{i=1}^{n} = \left\{\mathbf{X}, \mathbf{y}\right\}$
entropy	$H\left[\Theta \mid D\right]$

$$\arg\max_{\mathbf{x}^*} H\left[y^* \mid \mathbf{x}^*, D\right] - E_{\Theta \sim p(\Theta \mid D)} \left[H\left[y^* \mid \mathbf{x}^*, \Theta\right]\right]$$

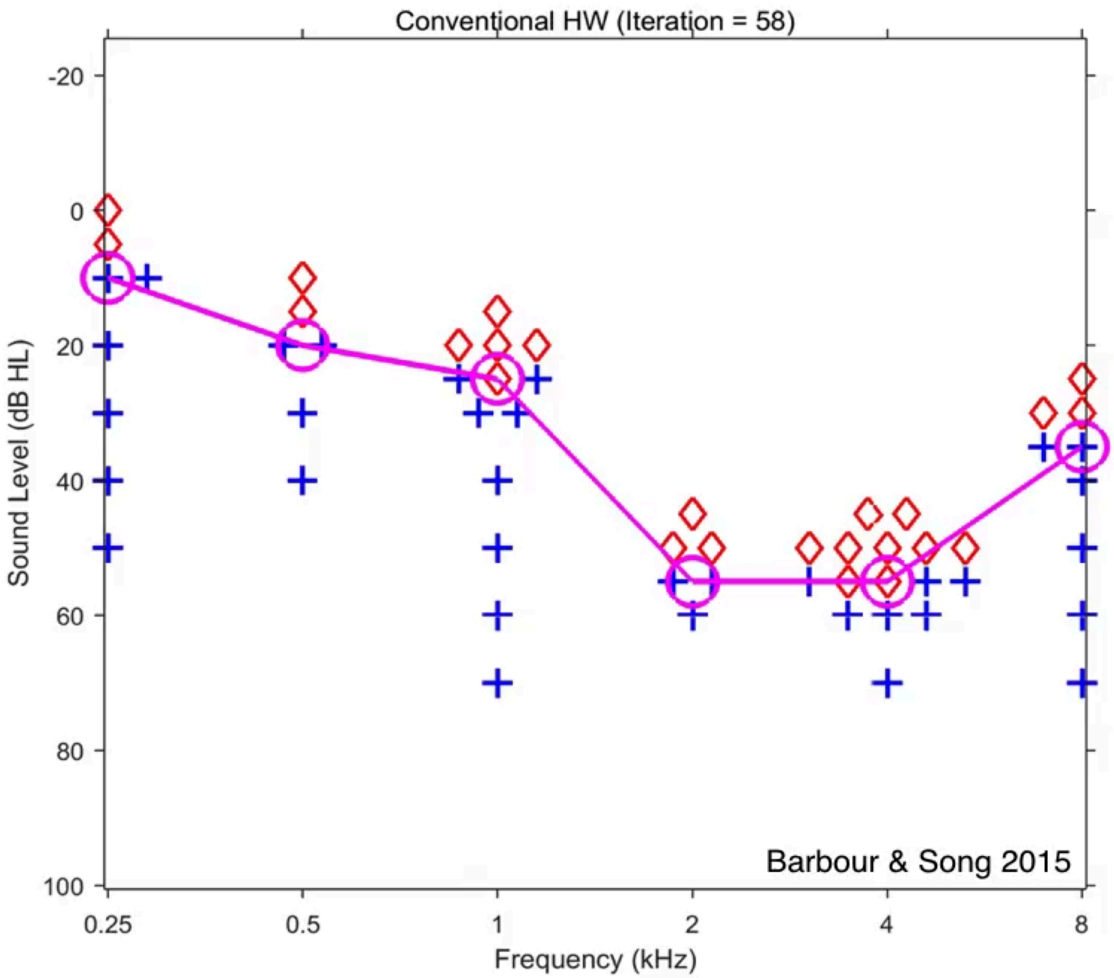
Additional observations can be selected where they would be most valuable



Up-down procedures estimate thresholds only



Hughson-Westlake audiometry (HWA) estimates tone detection thresholds only



The probabilistic classifier in this case is a Gaussian process

input data vectors $\mathbf{x} = \left(L, \omega\right), \mathbf{x}' = \left(L', \omega'\right)$

output observation $y\left(\mathbf{x}\right) \sim \mathrm{Bernoulli}\left(\Phi\left(f\left(\mathbf{x}\right)\right)\right)$

sigmoidal link function $\Phi\left(f\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{f} \exp\left(-\frac{z^2}{2}\right) dz$

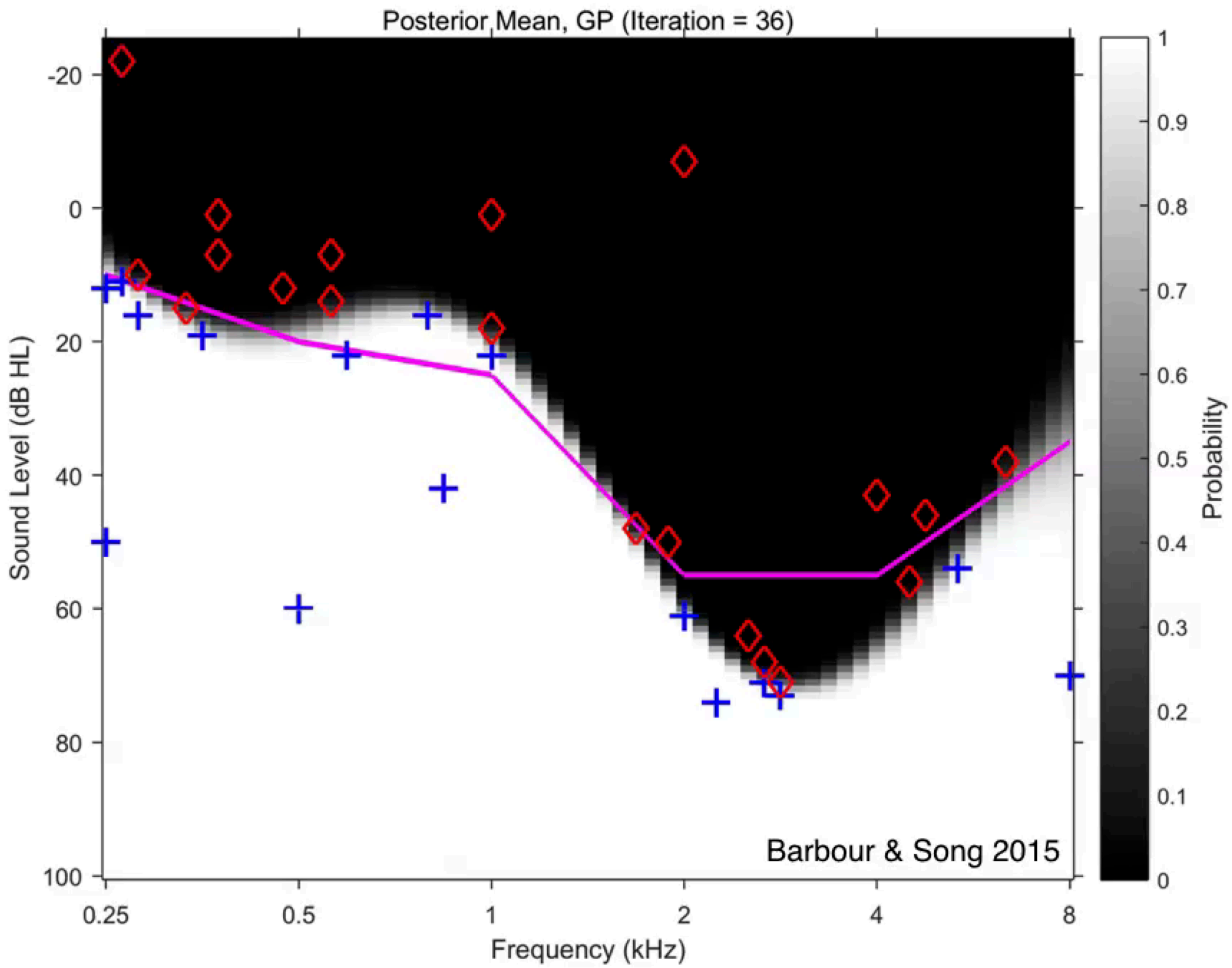
latent function $f\left(\mathbf{x}\right) \sim \mathcal{GP}\left(\mu\left(\mathbf{x}\right), K\left(\mathbf{x}, \mathbf{x}'\right)\right)$

sound level kernel $K_{L}\left(\mathbf{x}, \mathbf{x}'\right) = K_{L}\left(L, L'\right) = s_{L}^{2} L L'$

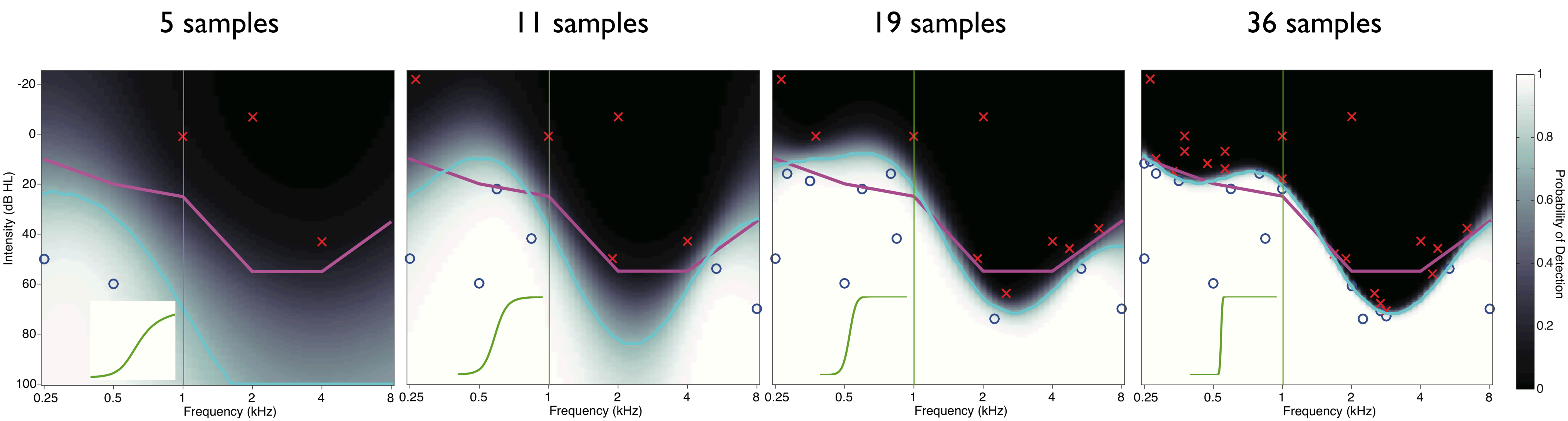
frequency kernel $K_{\omega}\left(\mathbf{x}, \mathbf{x}'\right) = K_{\omega}\left(\omega, \omega'\right) = s_{\omega}^{2} \exp\left(-\frac{\left(\omega - \omega'\right)^2}{2\ell^2}\right)$

complete kernel $K\left(\mathbf{x}, \mathbf{x}'\right) = K_{L} + K_{\omega}$

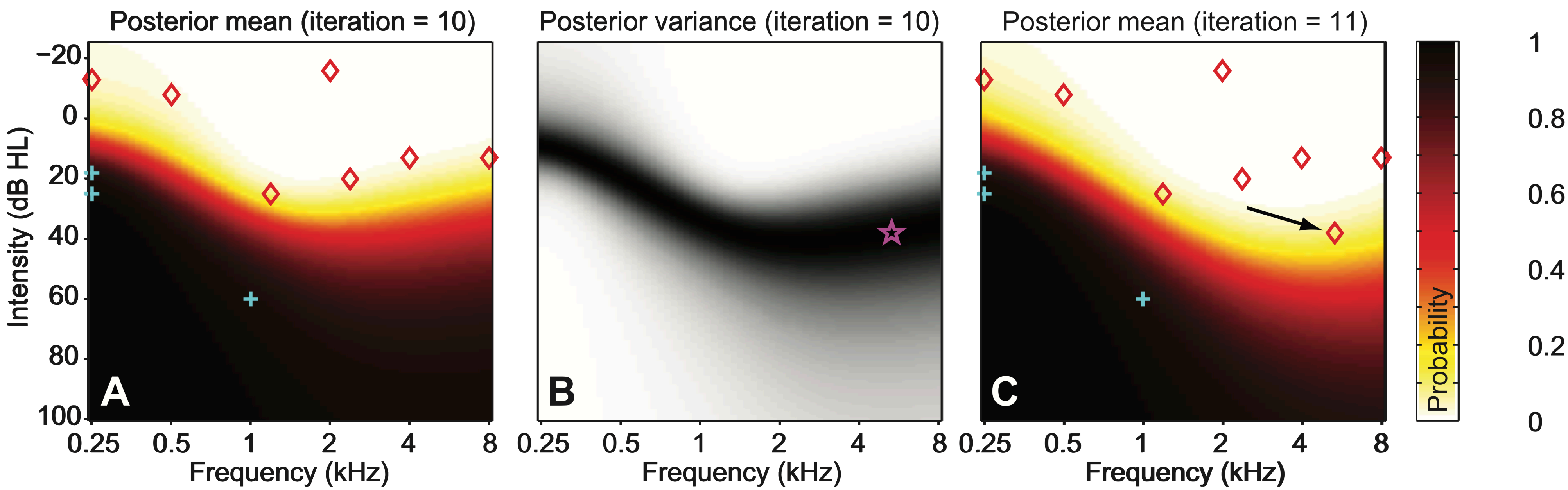
Gaussian process audiometry (GPA) estimates the complete audiometric function



GPA delivers both threshold and slope estimates across frequency

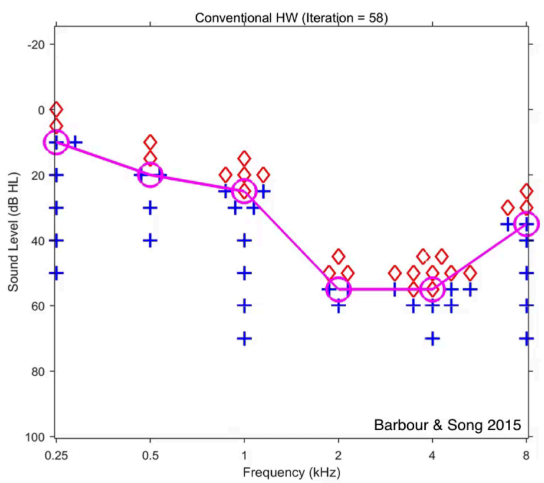


Bayesian active learning ensures that samples are acquired where needed most

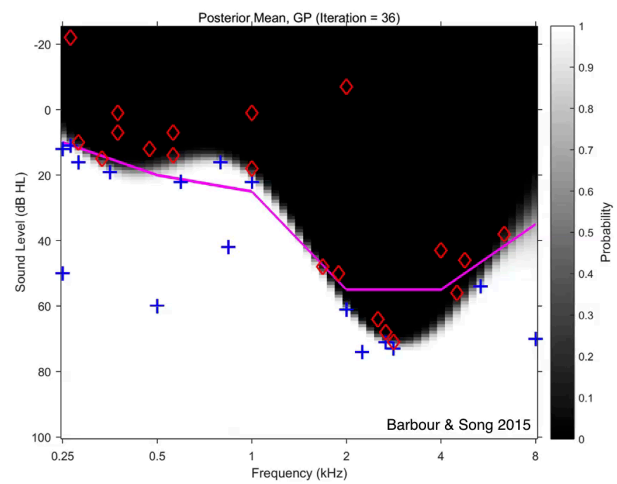


Song, et al, *Ear Hear*, 36(6):e326-335, 2015

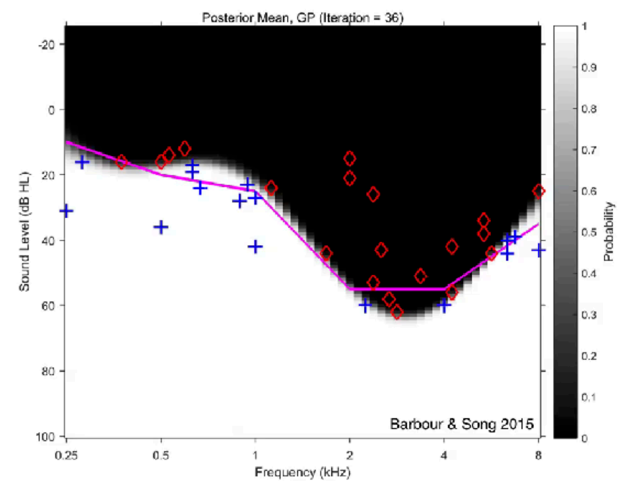
GPA is robust to stimulus selection



HWA

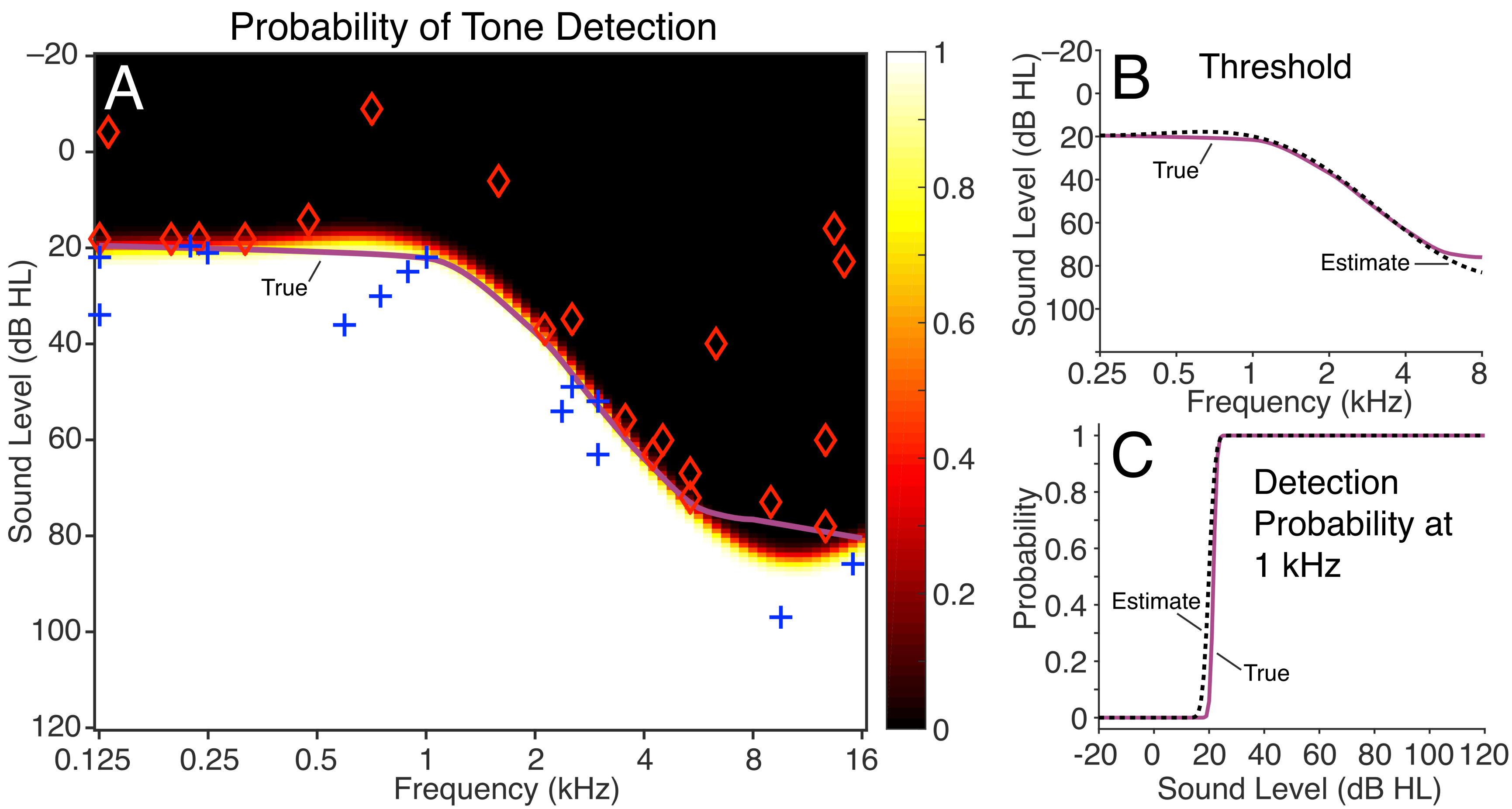


GPA 1

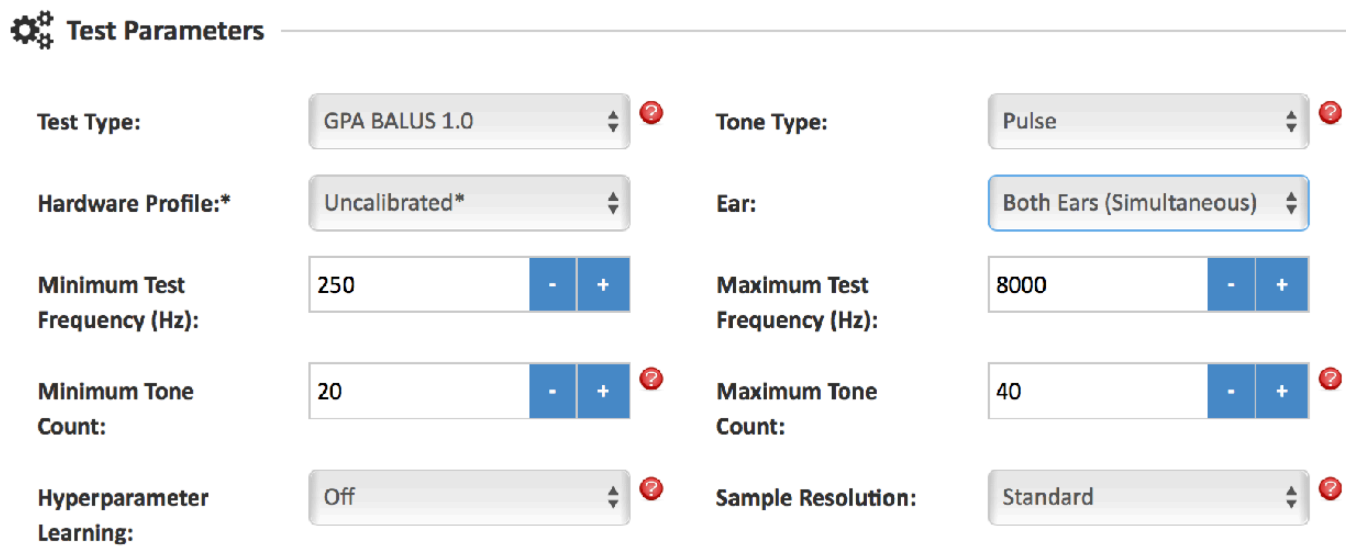


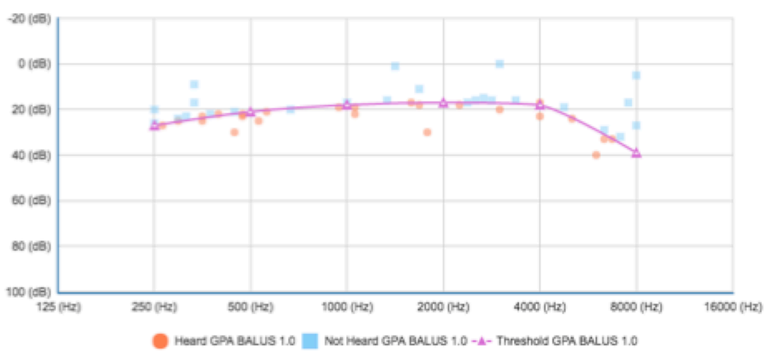
GPA 2

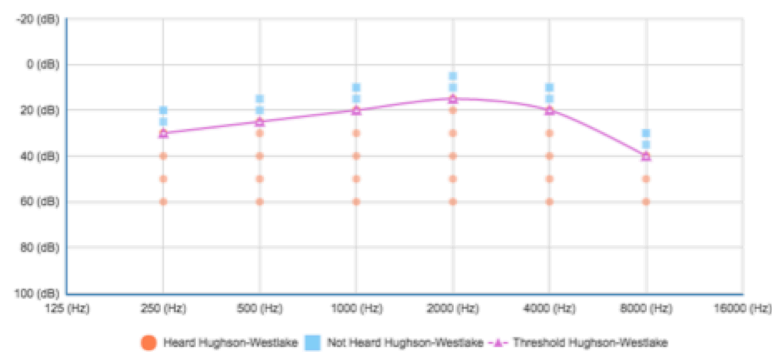
Gaussian process classification with Bayesian active learning yields accurate multidimensional psychometric threshold and slope estimates with few samples



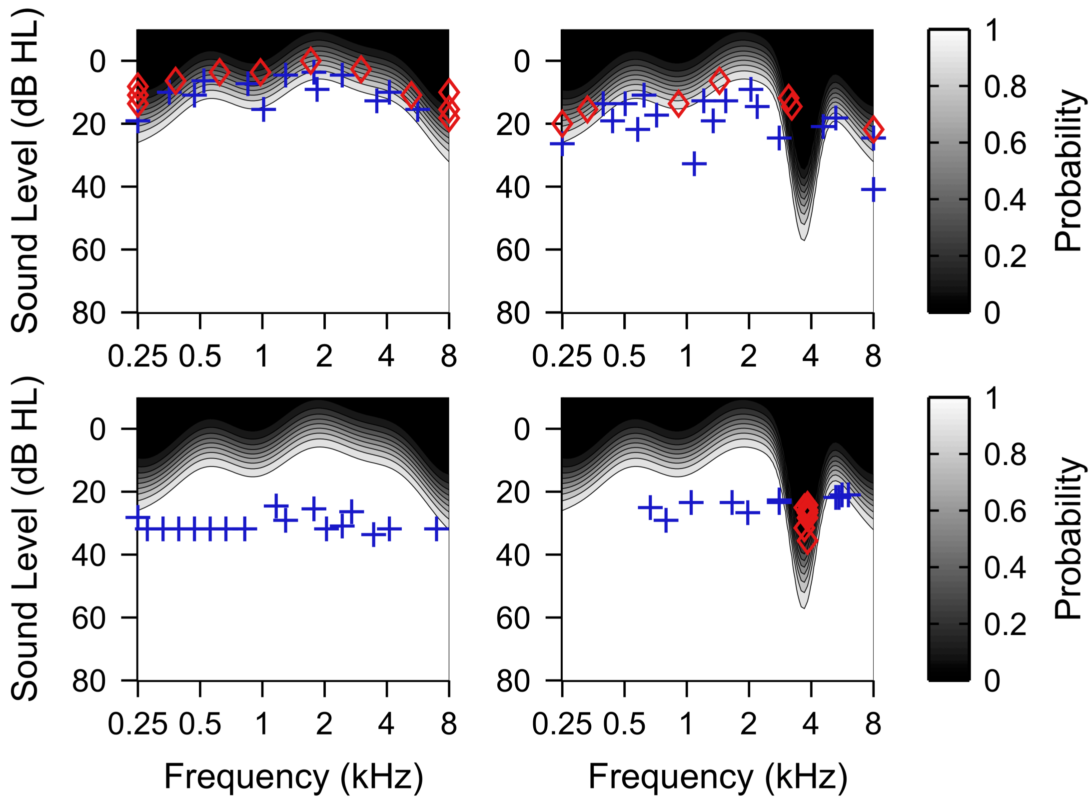
Online GPA can be compared directly to online HWA







Stimulus selection can be targeted toward real-time hypothesis testing



Bayesian active model selection

Gardner, et al, *NIPS*, 2377-2385, 2015

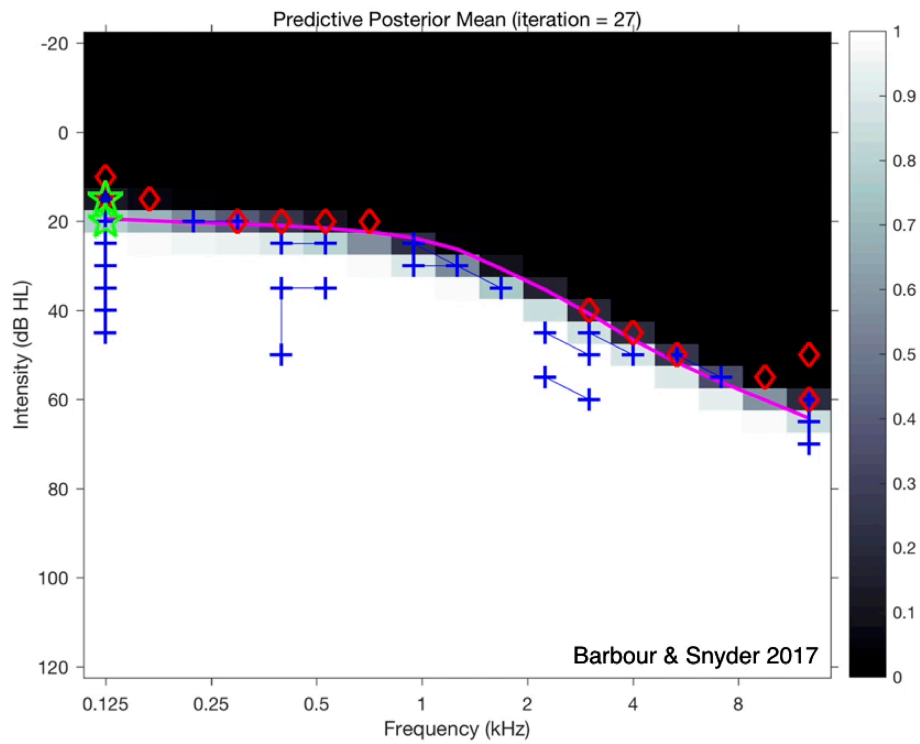
A test of human inference

Which answer, "yes" or "no," provides **more** information in response to the following question:

"Have you been anxious or depressed in the past 2 weeks?"

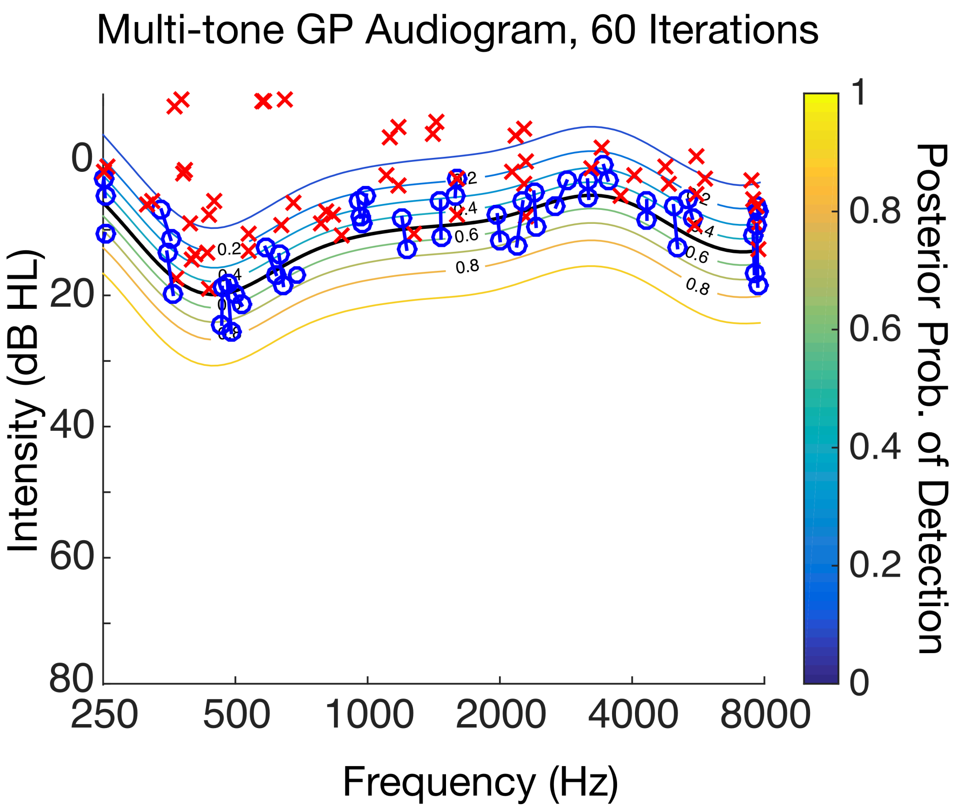
A. Yes
B. No

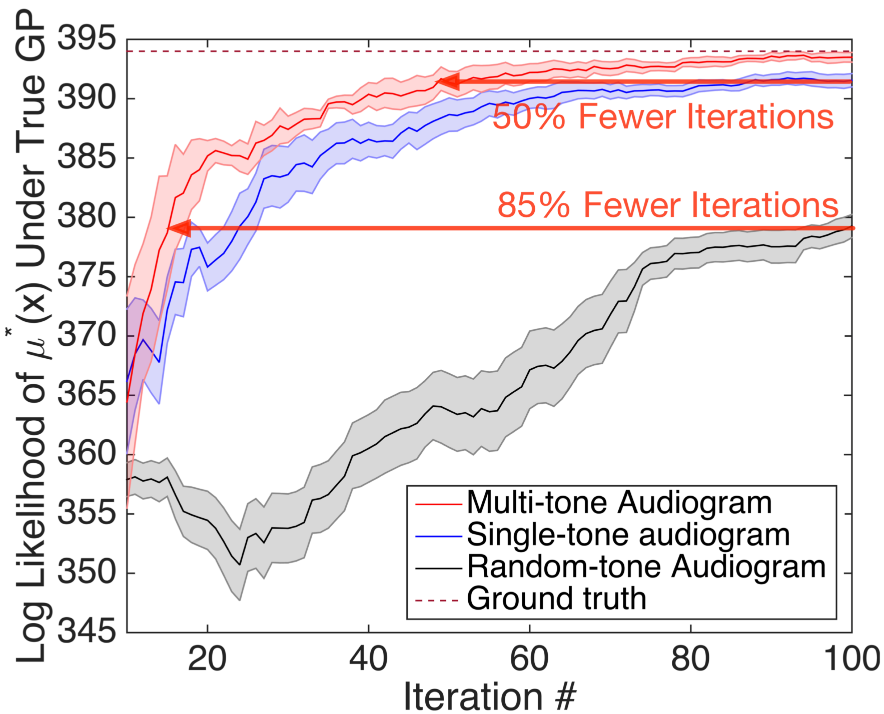
Multitone audiometry accelerates active learning



multiplexed psychometric estimation

Multitone audiometry accelerates active learning

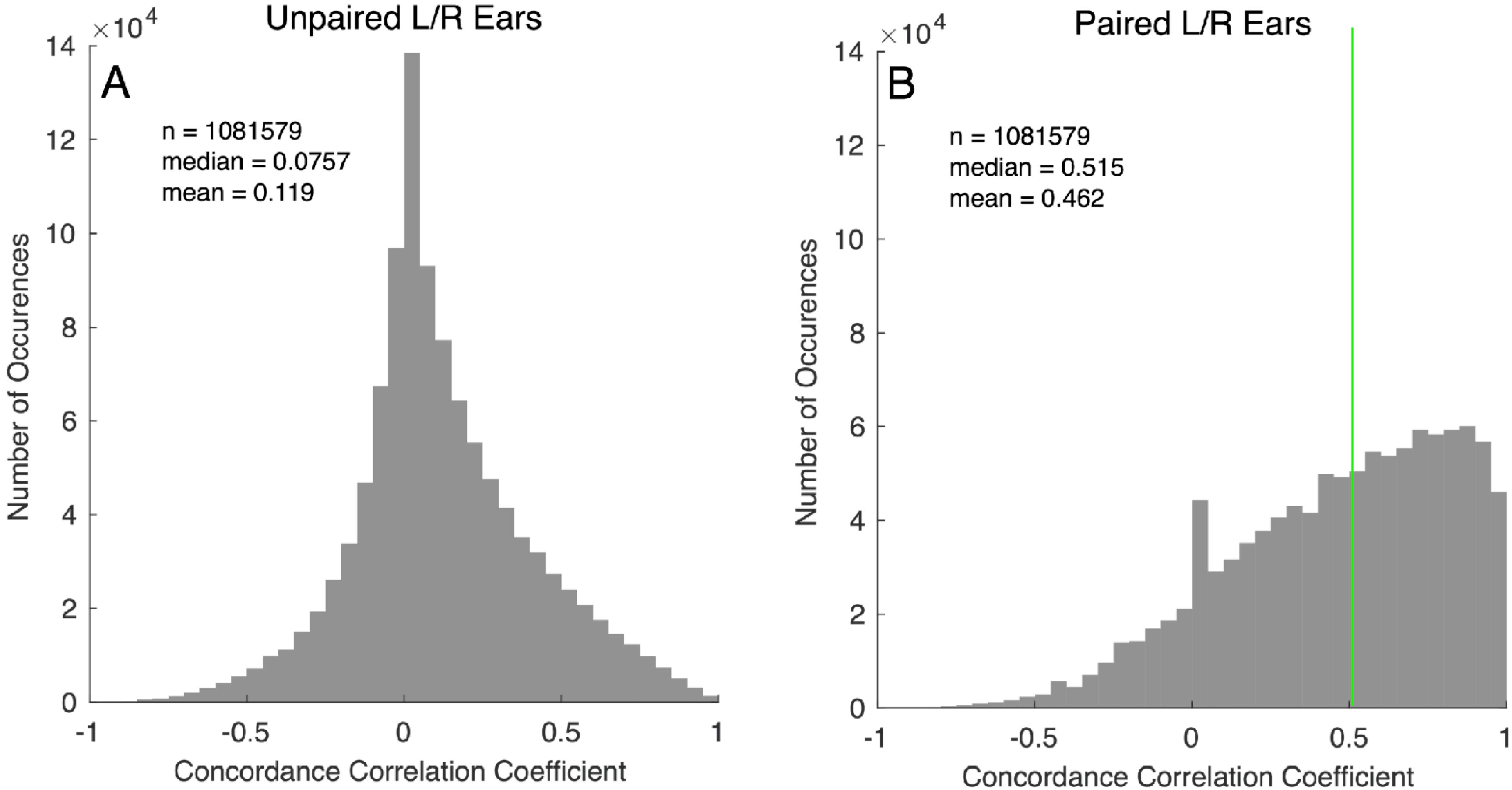




multiplexed psychometric estimation

Gardner, et al, *UAI*, 286-295, 2015

Hearing thresholds of a person's two ears are concordant



Barbour, DiLorenzo, *et al. Behav Res Meth* (in press).

Conjoint estimation is a natural extension of disjoint estimation with a GP

$$\mathbf{x} = \left(L, \omega, e\right), \mathbf{x}' = \left(L', \omega', e'\right)$$

$$K_e\left(\mathbf{x}, \mathbf{x}'\right) = \begin{cases} s_{11}^2 \text{ if } e = e' = 1 \\ s_{12}^2 \text{ if } e \neq e' \\ s_{22}^2 \text{ if } e = e' = 2 \end{cases}$$

$$K_{\text{conjoint}}\left(\mathbf{x}, \mathbf{x}'\right) = K_e\left(K_L + K_\omega\right)$$

$$\mathbf{x} = \left(L, \omega\right), \mathbf{x}' = \left(L', \omega'\right)$$

$$y\left(\mathbf{x}\right) \sim \text{Bernoulli}\left(\Phi\left(f\left(\mathbf{x}\right)\right)\right)$$

$$\Phi\left(f\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{f} \exp\left(-\tfrac{z^2}{2}\right) dz$$

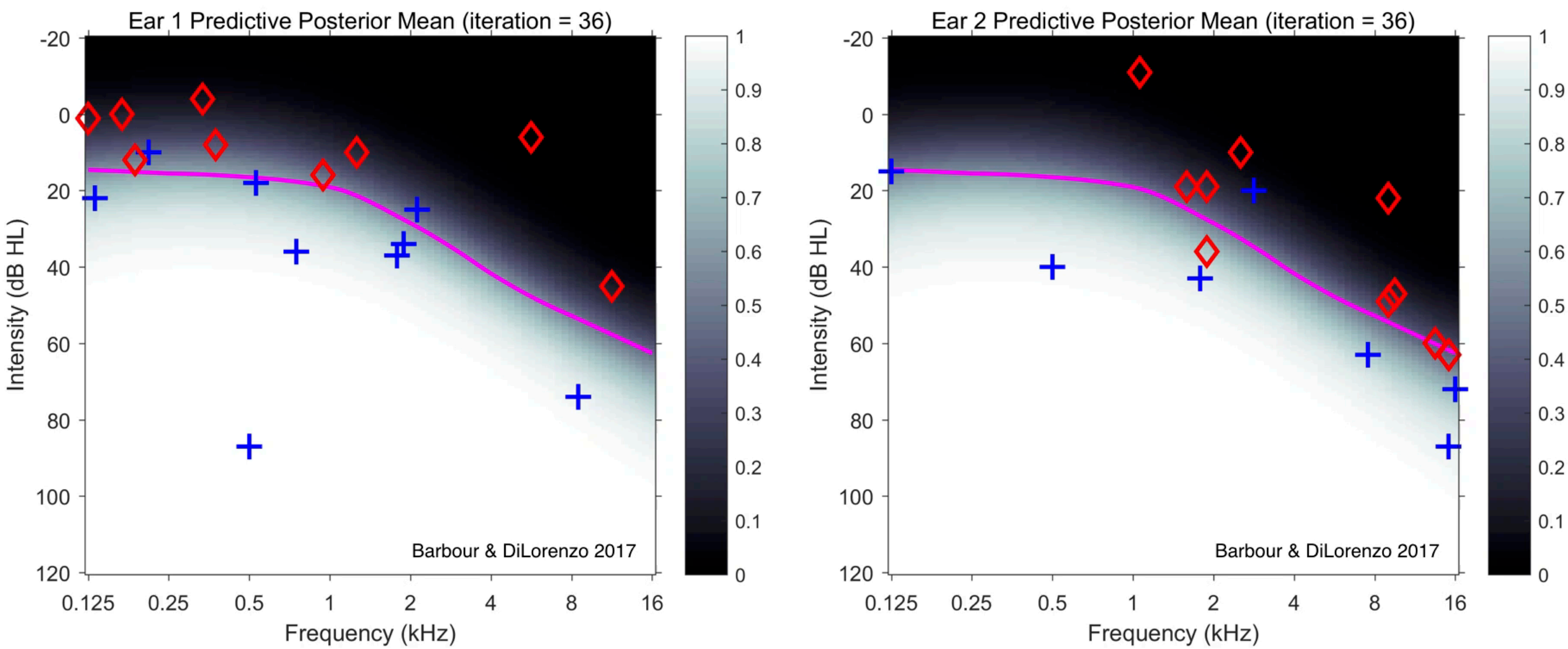
$$f\left(\mathbf{x}\right) \sim \mathcal{GP}\left(\mu\left(\mathbf{x}\right), K\left(\mathbf{x}, \mathbf{x}'\right)\right)$$

$$K_L\left(\mathbf{x}, \mathbf{x}'\right) = K_L\left(L, L'\right) = s_L^2 L L'$$

$$K_\omega\left(\mathbf{x}, \mathbf{x}'\right) = K_\omega\left(\omega, \omega'\right) = s_\omega^2 \exp\left(-\frac{\left(\omega - \omega'\right)^2}{2\ell^2}\right)$$

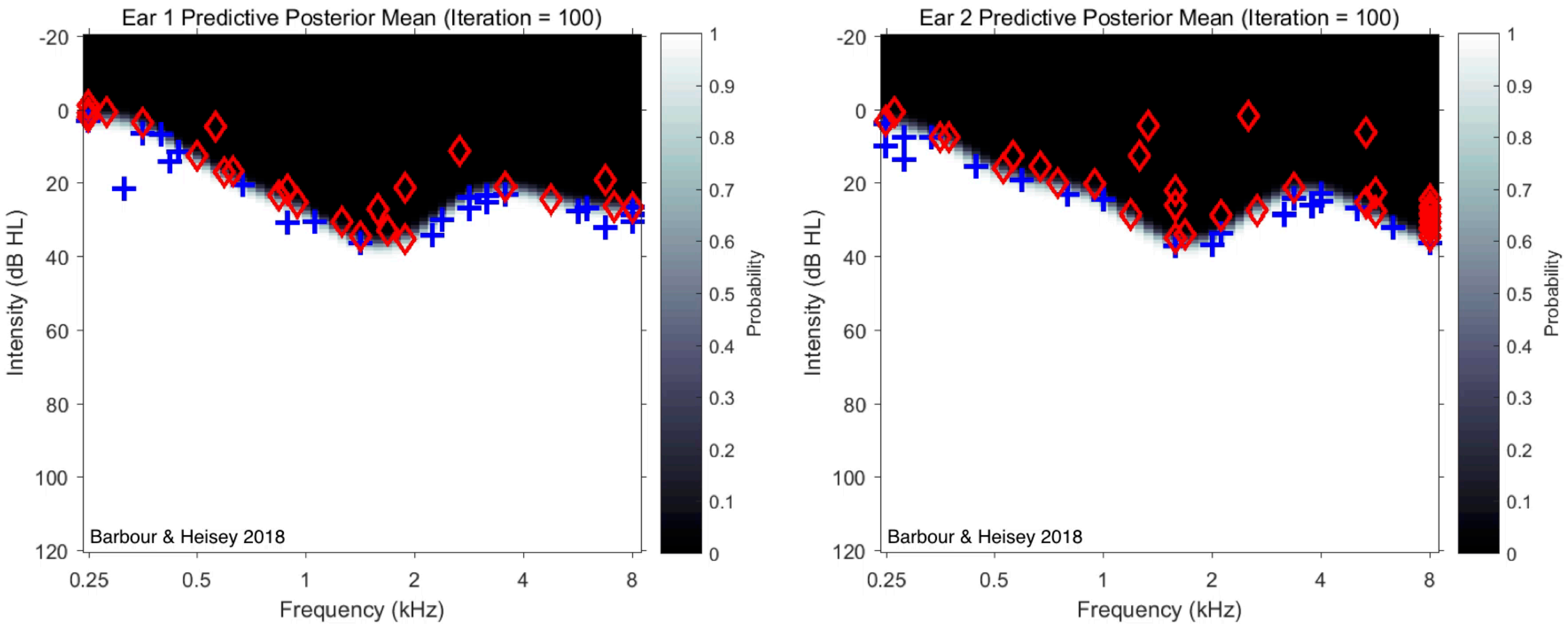
$$K_{\text{disjoint}}\left(\mathbf{x}, \mathbf{x}'\right) = K_L + K_\omega$$

Natural correlations can be exploited to augment estimator power

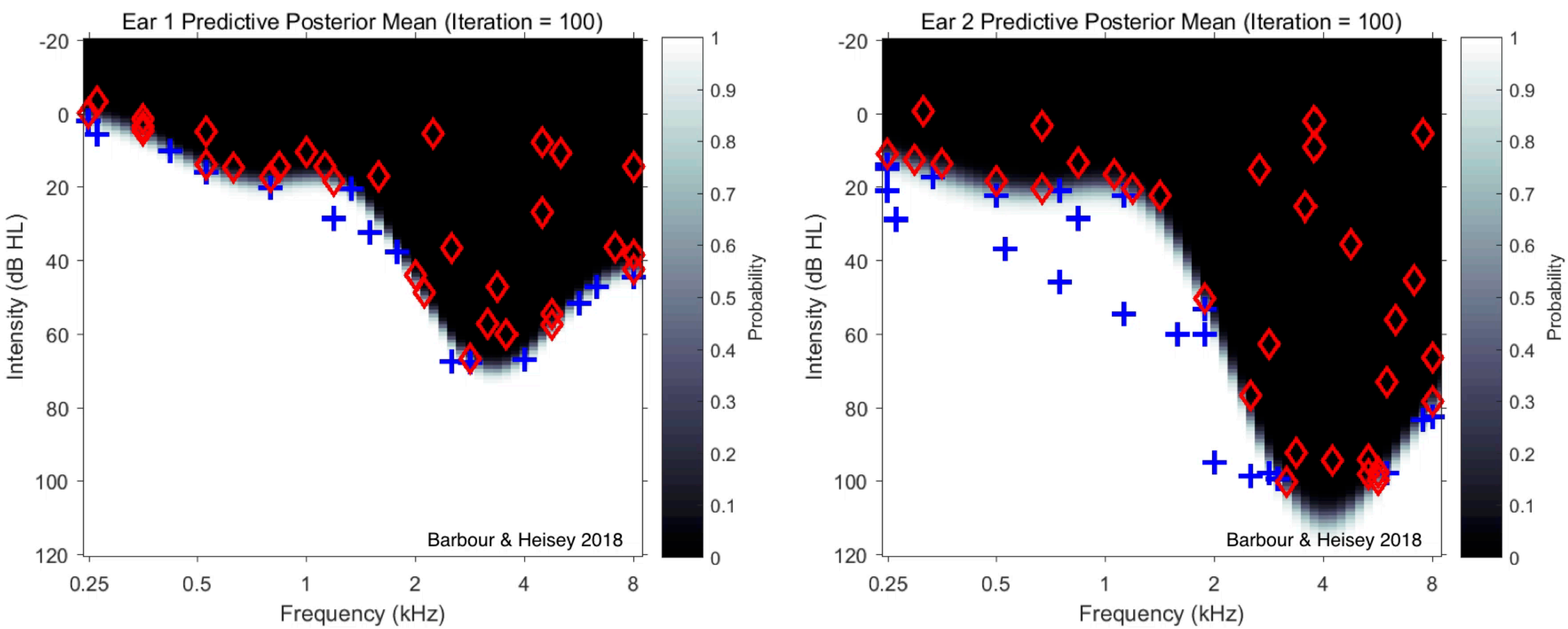


conjoint psychometric estimation

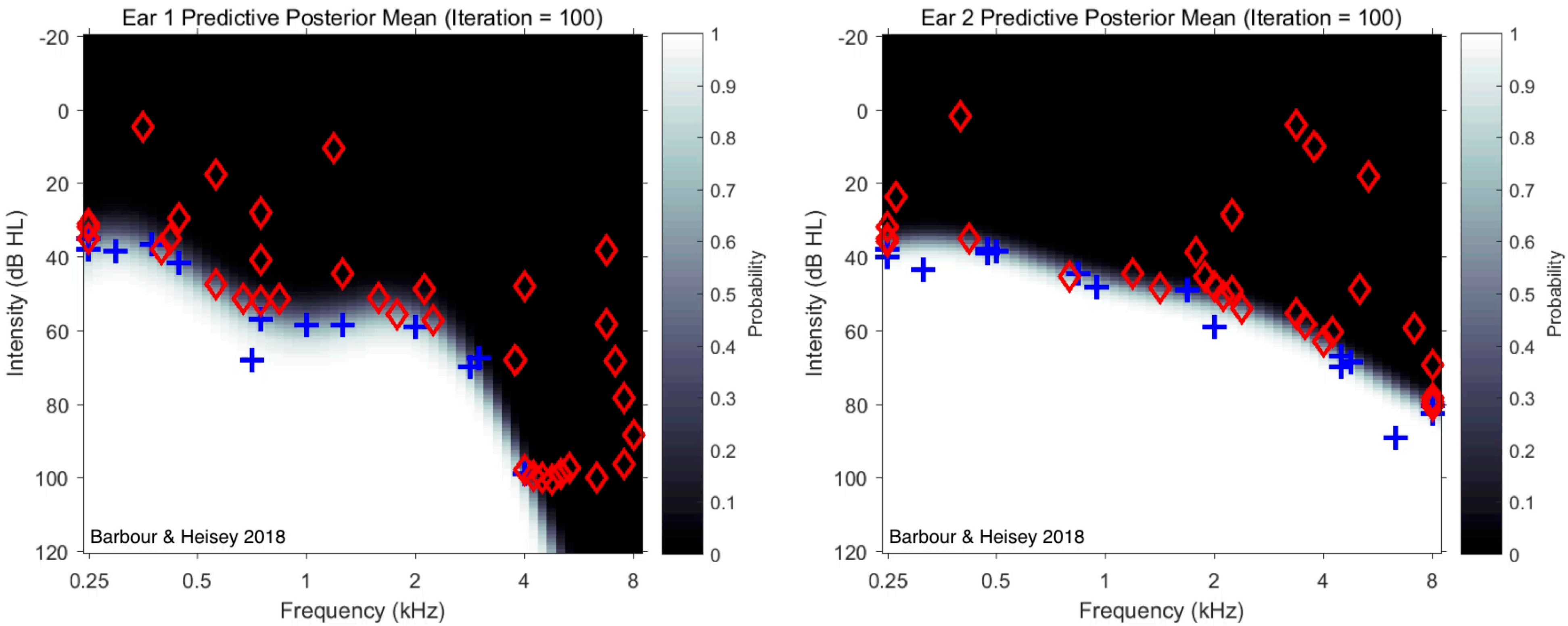
Active mutual conjoint estimation enables rapid assessment of both ears



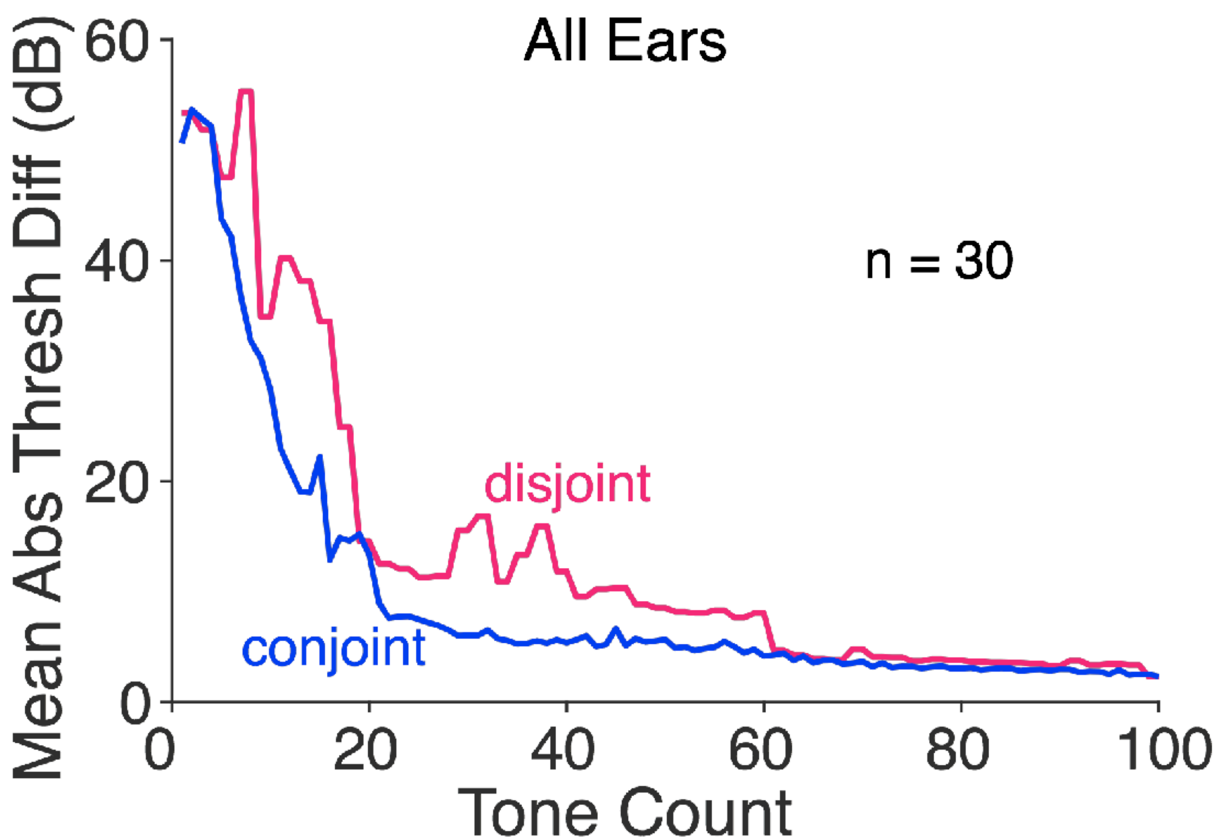
Active mutual conjoint estimation enables rapid assessment of both ears



Active mutual conjoint estimation enables rapid assessment of both ears

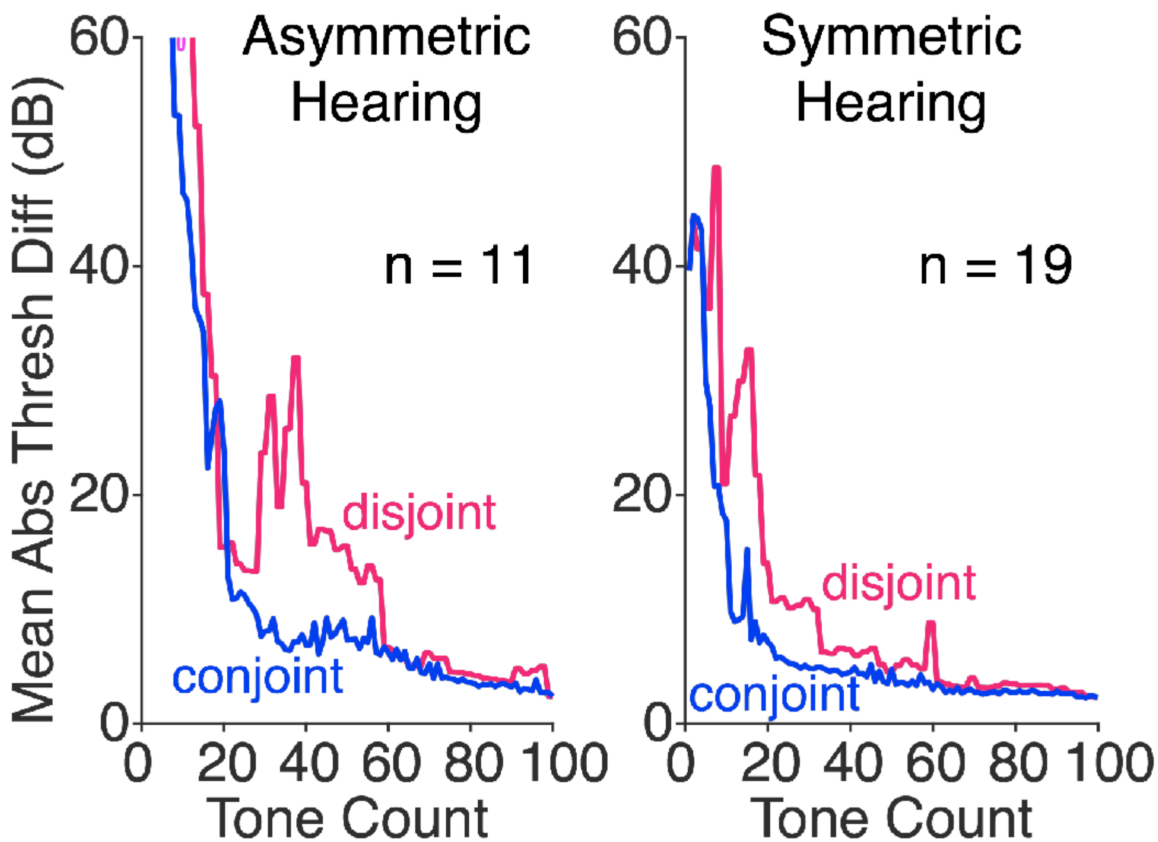


Active mutual conjoint estimation speeds up testing on average over disjoint



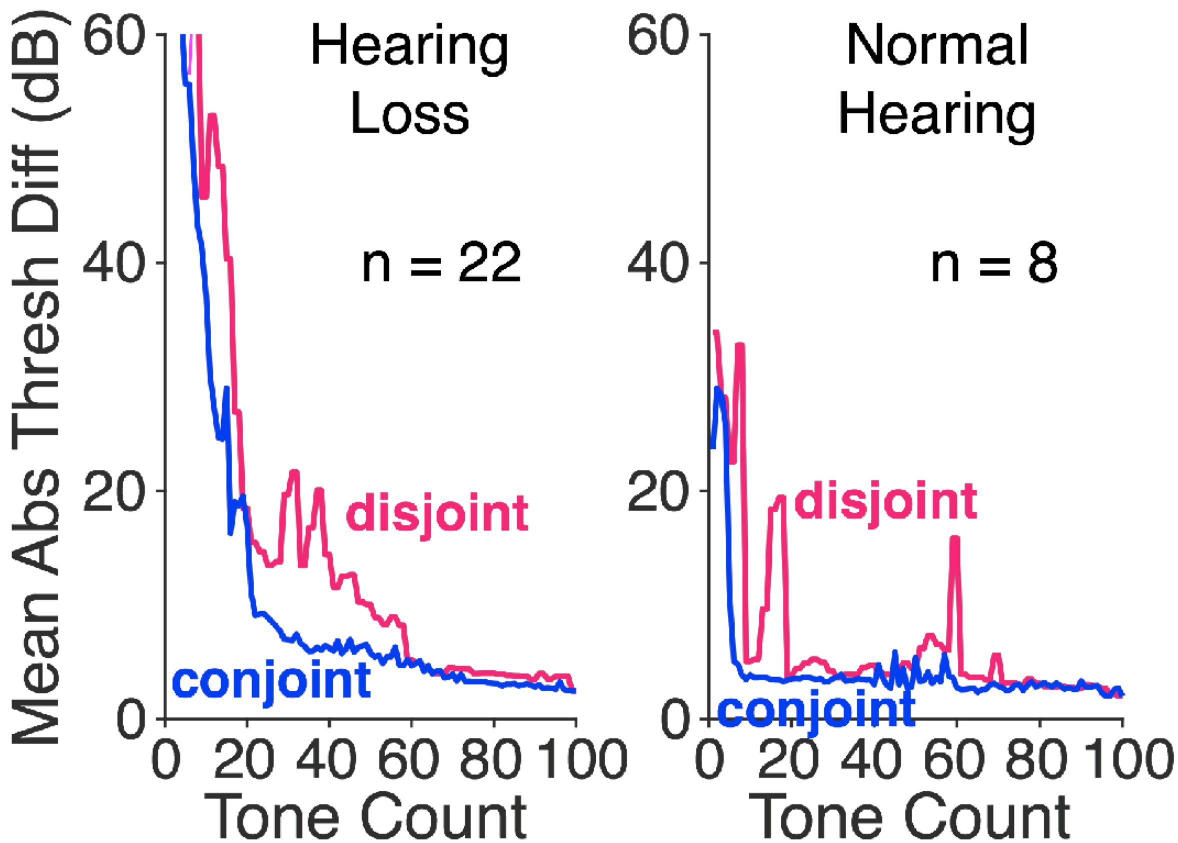
Heisey, Buchbinder & Barbour. *Acta Acustica* 104:762-765, 2018.

Active mutual conjoint estimation works for both symmetric and asymmetric hearing



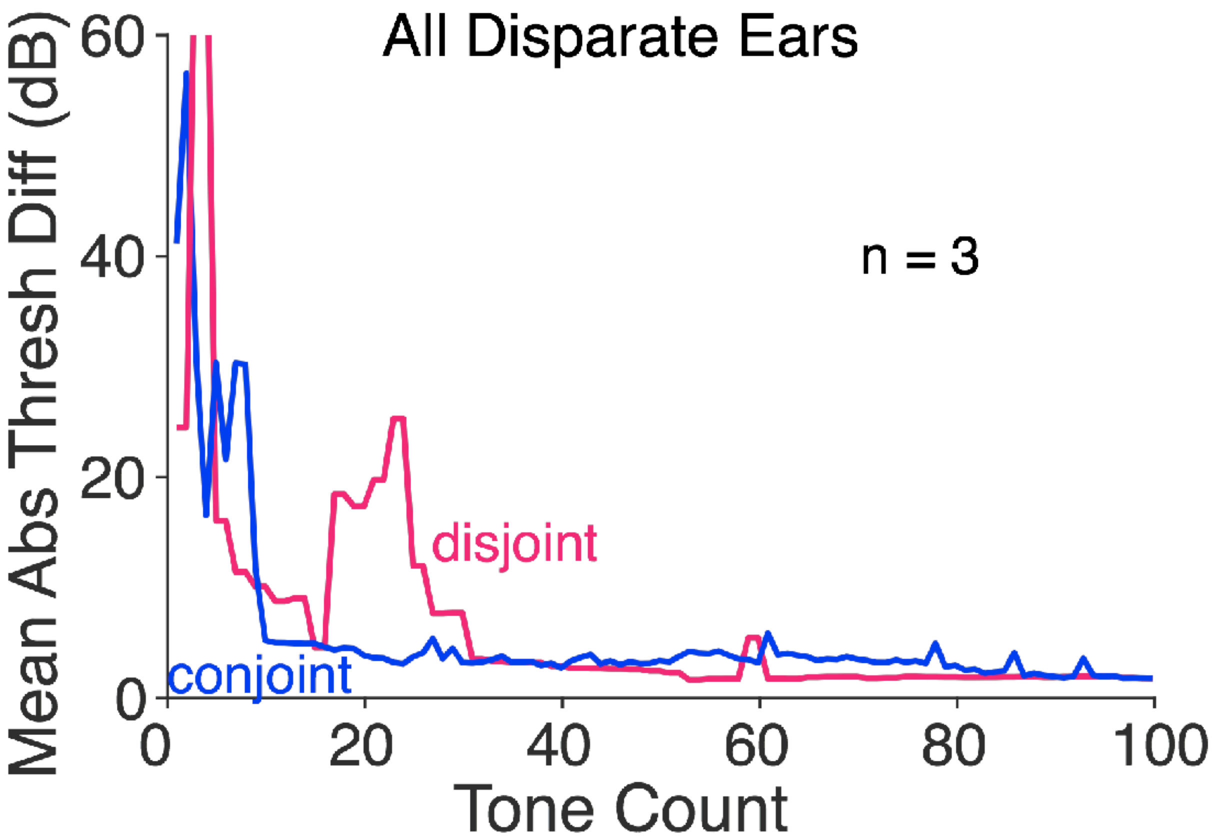
Heisey, Buchbinder & Barbour. *Acta Acustica* 104:762-765, 2018.

Active mutual conjoint estimation works for both normal and hearing loss



Heisey, Buchbinder & Barbour. *Acta Acustica* 104:762-765, 2018.

Active mutual conjoint estimation works across people



https://www.toonpool.com/cartoons/iPod_Sharing_65669; Heisey, Buchbinder & Barbour. *Acta Acustica* 104:762-765, 2018.

General active differential assessment

Online machine learning audiogram access

You can reach the test site here:

beta.bonauria.com

You can learn more about the web site, complete with instructions, here:

https://goo.gl/GcIpkK

And use these login credentials for the class:

Username: cse591
Password: HvpeCrkSv8zH45r5rAn

Acknowledgements

- David Song
- Katherine Heisey

- Jake Gardner
- Gustavo Malkomes

- Kilian Weinberger
- Roman Garnett
- John Cunningham

- Kiron Sukesan
- James DiLorenzo
- Braham Snyder
- Trevor Larson
- Steven Bosch

- Brittany Wallace
- Rebecca Howard
- Jenna Buchbinder

- Jeff Chen
- Jonathan Chen
- Ellie Degen
- Cate Jiang
- Kevin Xie
- Ramone Agard

Skandalaris Center, National Science Foundation, National Institutes of Health

Institute for Clinical and Translational Sciences

Center for Integration of Medicine and Innovative Technology