Sensory Cue Combination

Robert Jacobs
Department of Brain & Cognitive Sciences
University of Rochester

Sensory Redundancy

- Multiple sensory modalities
  - Vision
  - Audition
  - Touch
  - Proprioception
  - Smell
  - Taste

Visual Redundancy

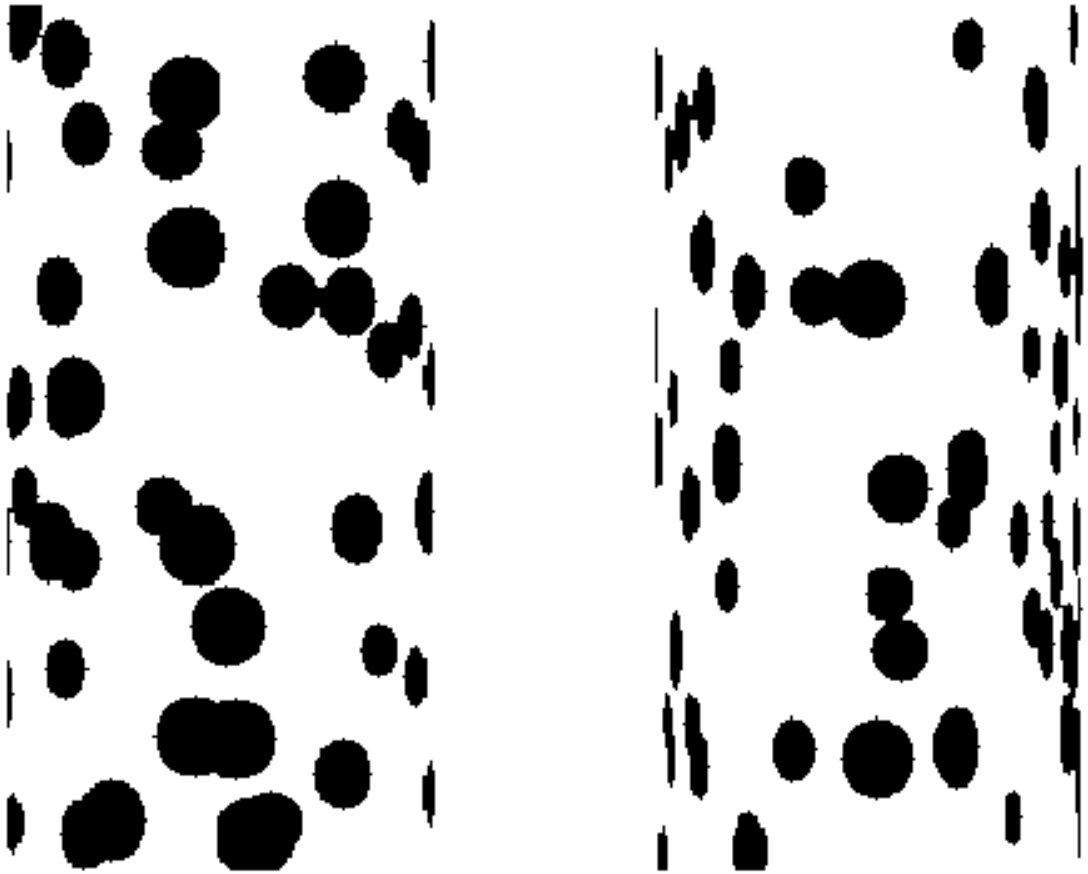
- Within visual modality, multiple cues
  - Motion
  - Binocular disparity
  - Texture gradients
  - Shading gradients
  - Linear perspective
  - Occlusion
  - Blur

Cue Combination

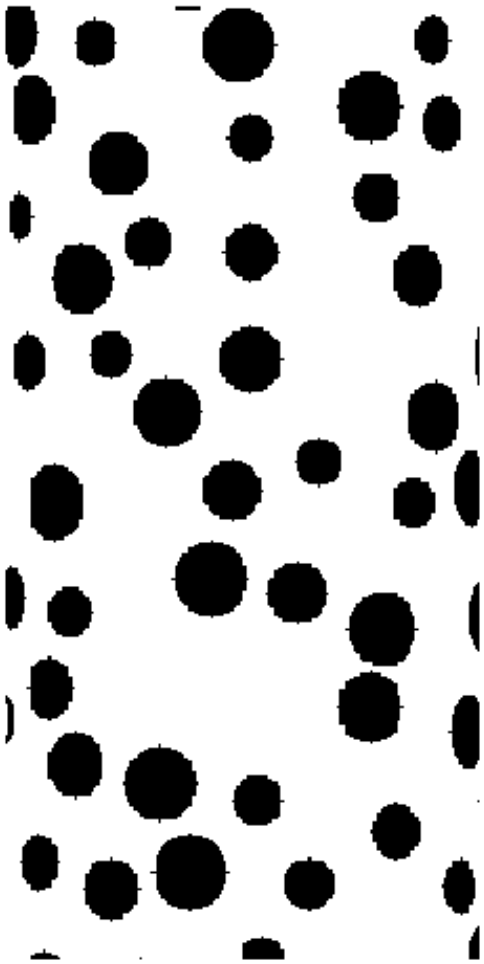
- Linear Cue Combination
  - Simplest (and most commonplace) sensory integration rule

- Example: Estimate visual depth based on motion and texture cues
  - Stage 1: Depth estimates based on individual cues are derived
  - Stage 2: Weighted combination of these estimates is used as the observer's composite depth percept

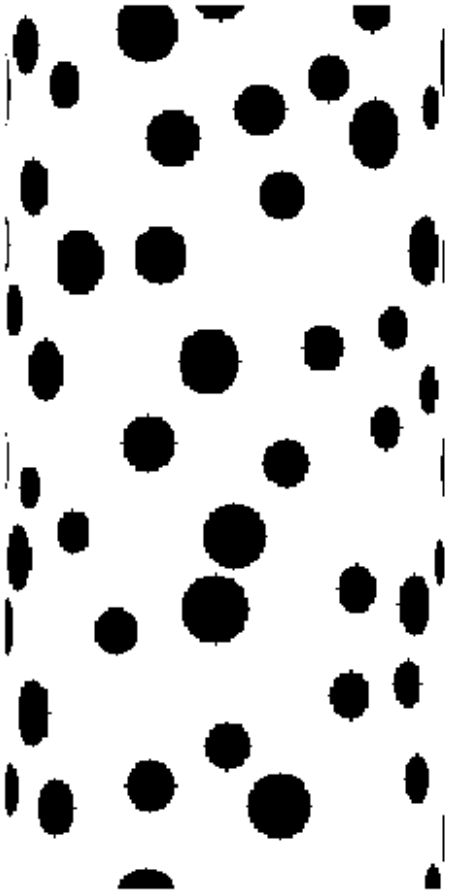
$$d(m,t) = w_M d_M(m) + w_T d_T(t)$$



Movie 1



Movie 2



Movie 3

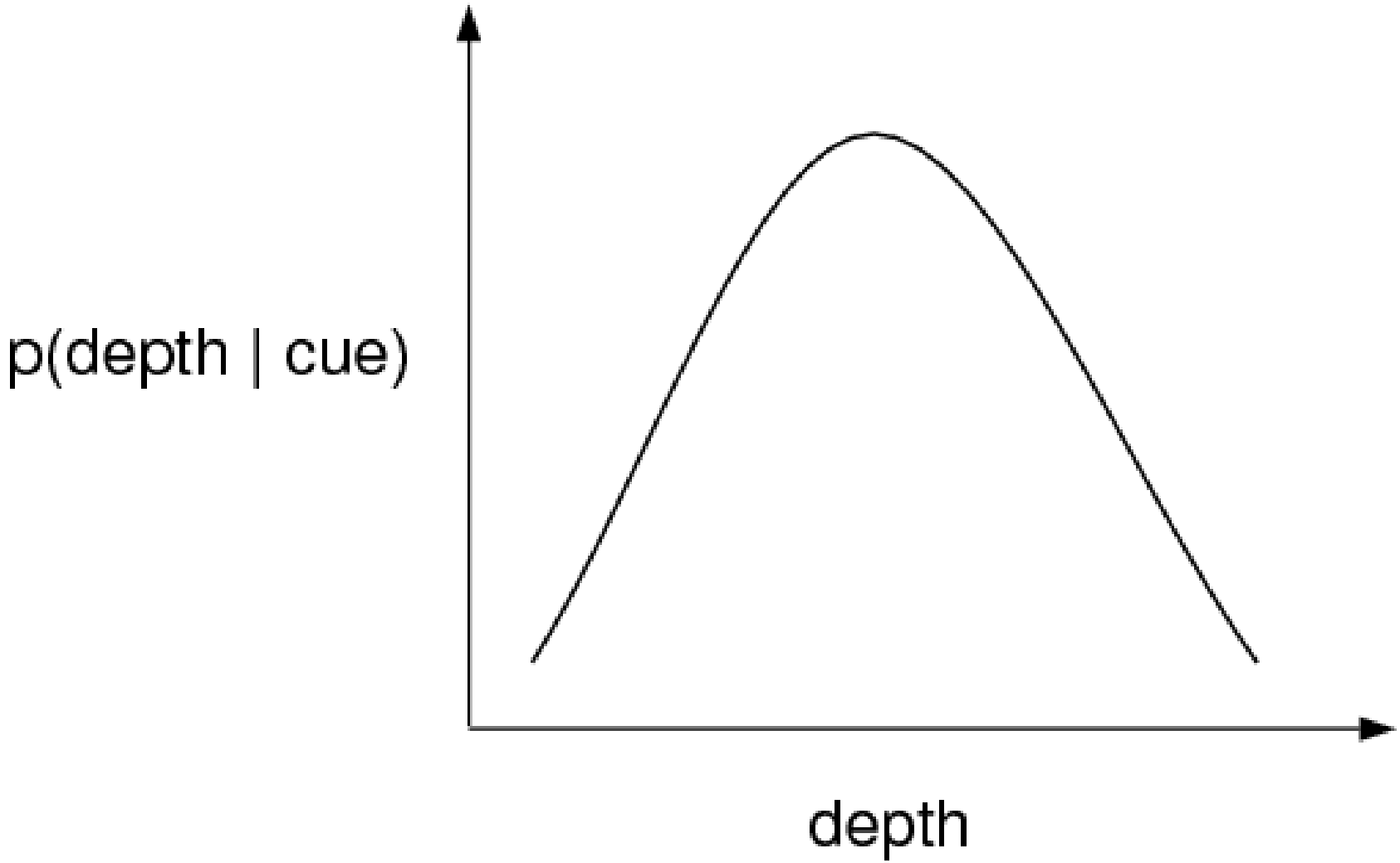
Linear Cue Combination

- Q: How do we choose the linear coefficients w_M and w_T?
- A: Cue weights should be based on cue reliabilities.

Linear Cue Combination

- Q: What is “cue reliability”?

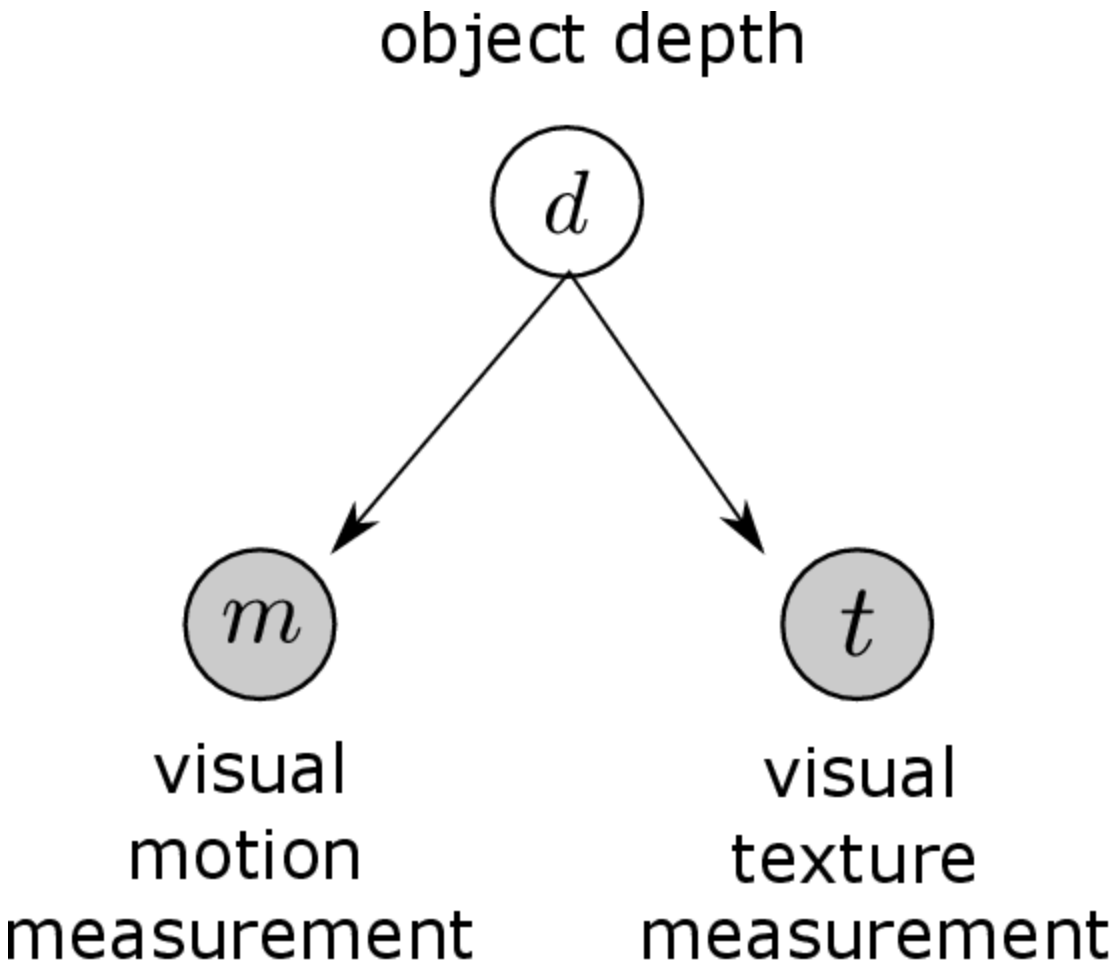
- Key observation: **EVERY CUE IS AMBIGUOUS!**
  - Physical factors: atmospheric or optical blurring
  - Biological factors: neural noise


p(depth | cue)
depth

Optimal Linear Cue Combination

- Maximum likelihood (or Kalman filter) theory of cue reliability
  - A cue is relatively reliable if the distribution of inferences based on that cue has a relatively small variance

Graphical Model



Mathematical Assumptions

- Prior probability *p(d)* is uniform
- Conditional probability distributions *p(m|d)* and *p(t|d)* are normal distributions with means equal to *d* and variances denoted σ_m^2 and σ_t^2
- Motion and texture cues are conditionally independent given scene parameter(s) of interest (i.e., *d*)

Tangent: Conditional Independence

- Conditional independence and independence are different
  - Motion and texture cues are **not** independent
    - Example: If motion cue indicates a small depth, then texture cue probably also indicates a small depth
    - However, after we've taken the depth d into account, then the motion and texture cues are conditionally independent
      - The noise in the motion cue (i.e., $d - d_m$) is independent from the noise in the texture cue (i.e., $d - d_t$).
      - In other words, the trial-by-trial variabilities (i.e., noise) in the motion and texture measurements are uncorrelated

- Conditional independence means:

$$p(m, t|d) = p(m|d)\, p(t|d)$$

- Independence would have been: $p(m, t) = p(m)\, p(t)$.
  However, this is not true here.

- Example: Alzheimer's disease and needing reading glasses are **not** independent (because they both tend to occur in older people). However, once we know that a person (or persons) is older, then Alzheimer's disease and needing reading glasses are conditionally independent.

- Example: Homicide rates in a city and ice cream sales are **not** independent (because they both tend to increase on hot days). However, once we take into account the outside temperature, the two are conditionally independent.

Posterior Distribution

- Posterior distribution: $p(d \mid m, t) \propto p(m \mid d)\, p(t \mid d)\, p(d)$
- Recalling that $p(d)$ is uniform:

$$p(d \mid m, t) \propto p(m \mid d)\, p(t \mid d)$$

- Posterior is proportional to the product of 2 normal likelihood functions
  - This is closely related to a situation that we discussed extensively in an earlier lecture

Optimal Linear Cue Combination

d_m^* = MLE estimate of visual depth based solely on the motion cue

= depth d that maximizes $p(m \mid d)$

d_t^* = MLE estimate of visual depth based solely on the texture cue

= depth d that maximizes $P(t \mid d)$

d^* = posterior depth estimate based on both cues

= depth d that maximizes $p(d \mid m, t)$

Optimal Linear Cue Combination

$$d^* = w_m d_m^* + w_t d_t^*$$

$$w_m = \frac{1/\sigma_m^2}{1/\sigma_m^2 + 1/\sigma_t^2}$$

$$w_t = \frac{1/\sigma_t^2}{1/\sigma_m^2 + 1/\sigma_t^2}$$

Note: Cue weights are non-negative and sum to one.

Optimal Linear Cue Combination

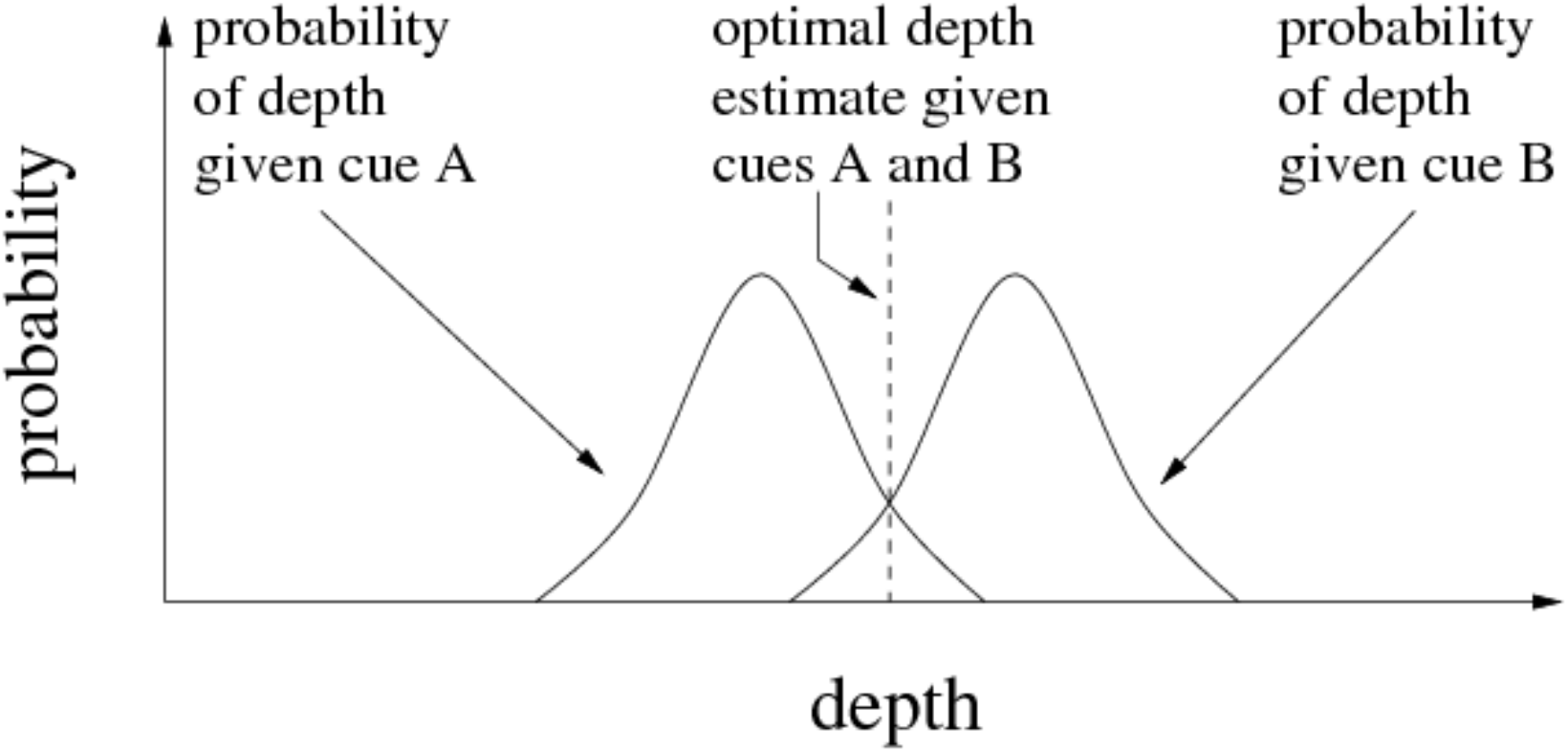
- Combine two Gaussian distributions [*p(m|d), p(t|d)*] to form new Gaussian distribution [*p(d|m,t)*]

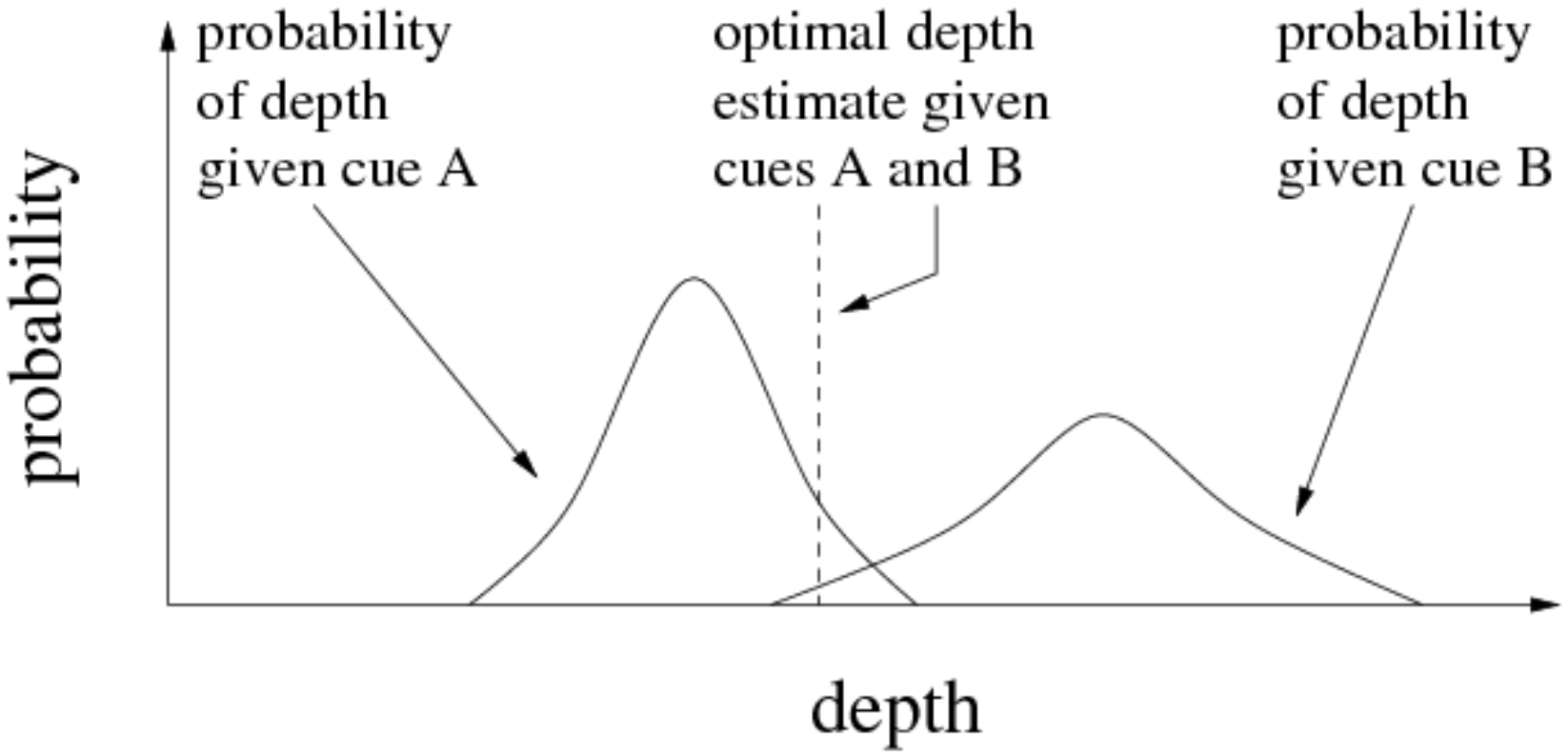
$$\text{Mean} = d^*$$

$$\text{Variance} = \frac{\sigma_m^2 \sigma_t^2}{\sigma_m^2 + \sigma_t^2}$$

$\rightarrow$ Variance is less than (or equal to) σ_m^2 or σ_t^2

- Combined depth estimate is more reliable than estimates based on individual cues!!!


probability
probability of depth given cue A
optimal depth estimate given cues A and B
probability of depth given cue B
depth


probability
probability of depth given cue A
optimal depth estimate given cues A and B
probability of depth given cue B
depth

Distribution of the MAP Estimate

- Because we assumed that the prior distribution $p(d)$ is uniform, the distribution of the MAP estimate $p\left(\hat{d}_{\text{MAP}}|d\right)$ is identical to the posterior distribution $p(d|m,t)$

  – Both the mean of the MAP distribution and the mean of the posterior distribution are unbiased estimates of d

- This stands in contrast to the situation discussed in a previous lecture in which the posterior distribution was proportional to the product of a normal prior distribution and a normal likelihood function.

  - Mean of the MAP distribution is equal to the mean of the posterior distribution
    - The means of these distributions are biased estimates

  - Variance of the MAP distribution **differs** from the variance of the posterior distribution

Battaglia, P.W., Jacobs, R.A, and Aslin, R.N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A, 20*, 1391-1397.

Integration of Visual and Auditory Signals for Spatial Localization

- Battaglia, Jacobs, and Aslin (2003)
- Cue-consistent versus cue-conflict environments
- Two models for cue-conflict situations:
  - Winner-take-all (Visual Capture)
  - Maximum Likelihood Estimation (MLE)

Maximum Likelihood Estimation

- Statistically optimal cue combination rule (given certain mathematical assumptions)

- Linear combination rule:

$$L^*(v,a) = w_V L_V^*(v) + w_A L_A^*(a)$$

Procedure

- Auditory-only trials
  - Estimate auditory mean and variance
- Visual-only trials
  - Estimate visual mean and variance
- Compute predictions of two models on visual-auditory trials:
  - Compute Visual Capture predictions
  - Compute MLE predictions
- Visual-Auditory trials

Virtual Reality Environment

Auditory Stimuli

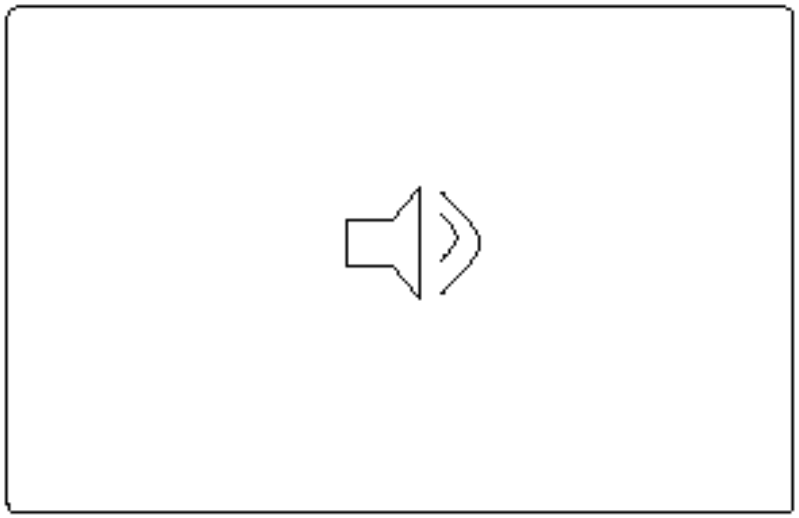
- Broadband noise filtered to mimic the spectral characteristics of a sound source external to the listener

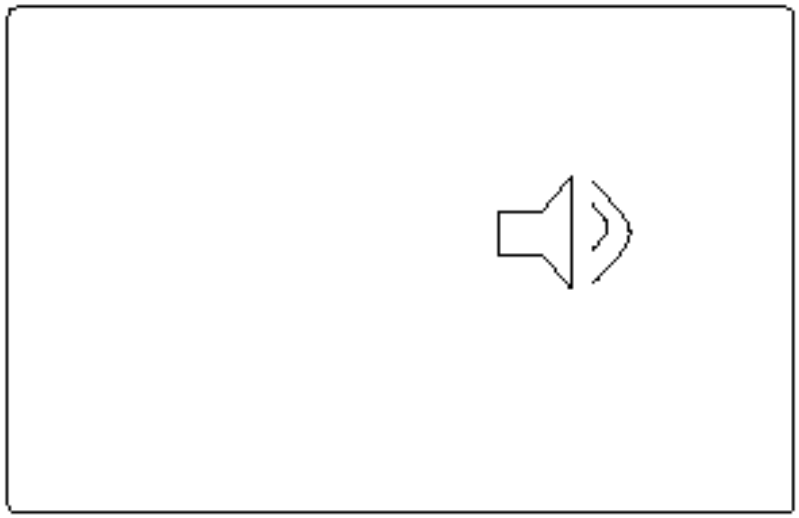
Visual Stimuli

- Random-dot stereogram of a bump protruding from a background surface

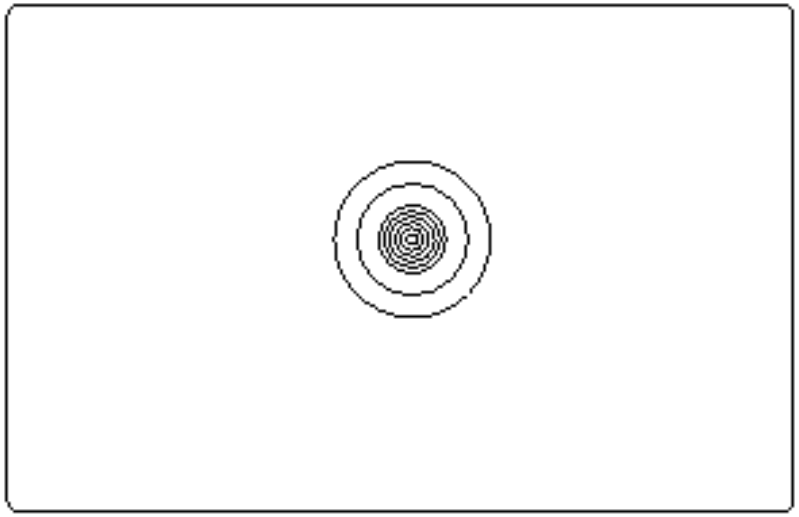
- Noise could be added to visual stimulus
  - Low noise: easy to detect and localize bump
  - High noise: difficult to detect and localize bump

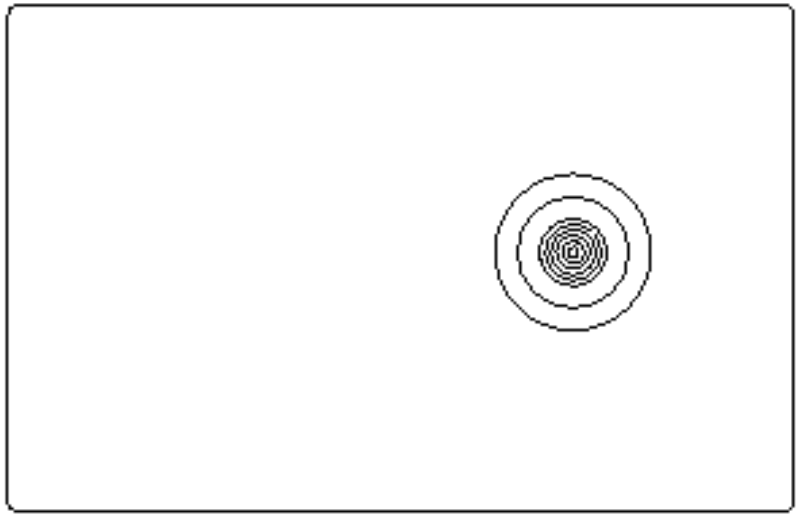
Auditory-Only Trial



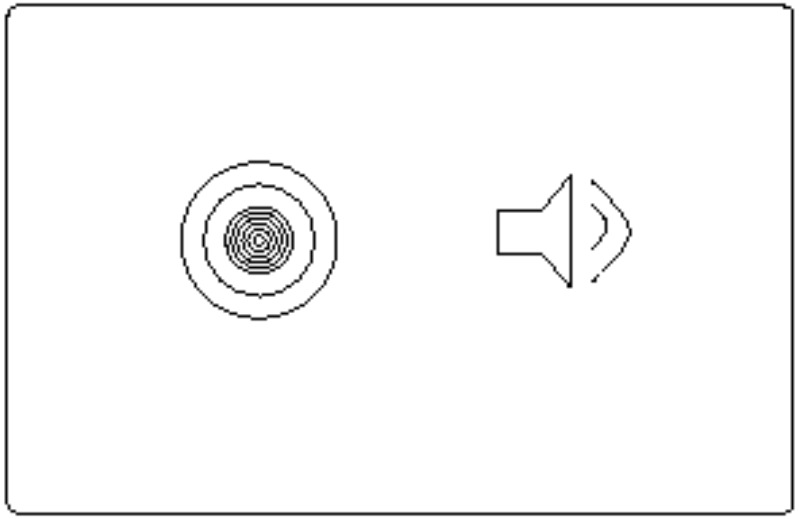


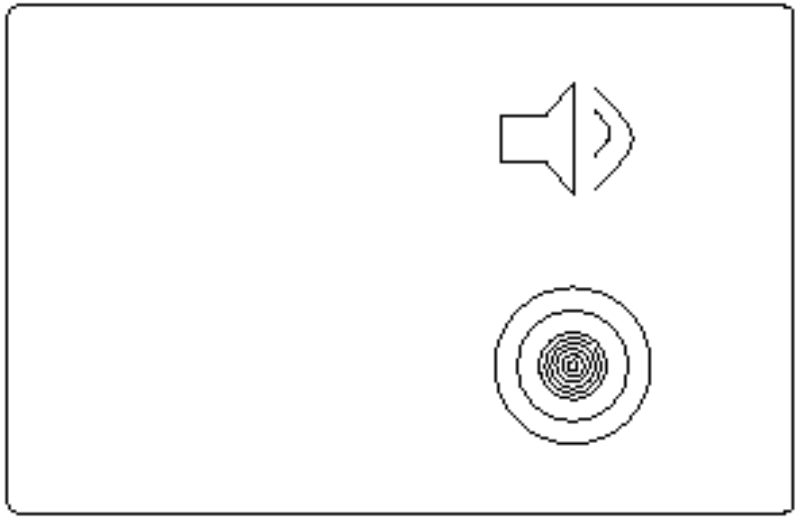
Visual-Only Trial



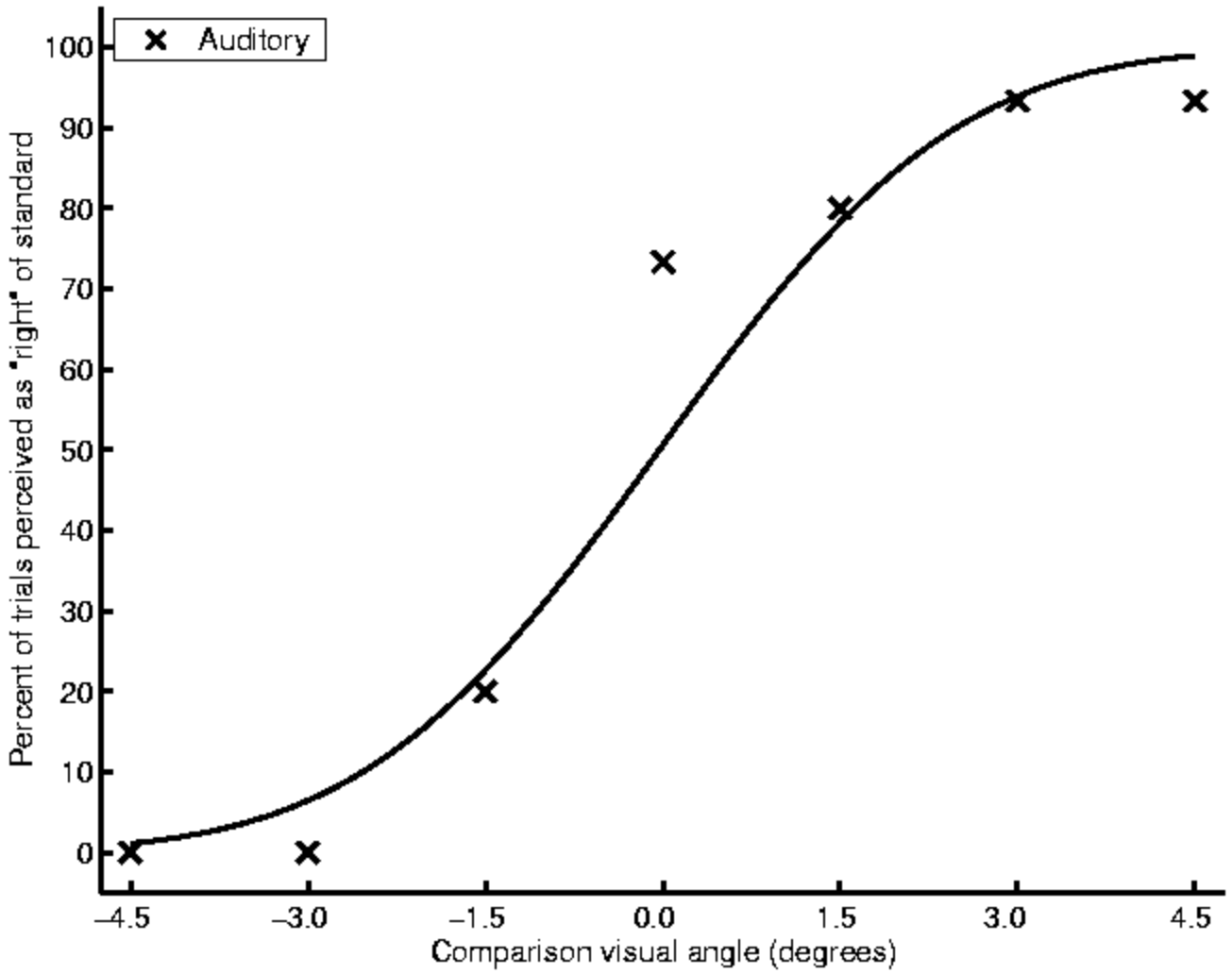


Visual-Auditory Trial

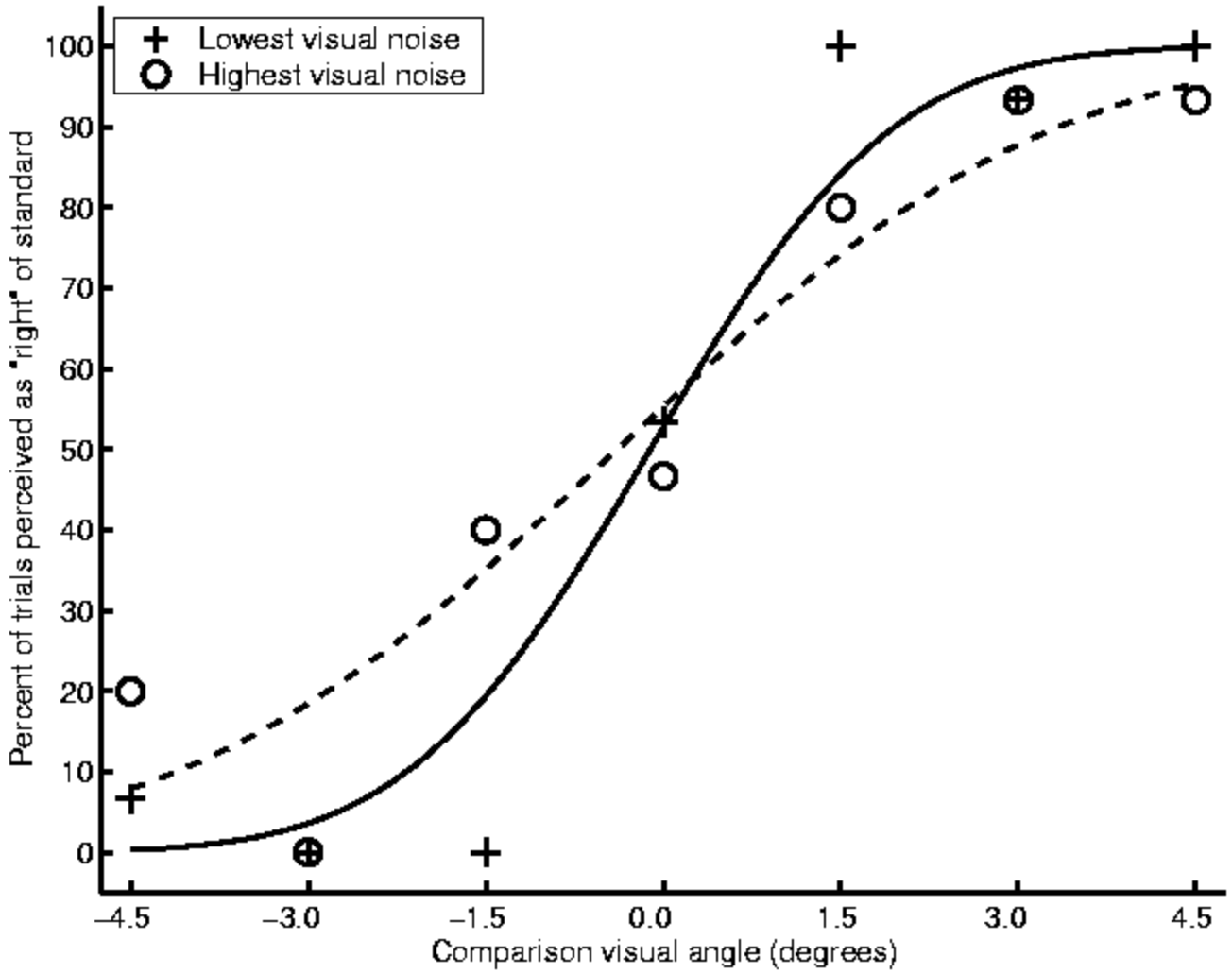




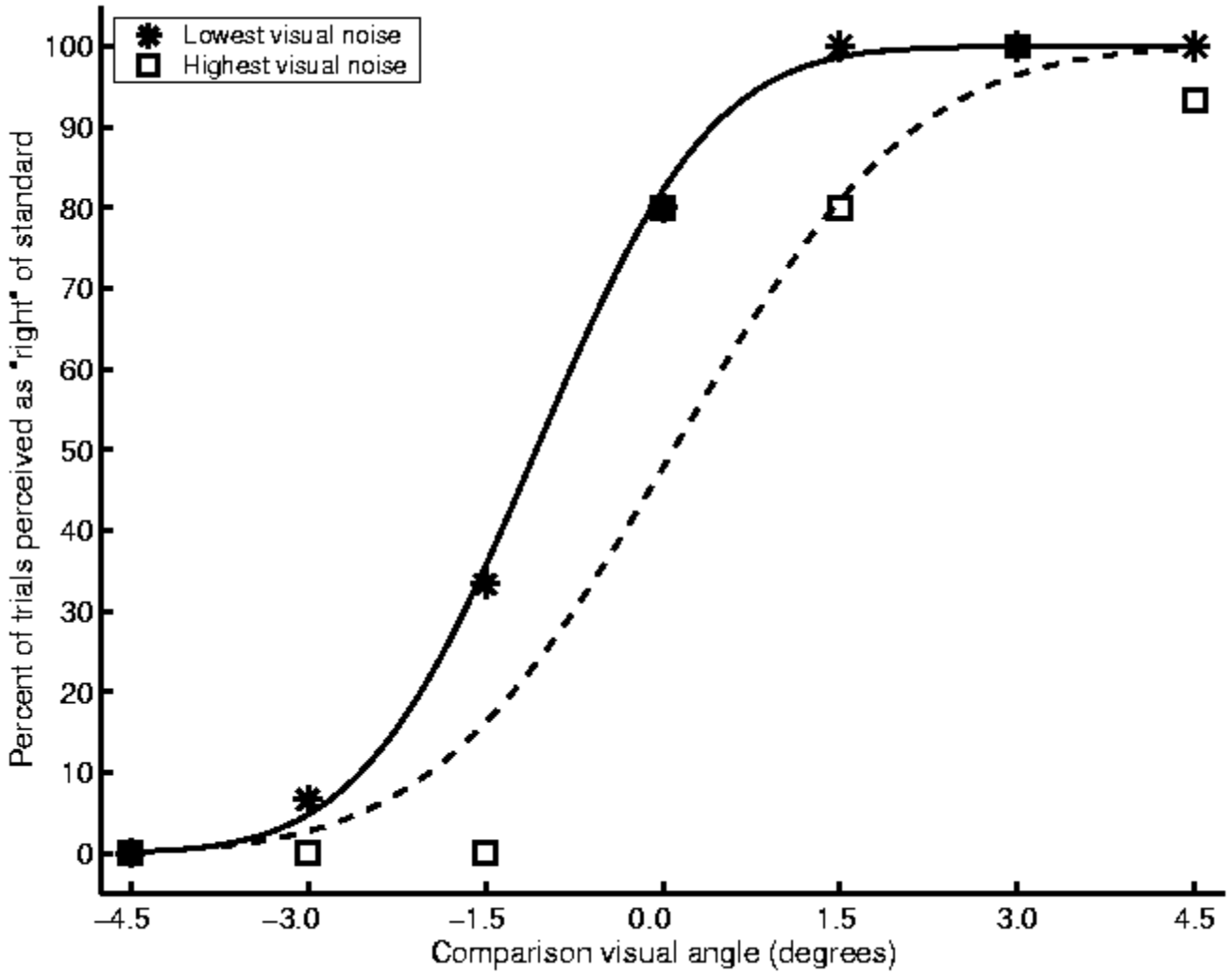
Subject EJA: Auditory-Only Trials



Subject EJA: Visual-Only Trials



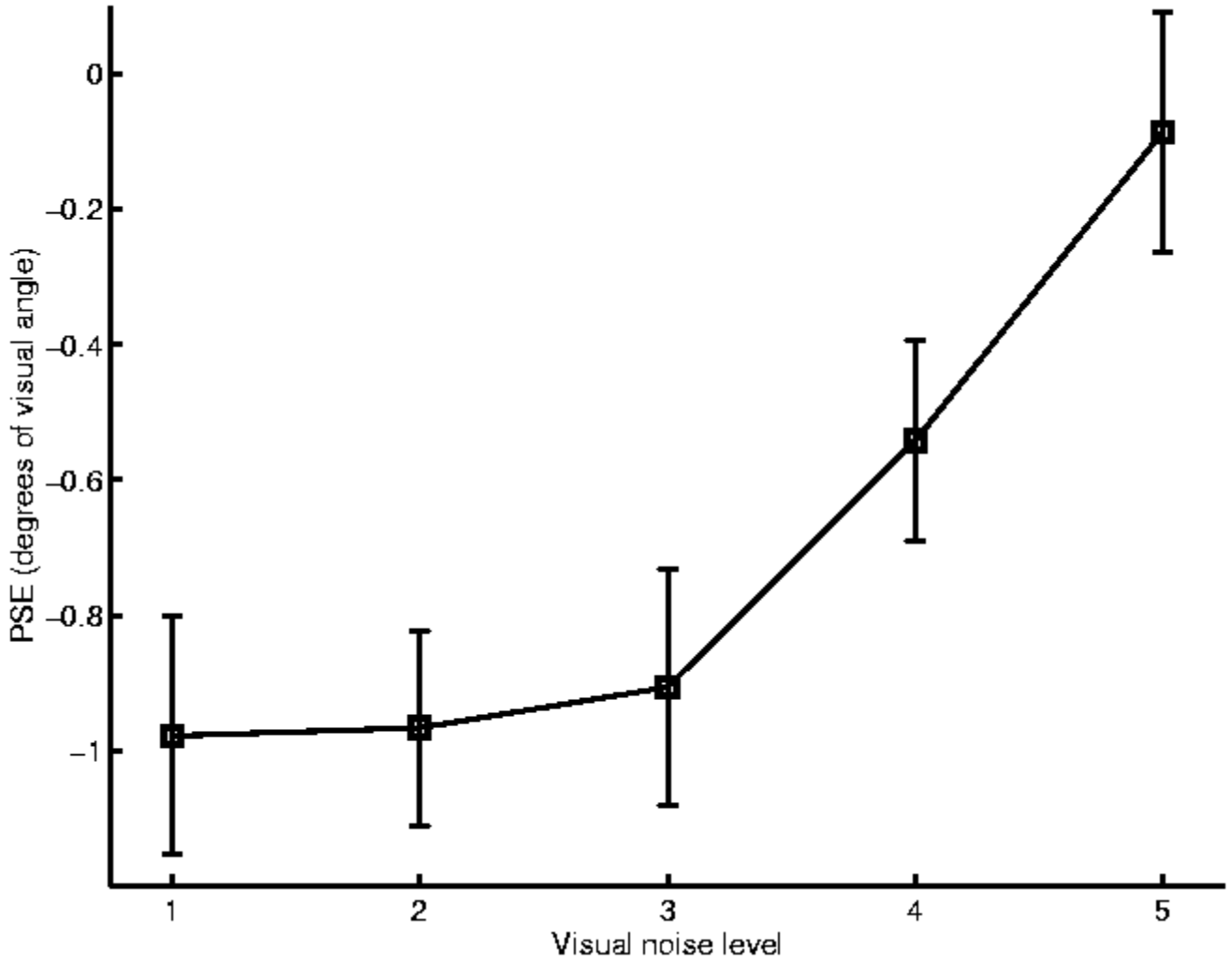
Subject EJA: Visual-Auditory Trials



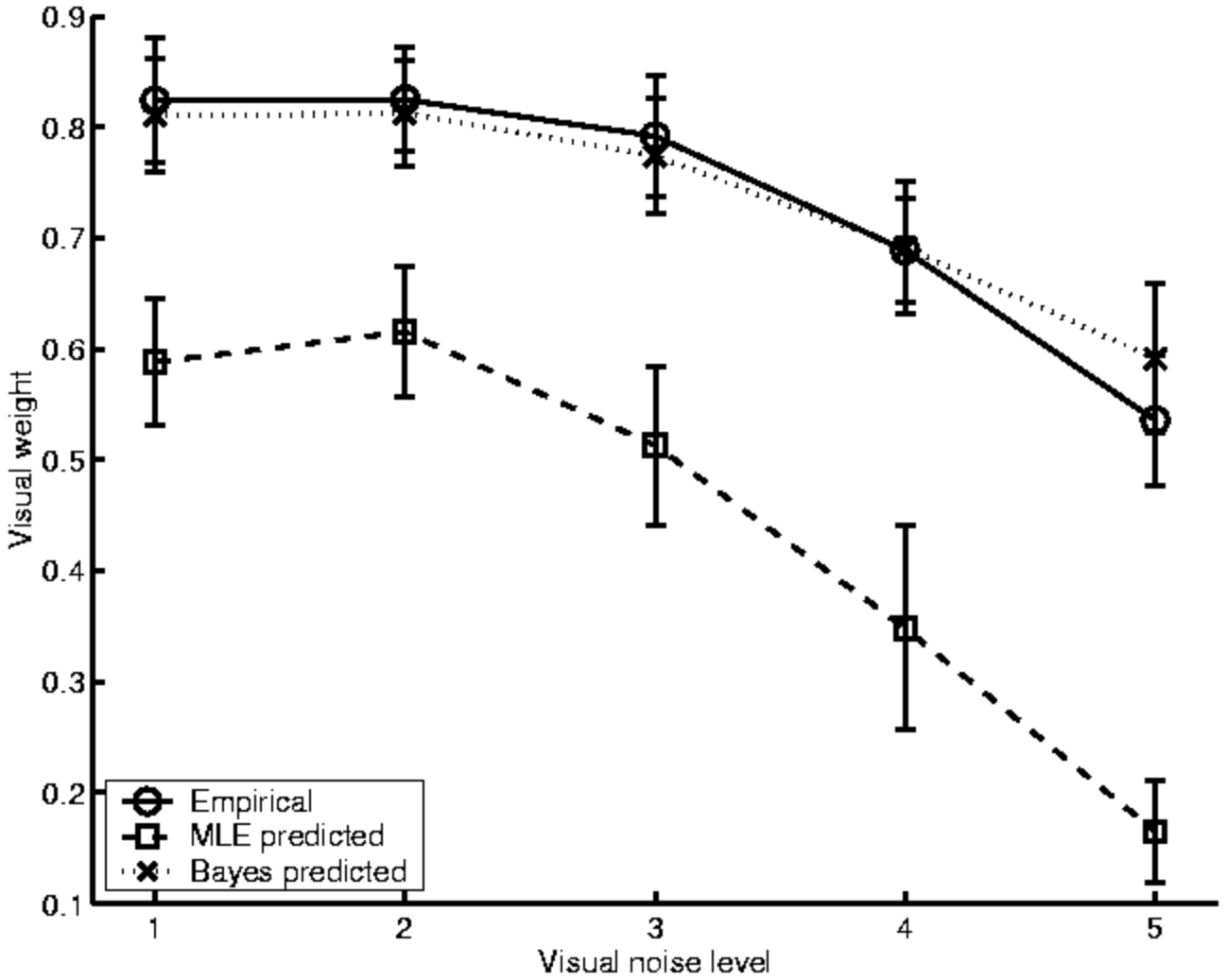
Subject EJA: Visual-Auditory Trials

- Low visual noise:
  - PSE near location of visual stimulus
  - Visual weight near one; auditory weight near zero
- High visual noise:
  - PSE at a location between visual and auditory stimuli
  - Visual and auditory weights each near one-half

PSE as a function of visual noise



Visual weight as a function of visual noise



Models

- MLE partially correct
  - Signal reliability significantly altered judgments of spatial location
- Visual Capture partially correct
  - Judgments are characterized by an overall bias to rely on visual over auditory information
- Hybrid model may be best

Vul, E. & Pashler, H. (2008). Measuring the crowd within. *Psychological Science, 19*, 645-647.

Wisdom of Crowds

- If people guess the weight of an ox, the error of the average response is smaller than the average error of individual estimates (Galton, 1907)

- Wisdom of crowds:
  - Crowd's average is more accurate as long as some of the error of one individual is statistically independent of the error of other individuals

- Q: Does a wisdom-of-crowds effect occur when we average responses from a single individual?

- **Experiment:** 428 subjects answered eight questions about world knowledge (e.g., What percentage of the world's airports are in the United States?)

  – Half the subjects were (unexpectedly) asked to make a second, different guess

  – Remaining subjects made a 2nd guess three week later

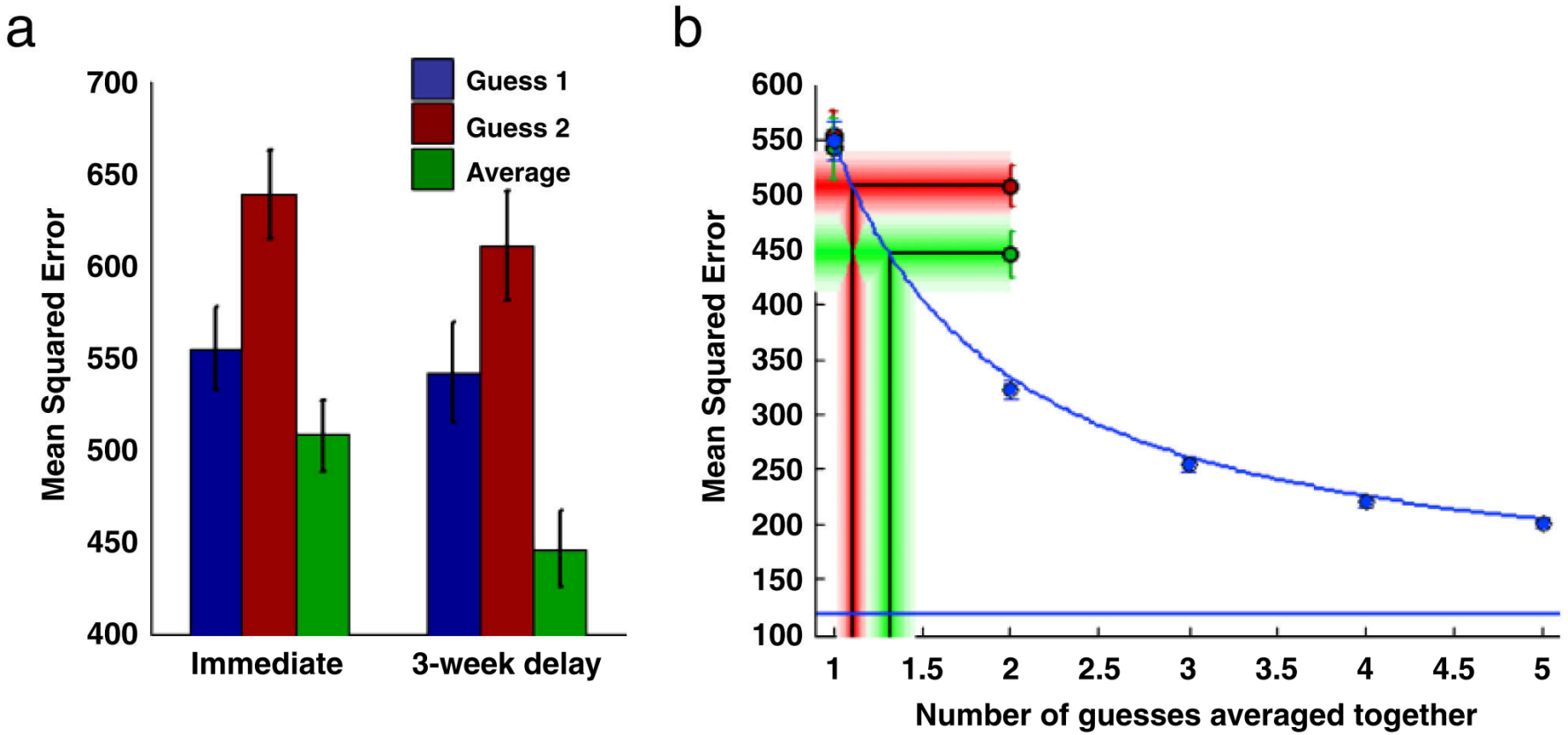


Fig. 1. Experimental results. The bar graph (a) presents mean squared error for the first and second guesses and their average, as a function of condition (immediate vs. 3-week delay). The line graph (b) shows mean squared error as a function of number of guesses averaged together. The data points show results for guesses from independent subjects (blue), a single subject in the immediate condition (red), and a single subject in the delayed condition (green). The blue curve shows convergence to the population bias, which is indicated by the horizontal blue line (the error of the guess averaged across all people). Through interpolation (black lines), we computed the value of two guesses from one person relative to two guesses from independent people, for both the immediate and the delayed conditions. The shaded regions are bootstrapped 90% confidence intervals. Error bars represent standard errors of the means.

- **Conclusion:** Although people assume that their 1st guess exhausts the best information available to them, a forced 2nd guess contributes additional information

- Suggests that:
  - Responses made by a subject are samples from an internal posterior probability distribution (rather than deterministically selected on the basis of all the knowledge a subject has)
  - Benefits obtained by averaging these samples over time